UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA

Programa de Pós-Graduação em Química

ALBERTO MALVEZZI

Modelos de *Virtual Screening* de Inibidores da Cruzaína: Desenvolvimento e Validação Experimental

São Paulo
18/03/2008

0
T
547.13
M262m

BIBLIOTECA
INSTITUTO DE QUÍMICA
Universidade de São Paulo
22.820

ALBERTO MALVEZZI

Modelos de *Virtual screening* de Inibidores da Cruzaína: Desenvolvimento e Validação Experimental

Tese apresentada ao Instituto de Química da Universidade de São Paulo para obtenção do Título de Doutor em Química Orgânica

Orientadora: Prof[a]. Associada Antonia Tavares do Amaral

São Paulo
18/03/2008

DEDALUS - Acervo - CQ

30100013565

Ficha Catalográfica
Elaborada pela Divisão de Biblioteca e
Documentação do Conjunto das Químicas da USP.

Malvezzi, Alberto
M262m Modelos de *Virtual screening* de inibidores da cruzaína: desenvolvimento e validação experimental / Alberto Malvezzi. -- São Paulo, 2008.
189p.

Tese (doutorado) – Instituto de Química da Universidade de São Paulo. Departamento de Química Fundamental.
Orientador: Amaral, Antonia Tavares do

1. Composto bioativo : Físico-química : Orgânica 2. Fármaco : Planejamento : Química farmacêutica 3. Química medicinal 4. Relações quantitativas entre estrutura química e atividade biológica (QSAR) : Físico-química : Orgânica I. T. II. Amaral, Antonia Tavares do, orientador.

547.13 CDD

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA

"Modelos de *virtual screening* de inibidores da cruzaína: desenvolvimento e validação experimental"

ALBERTO MALVEZZI

Tese de Doutorado submetida ao Instituto de Química da Universidade de São Paulo como parte dos requisitos necessários à obtenção do grau de Doutor em Química – Área: Química Orgânica.

Aprovado(a) por:

Profa. Dra. ANTONIA TAVARES DO AMARAL
(Orientadora e Presidente)

Prof. Dr. OMAR ABDEL MONEIM ABOU EL SEOUD
IQ – USP

Prof. Dr. SHAKER CHUCK FARAH
IQ – USP

Profa. Dra. ELIZABETH IGNE FERREIRA
FCF - USP

Prof. Dr. RICHARD CHARLES GARRATT
IFSC - USP

SÃO PAULO
09 DE MAIO 2008

Para minha esposa Cá,
meu Kilimanjaro.

AGRADECIMENTOS

Agradeço a Prof.ª Antonia Tavares do Amaral, que me deu tudo o que foi preciso para que esta tese se materializasse.

Agradeço ao amigo Leandro de Rezende, sem o qual a porção *virtual* desta tese não seria possível.

Agradeço a todos os colegas do laboratório, atuais e que já passaram, pelas discussões científicas e incentivos durante os anos de elaboração desta tese.

Agradeço ao Prof. Dr. Luiz Juliano do Departamento de Biofísica, Universidade Federal do Estado de São Paulo-SP (UNIFESP-SP) por ceder amostras da enzima Cruzaína, fundamentais para a validação experimental dos modelos propostos nesta tese.

Agradeço aos professores: Profª. Drª. Shirley Schreier do Departamento de Bioquímica do Instituto de Química – Universidade de São Paulo – SP; Prof. Dr. Willem Baader e Prof. Dr. Cassius Vinicius Stevani do Departamento de Química Fundamental do Instituto de Química – Universidade de São Paulo – SP; e Prof. Dr. Luiz Juliano do Departamento de Biofísica, Universidade Federal do Estado de São Paulo-SP, por disponibilizarem seus laboratórios para o uso dos respectivos aparelhos de fluorescência.

Agradeço à Auxiliar Nê e ao Técnico Enio (*in memorian*) pela convivência sempre agradável no laboratório.

Agradeço à FAPESP que financiou esta tese (Processo 02/09518-8) e o projeto temático (Processo 2001/01192-3) ao qual esta tese está inserida. Agradeço ao CCDC/CSD e às Agencias CAPES-DAAD e CNPq.

Agradeço aos meus pais e minhas irmãs que me deram o suporte necessário para iniciar e terminar esta tese.

Take what you need and do what you should,
and almost always,
you get what you want.

ÍNDICE

Página de Rosto ii
Folha de Aprovação iii
Dedicatória iv
Agradecimentos v
Epígrafe vi
Índice vii
Lista de Tabelas ix
Lista de Figuras xi
Lista de abreviaturas e Siglas xviii
Resumo xx
Abstract xxii
Objetivos xxiv
1. Introdução 1
1.1. A Descoberta de Fármacos 1
1.2. *HTS* e Química Combinatória 4
1.3. Resultados Falso-positivos 11
1.3.1. *Frequent hitters* 13
1.3.2. Compostos reativos frente à proteínas 14
1.3.3. O mecanismo promíscuo 15
1.4. *Virtual screening* 19
1.4.1. Filtros *drug-like* 23
1.4.2. Filtros farmacofóricos 30
1.4.3. Filtros de *Docking* 38
1.5. Descritores de interação molecular 51
1.6. Doença de Chagas 64
1.6.1. Epidemiologia, Transmissão e Sintomas 64
1.6.2. Terapias em uso 64
1.6.3. Planejamento de novos compostos antichagásicos 66
2. Materiais e Métodos 74
2.1. Reagentes e Solventes 74
2.2. Equipamentos 75
2.3. Bancos de dados de moléculas virtuais 76
2.4. Programas computacionais 77
2.5. Procedimentos computacionais 79
2.5.1. Manipulação e alinhamento das estruturas extraídas do *PDB* 79

2.5.2. Obtenção dos campos de interação molecular 80
2.5.3. Procura por cavidades similares à cruzaína 81
2.5.4. Dinâmica molecular 81
2.5.5. Preparação e aplicação do modelo farmacofórico (**Modelo I** e **II**) 83
2.5.6. Filtros de propriedades físico-quimicas (**Modelo I** e **II**) 84
2.5.7. *Docking* (**Modelo I** e **II**) 85
2.6. Teste de inibição da cruzaína (**Modelo I**) 86
2.7. Verificação do mecanismo promíscuo de inibição enzimática (**Modelo I**) 87
2.8. Teste de inibição da cruzaína na presença de detergente (**Modelo II**) 89
3. Resultados e discussão 91
3.1. Introdução 91
3.2. Reconhecimento da estrutura da cruzaína e características do sítio ativo 96
3.2.1. Reconhecimento por inspeção visual 99
3.2.2. Reconhecimento por campos de interação molecular 101
3.2.3. Reconhecimento por pseudocentros 103
3.2.4. Busca por cavidades similares 105
3.2.5. Reconhecimento por dinâmica molecular 110
3.3. Obtenção e validação de modelo de *virtual screening* de inibidores da cruzaína - **Modelo I** 115
3.3.1. Filtro Físico-Químico 116
3.3.2. Filtro farmacofórico 117
3.3.3. Filtro de *docking* 119
3.3.4. Filtro visual e seleção dos compostos para o teste de inibição enzimática 122
3.3.5. Testes Enzimáticos e Validação Experimental do Modelo de *Virtual screening* 124
3.4. Obtenção e Validação de modelo de *virtual screening* de inibidores da cruzaína - **Modelo II** 130
3.4.1. Filtro farmacofórico 132
3.4.2. Filtro Físico-Químico 137
3.4.3. Filtro de *Docking* 141
3.4.4. Filtro visual e seleção dos compostos para o teste de inibição enzimática 148
3.4.5. Testes de inibição enzimática e validação experimental do modelo de *virtual screening* 152
4. Conclusões 167
5. Referencias 171

LISTA DE TABELAS

Tabela 1. Resultados da busca por inibidores da tirosina fosfatase-PTP1B utilizando as estratégias de *HTS* e *Docking*. Número de compostos testados, número de compostos ativos identificados e porcentagem de compostos ativos [Doman *et al.*, 2002] 22

Tabela 2. Alvos moleculares submetidos ao procedimento de *docking*[Kubinyi, 2006] .50

Tabela 3. Valores dos parâmetros termodinâmicos K_b, ΔG^0, ΔH^0 e $T\Delta S^0$ para a formação do complexo com a streptavidina-ligante, determinados por microcalorimetria[WEBER *et al.*, 1992] 52

Tabela 4. Valores[a] da contribuição média de cada grupo para a afinidade do complexo *RL* [ANDREWS *et al.*, 1984] 60

Tabela 5. Principais tipos de interações entre um composto e o sistema biológico, um exemplo e as respectivas faixas de energias envolvidas.....61

Tabela 6. Códigos *PDB,* resolução e representação da fórmula estrutural dos inibidores presentes nos treze complexos da cruzaína, respectivamente. As estruturas dos inibidores foram modificadas para representar os compostos antes da formação do complexo covalente. Os círculos vermelhos indicam o átomo de carbono que se liga covalentemente à enzima. 97

Tabela 7. Valores de *RMSD* das 13 estruturas da cruzaína alinhadas átomo-a-átomo, em relação ao complexo 1AIM[a]. 99

Tabela 8. Códigos *PDB,* resolução e representação da fórmula estrutural dos inibidores presentes, respectivamente, nos 20 complexos selecionados pelo programa *CAVBASE*. As estruturas dos inibidores foram modificadas para representar os compostos antes da formação do complexo covalente. Os círculos em vermelho indicam o átomo de carbono que se liga covalentemente à enzima.....107

Tabela 9. Valores médios de *RMSD* (± desvios padrão) observados nas trajetórias[a] da cruzaína[b] sem o ligante e complexada com o ligante 1ME4, para todos os átomos, respectivamente, da proteína; do sítio ativo; de carbono-alfa e de cadeias laterais. 112

Tabela 10. Valores das distâncias, respectivamente, média, máxima e mínima, observadas nas trajetórias[a] da cruzaína[b] livre e a correspondente complexada, entre átomos do sítio ativo. 114

Tabela 11. Valores das propriedades físico-químicas utilizadas para selecionar inibidores da cruzaína. 117

Tabela 12. Resultados dos *dockings*[a] dos treze complexos do *PDB*[b], apresentando os correspondentes valores de *RMSD* da solução mais próxima à pose do cristal e a posição relativa [c] desta pose, em relação às outras possíveis poses encontradas. 120

Tabela 13. Representação esquemática dos 6 compostos selecionados pelo **Modelo I** de inibidores da cruzaína e que foram adquiridos e submetidos ao teste de inibição enzimática. São apresentados os correspondentes valores de massa molecular, logP[a] e empresa fornecedora, de cada um dos compostos adquiridos. Os círculos em vermelho indicam o

átomo de carbono eletrofílico susceptível ao ataque nucleofílico do enxofre catalítico da *Cys25*. 124

Tabela 14. Valores da massa molecular, do *logP*, do número de grupos doadores de H, do número de grupos aceptores de ligação de H, da soma dos átomos de Nitrogênios e Oxigênio, do número de halogênios, do número de anéis, do tamanho dos anéis, do número de ligações com rotação livre e, do número de grupos metalinas em cadeias alifáticas, para os 33 inibidores de cisteíno-protease selecionados..... 139

Tabela 15. Valores das propriedades físico-químicas utilizadas para selecionar inibidores da cruzaína. 140

Tabela 16. Valores de Fitness[(a)] e de RMSD[(b)], calculados para os correspondentes 33 compostos da série de validação docados [(c)] sobre o sitio de ligação da estrutura da cruzaína, extraída do complexo 1ME4. Os valores em vermelho indicam os ligantes que foram corretamente posicionados, segundo os critérios de avaliação dos resultados do *docking* descritos no texto. 143

Tabela 17. Valores de Fitness[(a)] e de RMSD[(b)], calculados para os 33 compostos da série de validação docados [(c)] sobre o sitio de ligação da estrutura da cruzaína 1ME4, modificada com a mutação *C25G*. Os valores em vermelho indicam os ligantes que foram corretamente posicionados, segundo os critérios de avaliação dos resultados do *docking* descritos no texto. 145

Tabela 18. Representação estrutural esquemática dos 19 compostos selecionados pelo **Modelo II** de *virtual screening* e que foram adquiridos para serem submetidos ao teste de inibição da cruzaína. São apresentados os correspondentes códigos do banco de dados ZINC , os valores de massa molecular e a empresa fornecedora, de cada um dos compostos adquiridos. Os círculos em vermelho indicam o átomo de carbono eletrofílico susceptível ao ataque nucleofílico do enxofre catalítico da *Cys25*. 150

LISTA DE FIGURAS

Figura 1. Representação esquemática das estruturas da Morfina e da Aspirina, respectivamente, o primeiro principio ativo natural isolado e o primeiro medicamento semi-sintético produzido em laboratório 1

Figura 2. Exemplo ilustrativo de aplicação do procedimento de *Mix & Split* partindo-se de três reagentes (A, B e C) e com três ciclos de repetição [FURKA *et al.*, 1991] 7

Figura 3. Representação esquemática da estrutura da spiropiperidina. *R1, R2 e R3* indicam as três posições, onde foram alternados diferentes substituintes para a preparação de cada uma das três bibliotecas[Bleicher *et al.*, 2002a] 9

Figura 4. Representação esquemática da estrutura da Sarcodictina isolada da *S. roseum*. Estão indicadas as características estruturais responsáveis pela atividade antitumoral identificadas após a preparação de uma biblioteca combinatória [Nicolaou *et al.*, 1998] ... 10

Figura 5. Grupos reativos frente à proteínas responsáveis por resultados falso-positivos em testes de *HTS*. [Rishton, 1997] 14

Figura 6. Exemplo de relação estrutura-reatividade, onde o composto mais potente é também o mais reativo frente a proteínas [Rishton, 1997] 15

Figura 7. Compostos visualizados por microscopia de transmissão eletrônica em tampão Tris (20mM) pH 7,2. (A-C) Tetraiodofenolftaleína (100 μM); (D) Vermelho Congo (50 μM), controle positivo de agregação; (E) Ácido 8-anilino-1-naftalenosulfônico (625 μM) controle negativo de agregação. Barra = 100 nm. [McGovern *et al.*, 2002] 17

Figura 8. Resultados do modelo de redes-neurais utilizado para distinguir fármacos (*ACD*) de não-fármacos (*WDI*). As linhas sólidas correspondem às séries de treinamento e as linhas pontilhadas correspondem à aplicação do modelo sobre os dois bancos de dados, respectivamente.[Sadowski & Kubinyi, 1998] 27

Figura 9. Representação esquemática da distribuição de compostos de diferentes classes terapêuticas, no universo de todos os compostos químicos. [Lipinski & Hopkins, 2004] ... 28

Figura 10. Gráfico das três componentes principais obtidas pelo modelo *ChemGPS* (*Chemical Global Positioning System*) mostrando a distribuição de compostos inibidores de quinase (esferas azuis), em relação a compostos que passaram pela "regra dos 5" (pontos amarelos) e outros compostos medicinais (pontos vermelhos).[Lloyd *et al.*, 2006] 30

Figura 11. Representação esquemática dos 6 diferentes modelos farmacofóricos possíveis da histamina, a partir da mudança da protonação do anel imidazólico e de uma rotação de 180° da cadeia lateral.[Kubinyi, 2005] 34

Figura 12. (A) Modelo farmacofórico de antagonistas de 5HT;[HIBERT *et al.*, 1988] (B) Modelo farmacofórico de antifúngicos, posicionado sobre a estrutura do 1-[(Aril)(4-aril-1*H*-pirrol-3-il)metil]-1*H*-imidazol.[Di Santo *et al.*, 2005] Grades em verde representam regiões de interação hidrofóbica; grades em vermelho representam posições para um anel aromático, grade em azul representa posição de um nitrogênio aromático, grades em preto representam regiões de exclusão de volume. 34

Figura 13. Modelo farmacofórico de inibidores da *Pfmrk*. Grades em verde representam regiões aceptoras de ligação de hidrogênio, grades em bege representam a posição de um anel aromático e a grade em azul representa uma região hidrofóbica alifática [Bhattacharjee *et al.*, 2004] 35

Figura 14. Representação esquemática dos mapas de contorno CoMFA, obtidos a partir de uma série de α-(N)-piridina carboxaldeído tiosemicarbazonas substituídas. **A**) Mapas de contorno eletrostático: O poliedro verde representa regiões onde cargas positivas aumentam a atividade, enquanto que o poliedro vermelho representa regiões onde cargas negativas são mais favoráveis para a atividade. **B**) Mapas de contorno estérico: poliedro verde representa regiões onde volume estérico aumenta a atividade biológica. enquanto que o poliedro vermelho representa regiões onde volume estérico diminui a atividade biológica. 36

Figura 15. (A) Representação esquemática da estrutura da HIV protease (1HHP) cristalizada sem o ligante em sua conformação "aberta"; (B) Representação esquemática da estrutura da HIV protease (1HVR) cristalizada com um ligante em sua conformação "fechada" 47

Figura 16. Representação esquemática da estrutura da Biotina e do HABA. Estão indicadas as interações com os resíduos do sítio de ligação da streptavidina[WEBER *et al.*, 1992] 53

Figura 17. Representação esquemática da estrutura de dois inibidores da trombina; e gráfico dos valores dos parâmetros termodinâmicos ΔG, ΔH e $-T\Delta S$ para a formação do complexo com a trombina-ligante, determinados por microcalorimetria[Klebe, 2006] 54

Figura 18. Representação esquemática das contribuições entrópicas e entálpicas que ocorrem durante o processo de formação do complexo [RL][ANDREWS *et al.*, 1984] Moléculas de água estão representadas por círculos azuis, o ligante está representado pelas esferas beges e o receptor está representado pela superfície verde. $\Delta H_{L\text{-}A}$; $\Delta H_{R\text{-}A}$; representam as entalpias de desolvatação do ligante e do receptor, respectivamente; $\Delta H_{L\text{-}R}$ representa a contribuição entálpica das interações ligante-receptor; $\Delta S_{A,}$ representa a variação da entropia resultante da desolvatação das superfícies hidrofóbicas; ΔS_{rot}, ΔS_{trans} e ΔS_{conf} correspondem a variação da entropia rotacional, translacional e conformacional, respectivamente 55

Figura 19. Gráfico da afinidade do complexo ligante-receptor ($-lnK_i$) em função do número (n) de ligações de hidrogênio entre o ligante e o receptor. Dados extraídos de 80 complexos [RL] depositados no *PDB*.[Bohm & Klebe, 1996] 56

Figura 20. Representação esquemática dos ligantes da termolisina. Estão indicadas as interações entre os ligantes e a proteína e os valores de K_i. 57

Figura 21. Representação esquemática dos inibidores da trombina, NAPAP e NAPAP-cíclico, respectivamente. Estão indicados os valores de K_i, determinados para cada um dos inibidores. [MACK *et al.*, 1995].....58

Figura 22. Gráfico do potencial de força média ($-\ln(g(R)/g_{i,j}(R))$) em função da distância (R[Å]) para três pares de átomos, respectivamente, O.3-O.3; O.3-O.co2 e O.co2-N.pl3 (nomenclatura para tipos de átomos utilizada no *SYBYL*[SYBYL, 2004]), derivados de 6026 complexos proteína-ligante depositados no *PDB*. $g_{i,j}(R)$ é a função de distribuição

radial dos átomos tipo *i*, e *j*; e g(r) é a função de distribuição radial de dois átomos de referência. [Gohlke *et al.*, 2000a] 63

Figura 23. Representação esquemática da estrutura dos compostos antichagásicos nifurtimox e benznidazole, respectivamente 65

Figura 24. Representação esquemática da ligação entre um peptídeo e uma enzima protease. Estão indicadas a terminologia usual para as cavidades da enzima e os correspondentes resíduos do substrato. A seta indica o ponto em que ocorre a hidrólise do peptídeo. 69

Figura 25. Representação esquemática do mecanismo de proteólise de cisteíno-proteases: (**A**) Forma ativa da enzima com o par iônico *SG⁻- Cys25/ImH⁺-His159*; (**B**) aproximação de um ligante; (**C**) formação do complexo não covalente com ligações de hidrogênio entre o oxigênio carbonílico do peptídeo e os hidrogênios da *Cys25* (cadeia principal) e da *Gln19* (cadeia lateral); (**D**) primeiro intermediário tetraédrico, formado pelo ataque nucleofílico do *SG-Cys25* ao carbono carbonílico do peptídeo, e estabilizado pelas ligações de hidrogênio; (**E**) esterificação da enzima, e saída da porção amina após catalise ácida do imidazol da *His159*; (**F**) ataque nucleofílico de uma molécula de água ao carbono acil da enzima; (**G**) Segundo intermediário tetraédrico; liberação do produto ácido e reconstituição da enzima livre (**A**). 70

Figura 26. Representação esquemática da estrutura do inibidor da cruzaína bis-2-hidroxi-benzaldeído-(carboxi)-hidrazona 72

Figura 27. Representação esquemática do mecanismo de ação proposto para a inibição da cruzaína por tiosemicarbazonas. [EAKIN *et al.*, 1992] 73

Figura 28. Estruturas dos 13 complexos da cruzaína alinhados átomo-a-átomo, em relação a estrutura do complexo 1AIM. Foram excluídos do alinhamento os inibidores e moléculas de água presentes nos arquivos *PDB*. Está indicado o resíduo catalítico *Cys25* no fundo da cavidade do sítio ativo da cruzaína. 98

Figura 29. (A) Representação gráfica da superfície de Van der Waals da cruzaína.; (B) Representação gráfica dos resíduos que compõem o sítio ativo da cruzaína. As regiões assinaladas *S3*, *S2*, *S1*, *S1'* correspondem às respectivas cavidades do sítio ativo 101

Figura 30. (A) Representação gráfica dos inibidores alinhados no sítio ativo da cruzaína; (B) Detalhe dos inibidores alinhados sobre a superfície da cruzaína, onde estão indicadas as ligações de hidrogênio entre os inibidores e os resíduos de *Gly66* e de *Asp158* no sítio ativo da enzima. 101

Figura 31. Representação gráfica dos contornos isopotenciais da cruzaína calculados pelo programa *GRID* [Goodford, 1985], usando-se diferentes sondas moleculares, sendo:. Branco - carbono alifático. nível de energia de –3,0 Kcal/mol); Vermelho: oxigênio carbonílico - nível de energia de -3,0 Kcal/mol; Azul: nitrogênio de amida - nível de energia de –6,0 Kcal/mol). A superfície em amarelo corresponde a superfície de Van der Waals da cruzaína (1ME4). As regiões assinaladas *S1*, *S2* e *S3* correspondem as respectivas cavidades do sítio ativo. 103

Figura 32. Representação dos 63 pseudocentros (PC) identificados pelo programa *CAVBASE*, alinhados sobre os resíduos que compõem a cavidade de ligação da cruzaína

(estrutura extraída do complexo *PDB 1ME4*). Os cubos em azul, em vermelho e, em cinza representam os PC de resíduos doadores de ligação de hidrogênio, aceptores de ligação de hidrogênio e hidrofóbicos alifáticos, respectivamente. As regiões assinaladas *S1*, *S2*, *S3* e *S1'* correspondem as respectivas cavidades do sítio ativo.....104

Figura 33. Representação dos 10 (dez) pseudocentros (PC) selecionados pelo programa *CAVBASE*, alinhados sobre os resíduos que compõem a cavidade de ligação da cruzaína (estrutura extraída do complexo *PDB 1ME4*). Os cubos em azul, em vermelho e em cinza representam as PC de resíduos doadores de ligação de hidrogênio, aceptores de ligação de hidrogênio e hidrofóbicos alifáticos, respectivamente. 105

Figura 35. Representação gráfica dos 20 inibidores presentes nos complexos selecionados pela busca de cavidades similares à cruzaína, utilizando o programa *CAVBASE*,[Schmitt *et al.*, 2001] alinhados no sítio ativo da cruzaína; estão indicadas as ligações de hidrogênio entre os inibidores e os resíduos de *Gly66* e de *Asp158* no sítio ativo da enzima.....110

Figura 36. Gráficos dos valores de *RMSD* em função do tempo, obtidos a partir das trajetórias[a] da cruzaína[b] sem o ligante (vermelho) e complexada com o ligante 1ME4 (azul), para todos os átomos da proteína (A); para os átomos do sítio ativo (B); para os átomos de carbono-alfa (C) e para os átomos de cadeias laterais (D), respectivamente. 111

Figura 37. Gráficos dos valores das distâncias em função do tempo, obtidos a partir das trajetórias[a] da cruzaína[b] livre (vermelho) e complexada (azul), respectivamente, entre os seguintes átomos: *CA-Asp60 – CA-Gly66* (A); *CA-Cys25 - CA-His159* (B); *CA-Met68 – CA-Val132* (C); *CA-Ser139 – CA-Asp158* (D); CD *Glu205 – CA-Cys25* (E); *CG-Asp158 – CA*-Gly65 (F); *CG*-Leu67 – *CA*-Gly65 (G); e entre os resíduos: Asp60[c] – Gly66[c] (H); *Cys25*[c] – *His159*[c] (I); *Met68*[c] - *Val132*[c] (J) e, *Ser139*[c] - *Asp158*[c] (K), respectivamente. ...113

Figura 38. Representação esquemática da seqüência filtros aplicados sobre a biblioteca de compostos *ZINC*,[Irwin & Shoichet, 2005] representando o **Modelo I** de *virtual screening*. Em parêntesis são indicados os números de compostos selecionados em cada etapa e ao lado a correspondente porcentagem de redução, por cada filtro.....116

Figura 39. (**A**) Representação do alinhamento dos ligantes[a] utilizados para a construção do modelo farmacofórico. (**B**) Representação do modelo farmacofórico, gerado com o programa *CATALYST*,[Catalyst, 2005] a partir dos ligantes[a] alinhados.....118

Figura 40. Sobreposição das poses de maior valor de *Fitness* obtidas pelo *GOLD 3.0* (branco) e as correspondentes poses encontradas no cristal (verde) para os ligantes dos complexos 1ME3 (**A**) e 1F2B (**B**), respectivamente. As regiões assinaladas *S3*, *S2*, *S1*, *S1'* correspondem às respectivas cavidades do sítio ativo.....120

Figura 41. Poses de maior valor de *Fitness* identificadas pelo programa *GOLD*, após o procedimento de *docking* dos compostos ZINC04506284 (**A**) e **ZINC03326243** (**B**), posicionados dentro da cavidade do sítio ativo da cruzaína. São indicados os valores das distâncias das ligações de hidrogênio entre os átomos dos ligantes e os resíduos Gly66 e Asp158 da cavidade do sítio ativo da cruzaína. As regiões assinaladas *S3*, *S2*, *S1*, *S1'* correspondem às respectivas cavidades do sítio ativo.....122

Figura 42. Gráficos dos valores da velocidade[a] de consumo do substrato[b] pela cruzaína[c] na presença de cada um dos compostos selecionados (usando pelo menos duas concentrações) e o efeito da adição de TritonX-100 (0,1%) nos testes onde foi observada

inibição da atividade enzimática (compostos **ZINC02663001**, **ZINC01936854** e **ZINC03326243**, respectivamente). 126

Figura 43. Gráficos dos valores[a] de consumo do substrato[b] pela cruzaína na presença dos compostos **ZINC02663001**, **ZINC01936854** e **ZINC03326243,** respectivamente. Os valores das concentrações da enzima foram aumentadas 10 vezes, respectivamente, de 0,18 nM (gráficos à esquerda) para 1,8 nM (gráficos à direita) para confirmar a ocorrência de inibição enzimática via mecanismo promíscuo. 128

Figura 44. Representação esquemática da seqüência filtros aplicados sobre a biblioteca de compostos *ZINC*,[Irwin & Shoichet, 2005] representando o **Modelo II** de *virtual screening*. Em parêntesis são indicados os números de compostos selecionados em cada etapa e ao lado a correspondente porcentagem de redução por cada filtro..... 132

Figura 45. Representação do modelo farmacofórico de 5 pontos gerado a partir das estruturas dos ligantes extraídos dos 30[(a)] complexos, selecionados para a criação e validação do modelo de *virtual screening* 133

Figura 46. Grupos funcionais eletrofílicos susceptíveis ao ataque do enxofre da *Cys25*. As setas apontam para o átomo eletrofílico, que sofre o ataque nucleofílico em cada um dos grupamentos apresentados: Aldeído; Cetona; α-Halo Metil Cetona; Epóxido; Semicarbazona; Nitrila; Azo Composto; Disulfeto; Vinil Sulfona; α-Amido Cetona; cetona αβ-insaturada..... 135

Figura 47. Representação do modelo farmacofórico de 5 pontos[(a)] com as correspondentes regiões de exclusão de volume, definidas a partir das coordenadas dos átomos dos resíduos que formam o sítio de ligação da cruzaína (estrutura extraída do complexo *PDB* 1ME4).137

Figura 48. Sobreposição da pose de maior valor de *Fitness* obtidas pelo *GOLD 3.0* (branco) e a correspondentes pose encontradas no cristal (verde) para os ligantes extraídos dos complexos 1ATK (**A**); 1AU3 (**B**), 1MS6 (**C**) e 2DCD (**D**), respectivamente., respectivamente. As regiões assinaladas *S3*, *S2*, *S1*, *S1'* correspondem às respectivas cavidades do sítio ativo..... 147

Figura 49. Poses de maior valor de *Fitness* identificadas pelo programa *GOLD*, após o procedimento de *docking* dos compostos ZINC02338249 (**A**) **ZINC02342142** (**B**), **ZINC02399860** (**C**) e **ZINC02400672** (**D**) posicionados dentro da cavidade do sítio ativo da cruzaína. São indicados os valores das distâncias das ligações de hidrogênio entre os átomos dos ligantes e os resíduos Gly66 e Asp158 da cavidade do sítio ativo da cruzaína. As regiões assinaladas *S3*, *S2*, *S1*, *S1'* correspondem às respectivas cavidades do sítio ativo. 148

Figura 50. Representação estrutural dos dois inibidores de cisteíno-protease identificados pelo modelo de *virtual screening* de inibidores da cruzaína, respectivamente, ZINC04899534 e ZINC01547017 149

Figura 51. Gráficos dos valores da velocidade[a] de consumo do substrato[b] pela cruzaína[c] sem inibidor (preto) e na presença de três diferentes concentrações do composto ***E-64***. ..153

Figura 52. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC00890529** (*Asinex BAS02985564*). 154

Figura 53. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC01806509** (*LifeChemicals F3047-0312*)..... 154

Figura 54. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC04964356** (*ChemBridge 5230379*).....155

Figura 55. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto ZINC03376060 (*Enamine T0519-6812*).....155

Figura 56. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC03243155** (*Enamine T0507-1260*).....156

Figura 57. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC01794422** (*LifeChemicals F3145-4606*)..... 156

Figura 58. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC01794384** (*LifeChemicals F3145-4620*)..... 157

Figura 59. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC01414485** (*Toslab 60063*).....157

Figura 60. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC02768688** (*Vitas-M STK008910*).....158

Figura 61. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC01794844** (*LifeChemicals F3145-3436*)..... 158

Figura 62. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC02411240** (*IBScreen, STOCK4S-44363*). ... .159

Figura 63. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC02400672** (*IBScreen, STOCK4S-33348*). ... .159

Figura 64. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC02198693** (*IBScreen, STOCK2S-31029*). ... .160

Figura 65. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC02209285** (*IBScreen, STOCK2S-31935*). ... .160

Figura 66. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC02188675** (*Specs, AG-205/37047002*)..... 161

Figura 67. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC02399860** (*IBScreen, STOCK4S-32328*). ... .161

Figura 68. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC01758885** (*IBScreen, STOCK3S-57280*). 162

Figura 69. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC00874200** (*IBScreen, STOCKS2S-49269*). 162

Figura 70. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC02342142** (*IBScreen, STOCK2S-99288*). 163

Figura 71. Representação estrutural do composto **ZINC01794422** (*LifeChemicals F3145-4606*) identificado como inibidor da cruzaína, através **Modelo II** de *virtual screening* aplicado ao banco de *ZINC* 164

Figura 72. Gráfico do ajuste[a] (—) dos parâmetros da equação de *tight-binding*[b] aos valores (■) de velocidade[c] de consumo do substrato[d] pela cruzaína[e] em função da concentração do composto **ZINC01794422** (*LifeChemicals F3145-4606*). 164

Figura 73. Pose de maior valor de *Fitness* identificadas pelo programa *GOLD*, após o procedimento de *docking* do composto **ZINC01794422** (*LifeChemicals F3145-4606*) posicionado dentro da cavidade do sítio ativo da cruzaína. São indicados os valores das distâncias das ligações de hidrogênio entre os átomos dos ligantes e os resíduos Gly66 e Asp158 da cavidade do sítio ativo da cruzaína. As regiões assinaladas *S3*, *S2*, *S1*, *S1'* correspondem às respectivas cavidades do sítio ativo. 165

LISTA DE ABREVIATURAS E SIGLAS

ACD	Available Chemical Directory
ADME	Absorção, Distribuição, Metabolismo e Excreção
C25G	Cruzaína mutante, com uma Glicina substituindo *Cys25*
CCDC	Cambridge Crystallographic Data Centre
ChemGPS	Chemical Global Positioning System
CMC	Comprehensive Medicinal Chemistry
CoMFA	Comparative Molecular Field Analysis
CoMSIA	Comparative Molecular Similarity Indices Analysis
CSD	Cambridge Structural Database
DLS	Dynamic Light Scattering
DM	Dinâmica Molecular
DMSO	Dimetilsulfóxido
DTT	Ditiotreitol
E-64	Trans-epoxisuccinil-l-leucilamido-4-guanidino-butano
EDTA	Ácido Etilenediaminotetracético
FEP	Free Energy Perturbation
HABA	Ácido 2,4-hidroxifenilazo benzóico
HIV	Human Immunodeficiency Virus
HOMO/LUMO	Highest Occupied Molecular Orbital/Lowest Unoccupied Molecular Orbital
HTS	High-Throughput Screening
IC_{50}	Concentração Inibitória de 50%
K_i	Constante de afinidade
logP	Logaritmo do Coeficiente de Partição
MC	Monte Carlo
MDDR	MACCS Drug Data Report
MQ	Mecânica Quântica
MIF	Molecular Interaction Fields
MM	Mecânica Molecular
MUP	Mouse Urine Protein
NK-1	Receptor de Neuroquinina
PC	Pseudocentros
PCA	Principal Components Analysis
PDB	Protein Data Bank
Pfmrk	*Plasmodium falciparum* quinase

PLS	Partial Least Squares
QSAR	Quantitative Structure Activity Relationship
RL	Complexo Receptor-Ligante
RMN	Ressonância Magnética Nuclear
RMSD	Root Mean Square Distance
S1'*, *S1*, *S2*, *S3	Cavidades do sítio ativo da cruzaína
SAR	Structure Activity Relationship
SNC	Sistema Nervoso Central
TK	Timidina Quinase
WDI	World Drug Index
Z-FR-MCA	Benzoil-fenilalanina-arginina-7-amido-4-metil-cumarina
ΔG	Variação da energia livre de Gibbs
ΔH	Variação da energia Entalpia
ΔS	Variação da energia Entropia

Nota para a nomenclatura de resíduos e átomos em proteínas.

A nomenclatura usada para a identificação de resíduos de aminoácidos em proteínas será a nomenclatura *IUPAC* de três letras,[[Anon], 1984] seguido da posição do resíduo na seqüência de aminoácido da estrutura primária da proteína. Por exemplo, o termo "*Cys25*", representa o resíduo de aminoácido cisteína na posição 25.

A nomenclatura usada para a identificação de átomos em resíduos de aminoácidos será a derivada da nomenclatura *IUPAC*,[[Anon], 1984] em que o átomo de carbono entre o carbono carbonílico e o nitrogênio da amina é denominado *CA* (ou carbono-α). O próximo átomo ligado ao *CA* será denominado *CB* (ou carbono-β), o seguinte será denominado *CG* (ou carbono-γ) e assim por diante, seguindo o alfabeto grego. Após a indicação do átomo será indicado o resíduo a que este átomo pertence, separado por um hífen. Por exemplo, o termo "*CA-His159*" representa o átomo de carbono-α do resíduo de aminoácido histidina 159. Quando o átomo em questão não for um carbono, a nomenclatura se ajusta para indicar o tipo do átomo e a posição em relação ao *CA*. Assim, por exemplo, o termo "*NE-Arg115*" representa o átomo de nitrogênio-ε do resíduo de aminoácido arginina 115, ou o termo "*SD-Met96*" representa o átomo de enxofre-δ do resíduo de aminoácido metionina 96.

RESUMO

Malvezzi, A. **Modelos de *Virtual screening* de Inibidores da Cruzaína: Desenvolvimento e Validação Experimental**. 2008.
100p. Tese (Doutorado) - Programa de Pós-Graduação em Química Orgânica. Instituto de Química, Universidade de São Paulo, São Paulo.

Com o objetivo de buscar e identificar novo(s) inibidor(es) da cruzaína – uma cisteíno-protease do *Trypanosoma cruzi*, o agente etiológico da doença de Chagas – foram propostos, validados e, a seguir, aplicados sobre a biblioteca de compostos *ZINC* (3.294.714 compostos), dois modelos de *virtual screening* (**Modelos I** e **II**).

Os modelos de *virtual screening* propostos, contendo seqüências de filtros físico-químicos, farmacofóricos, de *docking* e de seleção por inspeção visual, foram construídos a partir de informações de 13 complexos da cruzaína e de 20 complexos de outras cisteíno-protease, cujas estruturas estão disponíveis no *PDB*.

Numa primeira etapa, o reconhecimento detalhado das características estruturais da cruzaína foi realizado por inspeção visual; pelos campos de interação molecular, gerados pelo programa *GRID*; pela identificação das propriedades de interação molecular na superfície da cavidade, geradas pelo programa *CAVBASE* e; por simulações de dinâmica molecular.

O **Modelo I** de *virtual screening* – obtido a partir do reconhecimento das estruturas dos 13 complexos da cruzaína depositadas no *PDB* – foi aplicado sobre o *ZINC*, selecionando 10 compostos, dos quais 6 compostos foram adquiridos e submetidos ao teste de inibição enzimática da cruzaína, para a validação experimental do modelo. Observou-se que 3 destes compostos (**ZINC02470662**, **ZINC02682879** e **ZINC03192044**, respectivamente) não mostraram inibição significativa da cruzaína, nas condições experimentais utilizadas, até a concentração de 7 mM, enquanto que os 3 restantes (**ZINC02663001**, **ZINC01936854** e **ZINC03326243**, respectivamente) apresentaram inibição enzimática inespecífica, sugerindo que estes últimos agem pelo mecanismo promíscuo. O mecanismo promíscuo de inibição enzimática, foi verificado pela adição de 0,1% *Triton X-100* no ensaio enzimático, observando-se a correspondente perda de inibição

da cruzaína. Para estes compostos, a confirmação do mecanismo promíscuo foi feita observando-se a perda de inibição da enzima, após o aumento em dez vezes da concentração da cruzaína no ensaio enzimático.

O **Modelo II** – obtido a partir do reconhecimento das estruturas dos 13 complexos da cruzaína e dos 20 complexos de outras cisteíno-proteases, identificadas na busca por cavidades similares à cruzaína – foi aplicado sobre o banco de dados *ZINC* selecionando 55 compostos dos quais 19 foram adquiridos e submetidos ao teste de inibição enzimática da cruzaína, para validação experimental do modelo. Observou-se que o composto **ZINC01794422** apresentou inibição específica da enzima com constante de inibição no valor de K_i = 21 μM, enquanto que os demais 18 compostos não mostraram inibição significativa, nas condições experimentais utilizadas, até a concentração de 592 μM. O mecanismo promíscuo de inibição enzimática não foi observado, uma vez que todos os testes foram realizados com 0,1% de *Triton X-100*. O **Modelo II** identificou, ainda, mais dois inibidores da cruzaína (**ZINC04899534** e **ZINC01547017**) que, por serem estruturalmente semelhantes aos utilizados na construção do modelo e já terem sido descritos na literatura, não foram adquiridos ou testados nos ensaios enzimáticos. Considerando apenas o novo inibidor identificado, o **Modelo II** apresentou uma taxa de acerto de 5,3%. Este valor esta de acordo com as taxas de acerto encontradas na literatura que variam entre 1 a 50% .

Palavras-chave: Seleção virtual, cruzaína, *docking*; modelo farmacofórico; fármaco-similar, mecanismo promíscuo.

ABSTRACT

Malvezzi, A. **Virtual screening Models of Cruzain Inhibitors: Development and Experimental Validation**. 2008

100p. PhD Thesis – Graduate Program in Chemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

In order to search and identify new cruzain inhibitor(s) – a cysteine-protease of *Trypanosoma cruzi*, the etiologic agent of Chagas disease – two virtual screening schemes (**Models I** and **II**) were proposed, validated and applied to the *ZINC* database (3.294.714 compounds).

The proposed virtual screening models, bearing a sequence of different physical-chemical, pharmacophore and docking filters, as well as a visual inspection filter, were built from information taken from 13 cruzain complexes and from 20 complexes of other cysteine proteases, having their structures available in PDB.

In a first step, a detailed recognition of the cruzain structural features and characteristics was performed through visual inspection of the enzyme environment; followed by the analysis of GRID generated molecular interaction fields; through the identification of molecular interaction properties exposed at the enzyme cavity surface, generated by the *CAVBASE* program; and by molecular dynamics simulations.

The virtual screening **Model I,** – generated from the structural characteristics recognized from 13 *PDB* cruzain complexes – when applied to the *ZINC* database selected 10 compounds. For the experimental validation of the model, six of these compounds have been acquired and were tested as cruzain inhibitors. It was observed that three of the tested compounds (**ZINC02470662**, **ZINC02682879** and **ZINC03192044,** respectively) did not show any significant cruzain inhibition, up to 7 mM. Meanwhile the other three tested compounds (**ZINC02663001**, **ZINC01936854** and **ZINC03326243**, respectively) showed an unspecific cruzain inhibition, suggesting that an enzyme inhibition by promiscuous mechanism occurred. This mechanism was verified by the addition of 0.1% *Triton X-100* on the enzymatic assay with a concomitant loss of cruzain inhibition activity. For these compounds, the confirmation of the promiscuous mechanism was also done, observing the

loss of enzyme inhibition, after a ten times increase in the cruzain concentration on the enzymatic assay.

The virtual screening **Model II** – generated from the structural characteristics recognized from 13 cruzain complexes and 20 complexes of other cysteine proteases, that have been identified on a search for cavities similar to cruzain – selected 55 compounds, when applied to the *ZINC* database. In order to experimentally validate the model, nineteen compounds have been acquired and were tested as cruzain inhibitors. It has been observed that one compound, **ZINC01794422,** showed a specific cruzain inhibition (K_i = 21 μM), while the other eighteen showed no significant inhibition, up to 592 μM concentration. The promiscuous mechanism of enzymatic inhibition was not observed, since 0.1% of *Triton X-*100 was added in all assays. Additionally, **Model II** identified two other cruzain inhibitors (**ZINC04899534** and **ZINC01547017**). However, these compounds have not been acquired or tested, since they are known cruzain inhibitors - already described in the literature – and are structurally similar to the inhibitors used in the construction of the model. Referring to new inhibitors found, **Model II** showed a hit rate of 5,3%. This value is in agreement with those found in the literature, which ranges from 1 to 50%.

Keywords: Virtual screening, cruzain, docking; pharmacophore model; drug-like, promiscuous mechanism.

OBJETIVOS

O objetivo primário deste trabalho, destinado à tese de doutorado, é a busca e a identificação de novo(s) inibidor(es) da cruzaína - uma cisteíno-protease do *Trypanosoma cruzi*, o agente etiológico da doença de Chagas - através da aplicação de modelos de *virtual screening* sobre bibliotecas de compostos orgânicos de baixa massa molecular.

Para a seleção de compostos com determinadas propriedades ou características estruturais, típicas de inibidores da cruzaína, em banco de dados virtuais de compostos, pretendeu-se desenvolver, validar e aplicar modelos de *virtual screening* contendo uma seqüência de filtros virtuais de diferentes tipos: (*i*) filtros físico-químicos, que se baseiam em propriedades uni- e bi-dimensionais de compostos fármaco-similares; (*ii*) filtros farmacofóricos, que se baseiam na distribuição espacial das características responsáveis pela interação ligante-receptor; (*iii*) filtros de *docking*, que se baseiam em medidas da afinidade entre o composto e a estrutura do alvo biológico e; (*iv*) filtros de seleção por inspeção visual, que se baseiam no conhecimento e intuição química do pesquisador sobre todos os aspectos envolvidos no desenvolvimento de moléculas bioativas.

Para a obtenção e a validação dos modelos de *virtual screening* pretendeu-se utilizar as informações disponíveis nos treze complexos da cruzaína cujas estruturas foram resolvidas por cristalografia de raios-X e estão disponíveis no banco de dados de proteínas *PDB*, considerando informações tanto da estrutura da cruzaína cristalizada (*structure based drug design*), como da estrutura dos inibidores complexados com a enzima (*ligand based drug design*).

Para o reconhecimento detalhado das características estruturais da cruzaína e de seu sítio de ligação, consideraram-se nesta análise aspectos como: a resolução cristalográfica do complexo, o reconhecimento da função da enzima e de seu mecanismo catalítico; a identificação da cavidade de ligação e de suas propriedades de interação molecular; a conformação e a flexibilidade da proteína; e a presença de moléculas de água conservadas em sua estrutura. Concomitantemente, pretendeu-se considerar as estruturas dos inibidores complexados com a cruzaína, para que sejam identificadas as características compartilhadas por todos estes inibidores e que, ao mesmo tempo, contribuíssem para a interação com a enzima.

Para a validação experimental do modelo de *virtual screening* proposto, pretendeu-adquirir os compostos sugeridos pelo modelo para que estes sejam, a seguir, submetidos ao teste de inibição enzimática *in vitro* da cruzaína.

Finalmente, pretendeu-se estudar a possibilidade da ocorrência do mecanismo promíscuo nos testes de inibição enzimática, para que sejam reduzidas as probabilidades da ocorrência de resultados falso-positivos nos testes experimentais *in vitro*.

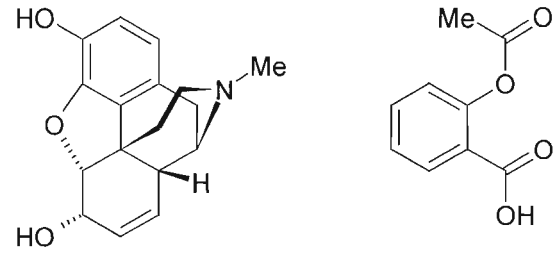
1. Introdução

1.1. A Descoberta de Fármacos

Da Pré História a Era do Computador

A descoberta de fármacos é um processo em contínua evolução que se modifica e agrega novas estratégias à medida que novos conhecimentos se acumulam. Não existe um início claro de como os primeiros fármacos foram descobertos, mas sabemos que em todas as sociedades humanas primitivas, substâncias naturais já eram usadas para os mais diversos fins terapêuticos. Este conhecimento milenar está incorporado naquilo que hoje conhecemos como a medicina tradicional, em parte, utilizada até os dias de hoje.[Viegas, 2006]

Até um pouco mais de cem anos atrás, todos os medicamentos utilizados pelo homem eram exclusivamente de origem natural (vegetal, animal e/ou mineral).[Kubinyi, 2005] A descoberta de novos medicamentos era feita, em geral, testando-se aleatoriamente diferentes produtos naturais contra as mais diversas doenças, e observando seus efeitos farmacológicos. Com o avanço das ciências naturais, em especial a química e a biologia, o homem passou a estudar sistematicamente as plantas de uso medicinal. Inicialmente, os princípios ativos naturais eram simplesmente isolados e caracterizados, mas logo em seguida estas substâncias isoladas passaram a ser modificadas em laboratório, com o objetivo de melhorar seus efeitos terapêuticos e/ou diminuir os seus efeitos colaterais. Como exemplos, cita-se a morfina (**Figura 1**), que foi o primeiro principio-ativo natural isolado em 1806, e o ácido acetilsalicilico (Aspirina) (**Figura 1**), que foi o primeiro medicamento semi-sintético produzido em laboratório, em 1897. A partir deste momento, o homem deixou de depender exclusivamente da natureza como única fonte de moléculas bioativas e passou a propor moléculas estruturalmente inéditas.



Morfina Ácido acetilsalicílico

Figura 1. Representação esquemática das estruturas da Morfina e do Ácido acetilsalicílico (Aspirina), respectivamente, o primeiro principio-ativo natural isolado e o primeiro medicamento semi-sintético produzido em laboratório

A primeira estratégia utilizada para a descoberta racional de novos fármacos foi a modificação estrutural de protótipos naturais e/ou de compostos bioativos. Até hoje esta estratégia é utilizada e é denominada *ligand based drug design*, ou seja, desenvolvimento de fármacos com base no ligante. Nesta tese de doutorado, o termo "ligante" será usado indiscriminadamente para se referir a qualquer molécula (agonista, antagonista, substrato ou inibidor) que se liga à uma macromolécula biológica.

Durante a primeira metade do século XX, o fator limitante para a descoberta de novos fármacos foi a limitada capacidade sintética.[Kubinyi, 2005] Por exemplo, a penicilina que havia sido descoberta por Fleming, em 1929, teve seus primeiros derivados puramente sintéticos produzidos apenas em 1959. Entretanto, após a segunda grande guerra, com o avanço das técnicas de síntese orgânica e com a adoção de padrões éticos mais rígidos para os testes de atividade em humanos, o fator limitante para a descoberta de fármacos passou a ser a falta de modelos experimentais para a validação dos novos candidatos a fármacos.[Kubinyi, 2005] Até então, os testes de atividade eram feitos diretamente em seres humanos ou em modelos animais vivos. Com o aumento da oferta de novas moléculas candidatas, estes modelos *in vivo* de atividade biológica se tornaram caros, lentos e inespecíficos, impossibilitando o teste das moléculas na mesma velocidade com que estas eram sintetizadas.

Com a expansão do conhecimento sobre as bases moleculares da fisiologia humana, a partir da década de 70, surgiram os primeiros testes de atividade *in vitro*. Estes modelos bioquímicos de atividade eram mais baratos mais rápidos e mais específicos que os modelos *in vivo* utilizados anteriormente. O aumento da velocidade com que os testes de atividade eram realizados, principalmente após o surgimento dos testes ultra-rápidos em larga escala (*High Throughput Screening - HTS*) nos anos 90, tornaram novamente a capacidade de síntese orgânica e a velocidade de isolamento de novos princípios ativos os fatores limitantes para a descoberta de novos fármacos.[Kubinyi, 2005]

Em resposta a esta demanda gerada pelos testes de *HTS*, surgiu a química combinatória, que propunha a síntese de centenas de milhares de novos compostos, utilizando-se técnicas de automação e de reações multi-componentes.[Kubinyi, 1998a] Em seus primeiros anos, a química combinatória criou grandes bibliotecas de compostos especificamente para serem usados nos testes de *HTS* existentes. Entretanto, ao contrário do

que se havia imaginado, esta estratégia não acarretou um aumento do número de novos compostos ativos descobertos. O que se observou, no entanto, foi o inverso e ainda mais associado a um grande desperdício de recursos, principalmente devido ao grande número de compostos com atividades falso-positivas.[Wess *et al.*, 2001]

Uma revisão crítica do paradigma *HTS*/química combinatória, concluiu que a síntese e o teste de compostos orgânicos de modo aleatório não tornam maiores as probabilidades de se encontrar um fármaco, pois estes devem apresentar propriedades físico-químicas e estruturais específicas. A partir destas observações, surgiu o conceito de compostos *drug-like*, ou fármaco-similares, para representar os compostos com propriedades físico-químicas/estruturais características de fármacos. Estas propriedades *drug-like* passaram a ser usadas como filtros para a pré-seleção e o enriquecimento de bibliotecas de compostos, antes destes serem submetidos aos testes de atividade, ou seja, para eliminar compostos que não têm características de fármacos.[Shoichet, 2004]

Mais recentemente, graças à disponibilidade da estrutura molecular dos alvos biológicos, surgiu uma estratégia metodológica complementar à adotada até então para a descoberta racional de novos fármacos. Esta estratégia é denominada *structure based drug design*, ou seja, desenvolvimento de fármacos com base na estrutura do alvo biológico.[Klebe, 2006] Nesta tese de doutorado, os termos "alvo" ou "receptor" serão usados indiscriminadamente para descrever uma macromolécula biológica na qual o fármaco ou, o ligante, deve se ligar. Normalmente esta macromolécula é uma proteína, mas também pode ser RNA ou DNA. A partir do conhecimento da estrutura do receptor, tornou-se possível o desenvolvimento de ligantes específicos e com alta afinidade para um determinado alvo. Nos dias de hoje, este desenvolvimento é feito em conjunto com a otimização de outras propriedades igualmente importantes, como a solubilidade, absorção, distribuição, metabolismo, excreção e toxicidade.[Oprea, 2002]

Finalmente, apesar dos avanços tecnológicos, metodológicos e conceituais citados anteriormente, o descobrimento de novos medicamentos ainda é um desafio para qualquer pesquisador envolvido na área de química medicinal. Modelos cada vez mais complexos se tornam necessários a cada novo conhecimento que se agrega e assim como não existe um início claro de como os fármacos eram descobertos, também se pode definir um fim para este processo em constante evolução.

A seguir, os principais temas abordados nesta breve introdução histórica do desenvolvimento de fármacos serão aprofundados para dar suporte conceitual às técnicas e às abordagens metodológicas utilizadas neste trabalho.

1.2. *HTS* e Química Combinatória

A força bruta aplicada ao descobrimento de fármacos

No final do século passado, o desenvolvimento metodológico da biologia molecular e da síntese orgânica, associado ao desenvolvimento tecnológico nas áreas de computação e de robótica, possibilitou o surgimento do *HTS* e da química combinatória e a conseqüente aplicação destas estratégias no descobrimento de novos fármacos.[Sittampalam *et al.*, 1997] Em seus primeiros anos, a estratégia metodológica *HTS*/química combinatória representou uma revolução na velocidade com que químicos eram capazes de sintetizar compostos e conduzir testes bioquímicos e farmacológicos. Entretanto, mais do que uma revolução tecnológica, a introdução destas duas estratégias representou uma verdadeira panacéia entre pesquisadores envolvidos no desenvolvimento de fármacos.

> *"Considerando a grande diversidade estrutural acessível e a velocidade dos testes, é provável que a descoberta de novos compostos líder vai progressivamente deixar de ser um fator limitante no processo de desenvolvimento de fármacos."* [Williard *et al.*, 1996]

> *"O tema motivador da química combinatória deve ser a possibilidade de se fazer um milhão de variações de qualquer padrão estrutural e testar. Mais barato, mais rápido e melhor. Quando nós formos capazes de fazer isso em uma semana, então possivelmente os desafios intelectuais estarão acabados."* [Borman, 2006]

Estas duas citações exemplificam as expectativas e a euforia que acompanharam o surgimento do *HTS* e da química combinatória quando projetadas às possibilidades de aplicação destas ao desenvolvimento de fármacos. Ao mesmo tempo, associados aos testes de *HTS* foram anunciados os programas de seqüenciamento do genoma humano e de outras espécies e com estes a promessa de novos alvos para o desenvolvimento de terapias contra diferentes doenças.[Kubinyi, 2003; Drews, 2000]. Em relação à química combinatória, um verdadeiro "jogo de números" expandia exponencialmente o número compostos

potencialmente sintetizáveis. Estimativas do número de compostos orgânicos possíveis de serem sintetizados, com massa molecular de até 500 Daltons, atingem valores de 10^{60}, 10^{100}, e até 10^{200}![Lahana, 1999] Em sua essência, estas estratégias representavam a aplicação da força bruta na busca de estruturas líder (*lead structures*). Estruturas líder são compostos que apresentam atividade biológica e que são potencialmente capazes de evoluírem para novos fármacos.

Os testes de *HTS* são realizados em sistemas bioquímicos isolados, adaptados para a automação das etapas realizadas anteriormente por um técnico/cientista. Durante o processo de adaptação para a automação, as escalas são reduzidas para volumes mili/micro/nano. Para evitar problemas de solubilidade, os compostos são normalmente dissolvidos em solventes orgânicos como DMSO.[Dunn & Feygin, 2000] A velocidade com que estes testes são realizados pode atingir 100.000 compostos por dia.[Sittampalam *et al.*, 1997] Os testes de *HTS* representam atualmente a tecnologia mais utilizada pela indústria farmacêutica para a busca de estruturas líder em programas de desenvolvimento de fármacos.[von Ahsen *et al.*, 2008; Bleicher *et al.*, 2003] Entretanto, apesar de amplamente utilizados, os testes de *HTS* ainda são muito criticados, principalmente por sua crescente especialização. De acordo com Horrobin,[Horrobin, 2003] os testes *in vitro* utilizados em *HTS* refletem pouco a realidade, enquanto que para Rishton,[Rishton, 2003] os testes bioquímicos são distantes dos sistemas biológicos e o autor os define como testes de interação molecular de alta sensibilidade, vulneráveis a artefatos experimentais. Alguns destes artefatos, bem como suas implicações na descoberta de fármacos e alternativas existentes, serão apresentados mais a diante (item **1.3.**).

A química combinatória envolve a reação de n reagentes do tipo A, com m reagentes do tipo B, resultando na síntese de (n x m) produtos do tipo C. Em reações que podem ocorrer entre um único tipo de reagente e em múltiplas etapas, como a condensação de aminoácidos para a formação de um polipeptídio, este procedimento resulta em um crescimento exponencial de produtos, nos quais o valor do expoente é o número de etapas executadas e a base é o número de diferentes reagentes em cada etapa. Assim, partindo-se dos 20 diferentes tipos de aminoácidos naturais, é possível em apenas 6 etapas de reação, produzir 64 milhões (ou seja, 20^6) de hexapeptídio diferentes. Reações desta natureza produzem misturas de produtos, que podem ser testados diretamente na forma impura.

Quando a mistura se mostra biologicamente ativa, seus componentes individuais são identificados e separados utilizando técnicas de deconvolução.[Dias & Correa, 2001]

A química combinatória encontrou na síntese em fase sólida as condições ideais para seu desenvolvimento. De acordo com o princípio geral desenvolvido por Merrifield,[Merrifield, 1963] um reagente ***A*** é covalentemente ligado a um suporte polimérico insolúvel. Um segundo reagente ***B*** é adicionado na fase solúvel e este reage com ***A*** para formar o produto ***A-B,*** covalentemente ligado ao suporte insolúvel. O excesso de ***B,*** que não reagiu, e o solvente, são facilmente separados por uma simples filtração. Ao final da reação, o produto é liberado do suporte sólido. Devido a facilidade de purificação dos produtos, a síntese em fase sólida permite o uso de um excesso de reagentes para deslocar o equilíbrio da reação na direção dos produtos, obtendo-se assim, altos rendimentos e produtos com um alto grau de pureza. Entretanto, a principal vantagem da síntese em fase sólida, quando aplicada à química combinatória, é a possibilidade de automação e de repetição do ciclo quantas vezes forem necessárias para a obtenção do produto final desejado.

Um exemplo de aplicação da síntese em fase sólida associada à química combinatória é o procedimento de *Mix & Split* desenvolvido por Furka.[Furka *et al.*, 1991] Neste procedimento, a resina de suporte insolúvel é dividida em ***n*** porções iguais e cada uma destas reage com um único reagente. O número de porções depende do número de reagentes disponíveis. Finalizada a reação e a subseqüente purificação para a remoção do excesso dos reagentes, as porções individuais de resina são recombinadas, misturadas e novamente separadas em porções iguais. Cada uma das porções reage novamente com os mesmos reagentes. O processo pode ser repetido quantas vezes for necessário. O número de compostos obtidos ao final do processo cresce exponencialmente com o número de repetições do ciclo. A **Figura 2** mostra um exemplo de aplicação do procedimento de *Mix & Split* partindo-se de três reagentes (A, ■ e C) e com três ciclos de repetição. Ao final dos três ciclos são obtidos 27 compostos diferentes.

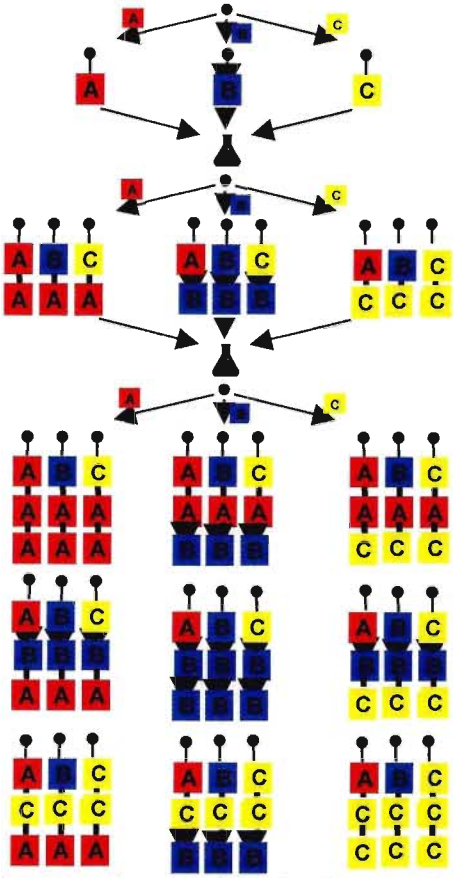


Figura 2. Exemplo ilustrativo de aplicação do procedimento de *Mix & Split* partindo-se de três reagentes (A, ■ e C) e com três ciclos de repetição [Furka *et al.*, 1991]

A química combinatória, nos seus primeiros anos, teve como foco principal a produção de bibliotecas com centenas de milhares de compostos. A ênfase na quantidade superava as preocupações com a qualidade e pureza dos compostos, e em função disso foram produzidas, inicialmente, bibliotecas de polímeros e de misturas de compostos orgânicos de alta massa molecular e de baixa diversidade estrutural.[Ortholand & Ganesan, 2004] Os resultados dos testes de *HTS* com estes compostos foram frustrantes.[Wess *et al.*, 2001] Foram verificadas dificuldades na identificação dos compostos ativos presentes nas misturas e, em geral, estes compostos ativos eram insolúveis em água e com propriedades de ADME incompatíveis com as de um fármaco.[Koppitz & Eis, 2006] Assim, após os primeiros anos de euforia e de resultados frustrantes, o foco da química combinatória se modificou para a produção de bibliotecas menores, com compostos orgânicos de baixa massa molecular e com maior diversidade estrutural.[Kubinyi, 1998a] Adicionalmente, a química combinatória passou a dar mais ênfase para a produção de compostos com alto

grau de pureza através da síntese paralela, em que as etapas sintéticas são realizadas simultaneamente para grupos de reagentes, mas cada produto é preparado em um compartimento de reação individual. A síntese paralela pode ser igualmente automatizada valendo-se das vantagens já mencionadas da síntese em fase sólida.

Atualmente, apesar dos contínuos avanços em direção à maior acessibilidade sintética,[Williard *et al.*, 1996] a química combinatória tem sido aplicada com melhores resultados na produção de bibliotecas focadas em alvos específicos ou em compostos derivados de uma estrutura líder.[Koppitz & Eis, 2006; Kubinyi, 2004]. A idéia inicial de utilizar a química combinatória associada ao *HTS* para a procura randômica de novos compostos bioativos se mostrou uma estratégia equivocada. Segundo Roger Lahana, editor da revista *Drug Discovery Today*:

> "Quando se procura uma agulha no palheiro, a melhor estratégia não é aumentar o tamanho do palheiro" [Lahana, 1999]

Bibliotecas focadas em alvos são coleções com até algumas centenas compostos, desenvolvidas por química combinatória, com o objetivo de encontrar ligantes para um determinado alvo específico.[Koppitz & Eis, 2006; Kubinyi, 2004] Estas são formadas por compostos orgânicos de baixa massa molecular derivados de um composto com atividade conhecida. Uma biblioteca focada deve promover uma diversidade de substituições na estrutura fundamental para explorar ao máximo o espaço químico ao redor desta e, assim, aumentar as chances de se encontrar derivados mais ativos.[Kubinyi, 2004] A preparação de diversas bibliotecas focadas pequenas produz maior diversidade estrutural que a produção de uma única biblioteca grande.[Kubinyi, 2004] No exemplo apresentado a seguir, três bibliotecas focadas foram preparadas por química combinatória com o objetivo de encontrar ligantes para o receptor NK-1 de neuroquinina.[Bleicher *et al.*, 2002b] O NK-1 é um receptor acoplado à proteína-G expresso tanto no SNC quanto em sistemas periféricos. A preparação das três bibliotecas foi feita por síntese em fase sólida e em fase líquida com base na espiropiperidina, uma estrutura privilegiada que é conhecida por se ligar a vários receptores análogos ao NK-1. Adicionalmente, considerando sua estrutura tridimensional pouco flexível, sua baixa massa molecular e, a possibilidade de modificações estruturais em três posições substituintes, a Spiropiperidina apresenta uma arquitetura ideal para o planejamento de fármacos empregando a química combinatória.

Como resultado desta busca foram, identificados novos ligantes de receptores NK-1 com afinidades nanomolares (nM).[Bleicher *et al.*, 2002a] A **Figura 3** apresenta a estrutura da espiropiperidina utilizada na preparação de cada uma das três bibliotecas.

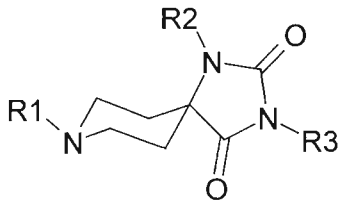


Figura 3. Representação esquemática da estrutura da espiropiperidina. *R1, R2 e R3* indicam as três posições, onde foram alternados diferentes substituintes para a preparação de cada uma das três bibliotecas[Bleicher *et al.*, 2002a]

Uma outra aplicação da química combinatória é a preparação de séries de análogos de um composto ativo para a obtenção de relações entre a estrutura química e a atividade biológica (*SARs – Structure Activity Relationships* ou *QSARs – Quantitative Structure Activity Relationships*). Em função da rapidez com que um grande número de compostos pode ser preparado por química combinatória, esta é uma estratégia ideal para a preparação de séries de análogos com uma grande faixa de variação de atividade, necessários para a elucidação do mecanismo de ação de um composto bioativo, ou para a compreensão das interações envolvidas na interação composto-alvo. Adicionalmente, a preparação de análogos de um composto ativo por química combinatória pode contribuir para a otimização das propriedades farmacocinéticas (ADME) de um composto candidato a fármaco, ou ainda auxiliar na proteção da patente do futuro produto comercial.[Kubinyi, 2004]

No exemplo a seguir, um produto natural foi utilizado como estrutura fundamental para a preparação, por química combinatória, de uma biblioteca de análogos estruturais da sarcodictina.[Dias & Correa, 2001] A sarcodictina é um produto isolado de corais da espécie *Sarcodictyion roseum* e possui potente ação antitumoral. A preparação da uma coleção combinatória com 67 análogos da sarcodictina empregou síntese em fase sólida e em fase líquida. Como resultado da busca, foi possível o estabelecimento de relações estrutura-atividade e, ainda, a identificação de alguns compostos mais potentes que os produtos naturais.[Nicolaou *et al.*, 1998] A **Figura 4** apresenta a estrutura da sarcodictina e

as características estruturais responsáveis pela atividade antitumoral, identificadas após a preparação de uma biblioteca combinatória.

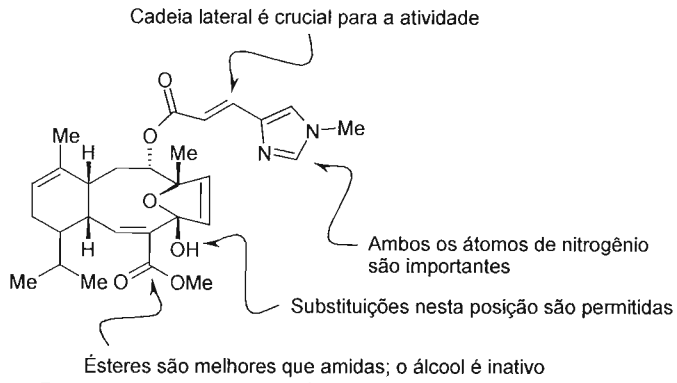


Figura 4. Representação esquemática da estrutura da sarcodictina isolada da *S. roseum*. Estão indicadas as características estruturais responsáveis pela atividade antitumoral identificadas após a preparação de uma biblioteca combinatória [Nicolaou *et al.*, 1998]

Apesar de existirem na literatura muitos outros exemplos de aplicação da química combinatória na busca de ligantes e de estruturas líder candidatas a fármacos, ainda são incipientes os exemplos de compostos que efetivamente chegaram no mercado e que foram identificados a partir de um programa de química combinatória.[Kubinyi, 2006] Entretanto, os resultados da química combinatória não podem ser avaliados somente pelo número de novos fármacos produzidos, mas por todas as contribuições que a sua implementação promoveu para a química medicinal. Adicionalmente, deve-se notar que a química combinatória é uma estratégia relativamente recente e que o processo entre a descoberta de uma estrutura líder e o lançamento do medicamento no mercado pode durar mais de uma década.

Do ponto de vista tecnológico, a química combinatória pode ser considerada um sucesso. Nos dias de hoje, todas as grandes empresas farmacêuticas têm programas de síntese em paralelo automatizada, aplicados às mais diversas estratégias baseando-se no princípio da química combinatória.[Borman, 2006; Guillier *et al.*, 2000] A produtividade das indústrias farmacêuticas, em termos do número de compostos preparados em relação aos custos por composto, apresentou significativos saltos de qualidade com o advento da química combinatória.[Borman, 2006; Guillier *et al.*, 2000]

Outra grande contribuição da química combinatória para a química aplicada ao descobrimento de novos fármacos foi a introdução dos conceitos de similaridade/diversidade química, aplicados à grandes bibliotecas de compostos.[Wess *et al.*, 2001] Considerações sobre a diversidade/similaridade de um conjunto de compostos passaram a ser usadas para avaliar a qualidade de bancos de dados de moléculas, gerados por química combinatória ou de coleções "*in house*" de compostos.[Kubinyi, 1998b] Por exemplo, bibliotecas randômicas preparadas para encontrar compostos ativos deveriam apresentar grande diversidade estrutural, enquanto que bibliotecas focadas deveriam ser similares ao composto líder e, ao mesmo tempo, apresentar diversidade suficiente para que a relação estrutura-atividade fosse exaustivamente explorada.[Kubinyi, 2004] Diversas ferramentas para quantificar a diversidade ou a similaridade entre compostos foram propostas e aplicadas aos mais diferentes sistemas químicos e biológicos.[Oprea & Gottfries, 2001; Martin *et al.*, 2001] Com o aumento do espectro de aplicações dos conceitos similaridade/dissimilaridade ficou claro que o grau de diversidade/similaridade apresentado por conjunto de compostos poderia ser diferente, dependendo do alvo estudado,[Martin *et al.*, 2001] ou seja, um conjunto de compostos com estruturas e propriedades diversas para um determinado alvo biológico, pode não ser igualmente diverso perante um outro alvo.[Ajay *et al.*, 1998; Kubinyi, 1998b] Ainda mais, tornou-se consenso que uma biblioteca de compostos para ser usada no desenvolvimento de fármacos deve apresentar não somente uma diversidade estrutural, mas também estruturas biologicamente relevantes e com propriedades específicas típicas de compostos bioativos.[Wess *et al.*, 2001; Martin *et al.*, 2001; Ajay *et al.*, 1998]. Assim, a partir da necessidade de se criar bibliotecas de compostos com propriedades de fármacos, surgiu o conceito de compostos *drug-like,* ou fármaco-similares. A seleção de compostos com propriedades *drug-like,* que deu origem às estratégias de *virtual screening*, ou seleção virtual, será tema de uma secção apresentada mais adiante (Item 1.4.).

1.3. Resultados Falso-positivos

Fail fast. Fail cheap

Conforme apresentado na seção anterior, os testes de *HTS* representaram um grande avanço na velocidade com que dados de atividade biológica poderiam ser coletados. Este

aumento de velocidade só foi possível graças a fatores como a redução da escala dos testes executados, o aumento da sensibilidade dos métodos de detecção e a possibilidade de utilizar sistemas biológicos isolados para monitorar os sistemas responsáveis pela atividade biológica.[Rishton, 1997] Entretanto, estes fatores também foram os responsáveis por inúmeros contratempos associados aos testes de *HTS*, como por exemplo, a ocorrência de resultados falso-positivos.

Em um procedimento normal de busca de compostos ativos, utilizando-se testes em sistemas modelos de atividade (como, por exemplo nos testes de *HTS*), os compostos que apresentam resultados positivos são selecionados e encaminhados para a etapa de otimização, que é realizada interativamente, através da síntese de derivados, da proposição de modelos de estrutura-atividade, da avaliação das propriedades ADMET, da obtenção de compostos mais ativos e, assim sucessivamente, até que se obtenha uma estrutura líder adequada para ser utilizada nas etapas clínicas de desenvolvimento. Uma estrutura líder ideal deve tanto apresentar alta afinidade pelo alvo bem como deve formar com este uma ligação não-covalente, reversível, com mecanismo de interação competitivo e independente do tempo.[Rishton, 2003; Oprea *et al.*, 2001] Adicionalmente, é desejável dispor de séries de análogos estruturais da estrutura líder, com modelos *SAR* ou *QSAR* conhecidos. Quando, após estas etapas de otimização da estrutura inicial, ainda não se obtém um composto líder que agregue as condições ideais e necessárias para seguir adiante com as etapas clínicas, o resultado do teste de atividade, em que a estrutura inicial foi identificada, é classificado como um resultado falso positivo. Em outras palavras, um resultado falso positivo ocorre quando um composto não evolui - por diversos motivos - para se tornar uma estrutura líder candidata a fármaco, apesar de apresentar uma resposta positiva no teste de atividade.

Uma vez que durante as etapas de otimização da estrutura inicial são investidas quantidades significativas de tempo e de dinheiro, a identificação precoce de um resultado falso positivo pode representar uma economia de tempo e de dinheiro para aqueles envolvidos no processo de desenvolvimento do fármaco. Nas indústrias farmacêuticas este conceito é conhecido pela expressão:

"*Fail fast. Fail cheap*" [Kamath, 2006]

que pode ser traduzido por, "Falhe rápido. Falhe barato". Ou seja, resultados falso-positivos são uma preocupação para as indústrias farmacêuticas ou para qualquer grupo envolvido com o teste de compostos bioativos, pois a falha em identificar precocemente estes resultados enganosos leva invariavelmente a perdas dos investimentos feitos neles. Considerando-se a grande quantidade de dados e o grande número de compostos utilizados em testes de *HTS*, a remoção precoce e sistemática de compostos que provocam resultados falso-positivos tornou-se vital para o sucesso de programas que utilizam testes *HTS*. Assim, esta necessidade motivou para diversos trabalhos que procuram compreender, identificar e prever a ocorrência de resultados falso-positivos.[Malvezzi *et al.*, 2008; Rishton, 2003; McGovern *et al.*, 2002; Roche *et al.*, 2002; Rishton, 1997]

Numa revisão da literatura, Rishton[Rishton, 2003] discute as diferentes razões para a ocorrência de resultados falso-positivos em testes de *HTS*. O autor identificou e classificou as três principais razões, mais comumente citadas na literatura, como sendo: (*i*) o uso de compostos capazes de interferir diretamente com o método de detecção;[Roche *et al.*, 2002] (*ii*) o uso de compostos reativos frente à proteínas;[Rishton, 1997] e (*iii*) compostos capazes de formar agregados moleculares.[McGovern *et al.*, 2002] A seguir, será apresentada em detalhes, cada uma destas razões para a ocorrência de resultados falso-positivos.

1.3.1. *Frequent hitters*

Analisando alguns resultados de testes de *HTS* da *F. Hoffmann-La Roche & Co.*, Schneider[Roche *et al.*, 2002] selecionou compostos que haviam provocado resultados positivos (*Hits*) em pelo menos oito diferentes ensaios de *HTS*. Por serem compostos freqüentemente identificados em testes de *HTS*, Schneider chamou esta classe de compostos de "*Frequent hitters*". O autor identificou que estes, em sua maioria, eram capazes de perturbar ou o teste ou o método de detecção do teste, fazendo com que artefatos experimentais fossem - erroneamente - interpretados como resultados positivos. Fazem parte da classe de compostos "*Frequent hitters*" moléculas coloridas ou fluorescentes que interferem diretamente com o detector; compostos muito lipofílicos e/ou insolúveis em água que precipitam ou turvam o experimento impedindo a leitura correta do resultado e compostos capazes de precipitar proteínas presentes no meio reacional do

experimento. Exemplos de compostos que compõem a classe de "*Frequent hitters*" são compostos poliaromáticos, polifenólicos, altamente lipofílicos e altamente conjugados.

1.3.2. Compostos reativos frente à proteínas

Em sua tentativa de auxiliar a identificação de compostos que promovem resultados falso-positivos, Rishton[Rishton, 1997] identificou uma série grupamentos eletrofílicos, suscetíveis à hidrólise ou à solvólise e que apresentam especial reatividade frente a grupamentos nucleofílicos de proteínas. A **Figura 5** apresenta os grupamentos suscetíveis à reação que, segundo o autor, são os responsáveis por resultados falso-positivos em testes de *HTS*.

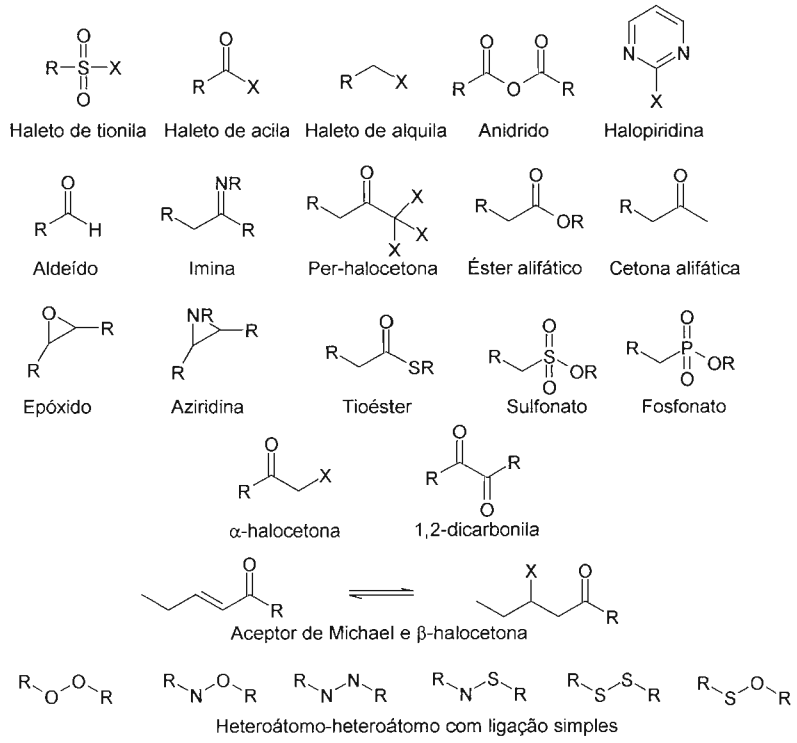


Figura 5. Grupos reativos frente à proteínas responsáveis por resultados falso-positivos em testes de *HTS*. [Rishton, 1997]

De acordo com Rishton, a inclusão de compostos contendo estes grupamentos alquilantes e/ou acilantes nos ensaios enzimáticos ou celulares (como ocorre nos testes de *HTS*) pode conduzir a inibição inespecífica das enzimas/proteínas presentes no teste. Esta reatividade pode ser, erroneamente, interpretada como a responsável pela atividade

desejada e, assim, o composto passa a ser considerado como um candidato à estrutura líder. Entretanto, durante as etapas subseqüentes de desenvolvimento, esta reatividade não-especifica acarreta em uma elevada toxicidade que, no entanto, não pode ser reduzida sem a concomitante perda da potência do composto. Assim, aquele resultado inicial observado, supostamente positivo, acaba se tornando um resultado falso-positivo nas etapas avançadas de desenvolvimento. Rishton sugere que sejam classificados como falso-positivos todos os compostos que apresentem uma relação estrutura-reatividade, ou seja, quando a "potência" do composto se correlaciona com um descritor da reatividade do grupo funcional eletrofílico. A **Figura 6** apresenta um exemplo hipotético da relação estrutura-reatividade, em que o composto mais potente é também aquele mais reativo, nas condições do experimento.

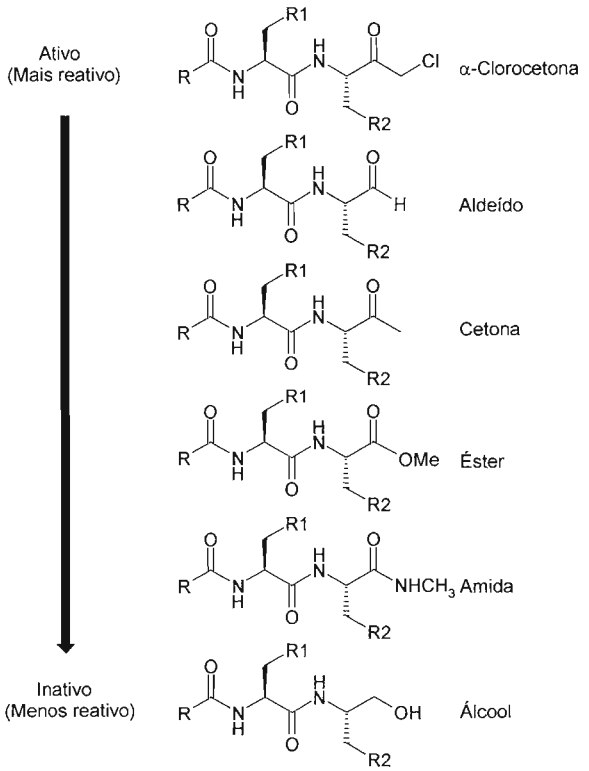


Figura 6. Exemplo de relação estrutura-reatividade, em que o composto mais potente é também o mais reativo frente a proteínas [Rishton, 1997]

1.3.3. O mecanismo promíscuo

Shoichet e colaboradores[McGovern *et al.*, 2002] identificaram compostos provenientes de programas de *HTS* que apresentavam as características estruturais

comumente descritas como *drug-like*[Lipinski *et al.*, 1997] capazes de inibir o alvo, em concentrações μM, ou sub-μM.[Shoichet, 2006] Estes compostos, no entanto, apresentavam mecanismo não-competitivo, correlações estrutura-atividade não-significativas, curvas dose-resposta íngremes e baixa especificidade. Estas propriedades são indesejadas (não *drug-like*) para uma estrutura líder e, portanto, estes compostos são considerados falso-positivos. Shoichet selecionou série inicial de 15 compostos para investigar as causas deste comportamento inesperado. Shoichet observou que independentemente do alvo inicial para o qual os compostos haviam sido inicialmente selecionados, todos eram capazes de inibir (em concentrações μM) outras enzimas não relacionadas (β-lactamase, quimotripsina, diidrofolato redutase e β-galactosidase) de um modo reversível, não-covalente e dependente do tempo (*i.e.*, observa-se um aumento da potência com pré-incubação do composto com a enzima). Adicionalmente, foi observado que estes compostos não eram desnaturantes de proteínas e que sua potência diminuía significativamente na presença de concentração dez vezes maior da enzima ou na presença de outras proteínas (*e.g.*, albumina bovina). A inibição indiscriminada de enzimas, de diferentes naturezas, motivou a denominação destes compostos como "inibidores promíscuos". Em um trabalho subseqüente, os autores[McGovern *et al.*, 2003] demonstraram a sensibilidade destes inibidores promíscuos frente a detergentes não-iônicos, como o *Triton X-100*. Shoichet observou que a adição de 0,1% (v/v) de *Triton X-100* era capaz de reverter a inibição provocada por um inibidor promíscuo. O efeito de detergentes na atividade de inibidores promíscuos foi estudada e corroborada por outros grupos de pesquisa.[Malvezzi *et al.*, 2008; Ryan *et al.*, 2003]

Sugeriu-se, então, que estes compostos estariam agindo através da formação de agregados moleculares e que estes agregados seriam os responsáveis pela inibição enzimática. Para testar a hipótese dos agregados moleculares foram feitos, entre outros, experimentos de *DLS* (*Dynamic Light Scattering*). Nestes experimentos, foram observadas partículas com valores de diâmetro variando de 95 a 400 nm, para todos os compostos testados. A seguir, experimentos de diálise mostraram que enzimas incubadas com os inibidores promíscuos não perdiam sua potência após a diálise, enquanto que enzimas incubadas com inibidores específicos competitivos recuperavam completamente a atividade após a diálise. Este comportamento era consistente com o modelo de agregação molecular proposto. Finalmente, os agregados moleculares foram diretamente observados por

microscopia eletrônica de transmissão. A **Figura** 7 apresenta exemplos de agregados moleculares formados por inibidores promíscuos, observados por microscopia eletrônica de transmissão.

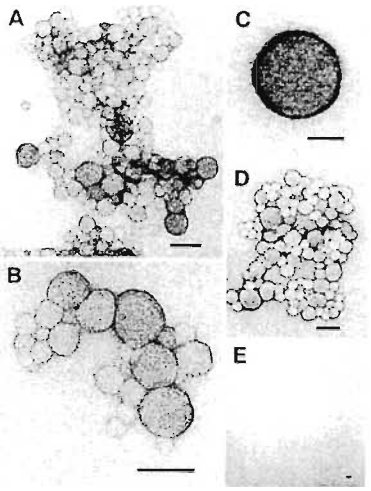


Figura 7. Compostos visualizados por microscopia de transmissão eletrônica em tampão Tris (20mM) pH 7,2. (A-C) Tetraiodofenolftaleína (100 μM); (D) Vermelho Congo (50 μM), controle positivo de agregação; (E) Ácido 8-anilino-1-naftalenossulfônico (625 μM) controle negativo de agregação. Barra = 100 nm. [McGovern *et al.*, 2002]

O mecanismo de inibição enzimática através da formação de agregados moleculares, proposto por Shoichet, pode ser aplicado a diversos grupos de moléculas e explica a baixa especificidade e as *SARs* não significativas observadas. Adicionalmente, esta proposta mecanística explica a variação da potência de inibição do composto com a concentração da enzima, uma vez que a concentração dos agregados é da mesma ordem de grandeza que a concentração das enzimas. Assim, um aumento de 10 vezes na concentração da enzima pode superar a capacidade de inibição destes agregados. Por fim, o modelo de formação de agregados pode explicar também a perda da potência de inibidores promíscuos na presença de detergentes, uma vez que os agregados são quebrados pela ação destes. Em duas publicações mais recentes, Shoichet e colaboradores[Feng & Shoichet, 2006; Feng *et al.*, 2005] propuseram série de testes que podem ser feitos para detectar e confirmar a ocorrência de inibição através do mecanismo promíscuo. Alguns destes testes podem ser realizados de modo *High-throughput*[Feng *et al.*, 2005] e podem ser aplicados a grandes bancos de dados, enquanto que outros são mais demorados e somente podem ser feitos com um pequeno número de compostos.[Feng & Shoichet, 2006] Dentre os testes sugeridos, o

mais confiável é o teste de sensibilidade da inibição dos compostos na presença de detergentes não-iônicos. Outras evidências experimentais da ocorrência de inibição pelo mecanismo promíscuo incluem: a sensibilidade da inibição dos compostos à concentração da enzima, a inibição inespecífica de enzimas não correlacionadas, a observação de inibição não-competitiva; a perda da potência após centrifugação, a observação de partículas de 50-1000 nm por *DLS* e a observação de curvas dose-resposta íngremes. Nenhum destes testes realizados isoladamente fornece confirmação definitiva do mecanismo promíscuo, entretanto quando combinados, são uma forte indicação da sua ocorrência. Outros trabalhos na literatura foram apresentados[Malvezzi *et al.*, 2008] corroborando com a hipótese de inibição enzimática por agregados moleculares.

Para os três tipos de resultados falso-positivos, citados anteriormente, as causas para tal comportamento são condições criadas pelo próprio experimento, independentemente do composto testado ser ou não um bom candidato a fármaco. Em outras palavras, compostos que são hoje usados na clínica médica e que são sabidamente bons fármacos podem provocar resultados falso-positivos, se testados nas condições dos testes *in vitro* de *HTS*. De fato, em todos os relatos, tanto de Schneider,[Roche *et al.*, 2002] em que foram identificados os compostos *frequent hitters*, como os de Rishton,[Rishton, 1997] que identificou os compostos reativos, quanto no de Shoichet,[McGovern *et al.*, 2003] que propôs o mecanismo de inibição de proteínas por agregados moleculares, são apresentados exemplos de fármacos conhecidos e com mecanismos específicos, que promovem resultados falso-positivos quando testados nas mesmas condições em que são realizados os testes de *HTS*.

Por exemplo, a ciclosporina e o paclitaxel são fármacos lipofílicos (CLOGP > 5) que poderiam ser classificados como *frequent hitters* por serem insolúveis em água e formarem precipitados que perturbam os métodos de detecção. Igualmente, a cisplatina, a ciclofosfamida e o clorambucil são exemplos de fármacos alquilantes antineoplásicos, altamente reativos frente a grupamentos nucleofílicos de proteínas e que seriam capazes de inibir, indistintamente, qualquer enzima. E por fim, a delavirdina, o benzoato de benzila e o miconazol, são exemplos de fármacos identificados como formadores de agregados moleculares[Seidler *et al.*, 2003] e, portanto, capazes de inibir proteínas de diferentes naturezas através do mecanismo promíscuo. Todos estes compostos, e muitos outros

antibióticos e/ou antineoplásicos que são hoje utilizados foram descobertos e introduzidos na terapêutica a partir de métodos clássicos de química medicinal, como ensaios biológicos *in vivo,* ensaios de citotoxicidade seletiva, ou descobertos ao acaso. Se estes compostos tivessem sido originalmente testados em ensaios bioquímicos de *HTS*, possivelmente seriam descartados dos bancos de dados por promoverem, repetidamente, resultados falso-positivos.

Esta aparente fragilidade de testes bioquímicos *in vitro* não diminui sua aplicabilidade ou sua importância para o descobrimento de novos fármacos, mas ressalta a importância dos limites dentre os quais os ensaios de *HTS* devem ser usados. Assim, os modelos *in vitro* de atividade biológica são testes de afinidade ligante-receptor ultra-sensíveis, portanto, sujeitos a artefatos e a reações indesejadas que podem mascarar um resultado positivo. Retomando o conceito de "*fail fast fail cheap*",[Kamath, 2006] os critérios para a seleção de resultados positivos em testes *HTS* devem considerar tanto os riscos como os benefícios de uma eliminação precoce e rápida de compostos, uma vez que tal procedimento pode minimizar o desperdício de recursos, mas ao mesmo tempo pode barrar o desenvolvimento de compostos que, se descobertos por outras tecnologias, poderiam vir a se tornar fármacos. Finalmente, o uso de filtros de compostos para eliminar possíveis resultados falso-positivos, ou a identificação de compostos com propriedades indesejadas, pode ser aplicado em bancos de dados de compostos antes mesmo destes serem submetidos a ensaios experimentais de atividade (*in vitro* ou *in vivo*), reduzindo-se ainda mais os custos e o tempo para se encontrar um composto líder. O sucesso desta estratégia, denominada *virtual screening*, vai depender da qualidade dos filtros utilizados e da qualidade do banco de dados. Estes aspectos relacionados ao *virtual screening* serão apresentados e discutidos a seguir.

1.4. *Virtual screening*

Testes 'in silico' de atividade biológica

Como discutido anteriormente nesta tese de doutorado, a pré-seleção de compostos antes que estes sejam submetidos aos testes de atividade se tornou uma necessidade em programas de desenvolvimento de fármacos, considerando-se o grande número de compostos disponíveis e os altos custos das etapas de desenvolvimento.[Bleicher *et al.*,

2003] Esta seleção de compostos, que classicamente era feita manualmente por um químico medicinal com base em sua intuição química ou conhecimentos e experiências acumuladas ou através de análises de *QSAR* considerando apenas um pequeno número de compostos, passou a ser feita automaticamente por computador (de modo virtual), aplicando-se filtros de diferentes naturezas. A crescente sofisticação na aplicação destes filtros permitiu que os testes de atividade, que eram realizados somente *in vivo* e/ou *in vitro* passassem a ser realizados *in silico*, ou seja, utilizando o computador. Os teste de atividade realizados no computador são denominados *virtual screening*, ou testes virtuais. Assim, para casos envolvendo um grande número de compostos, a estratégia de tentativa-e-erro foi substituída pela estratégia de tentativa-e-seleção.[Wess *et al.*, 2001] Assim, de modo análogo as estratégias utilizadas para séries pequenas, os dados de compostos com atividades conhecidas passaram a ser usados na construção de modelos de atividade biológica virtuais. Apenas os compostos selecionados nos testes virtuais são submetidos aos testes de atividade biológica experimental.[Shoichet, 2004]

Dentre as vantagens da aplicação do *virtual screening* em relação aos testes de *HTS*, estão a possibilidade do teste de compostos ainda não sintetizados e a economia de recursos, uma vez que estes testes são realizados apenas no computador sem o consumo de qualquer material.[Klebe, 2006] Outra vantagem é que questões tais como solubilidade, agregação, ou qualquer outra propriedade do composto que possa interferir com as condições do experimento não precisam ser consideradas em um primeiro momento, ou se consideradas, estas questões não interferem no resultado do teste. Adicionalmente, os testes de *virtual screening* podem ser rápidos (dependendo da complexidade o teste virtual pode levar de uma fração de segundo a muitos dias de cálculo, por composto) e podem ser prontamente repetidos ou refeitos caso novas informações sejam disponíveis ou novas condições sejam testadas.

O principal pré-requisito para um teste de *virtual screening* é o conhecimento das propriedades responsáveis pelas interações entre o composto e o sistema biológico,[Klebe, 2006] incluindo-se propriedades como a solubilidade do composto, a sua absorção e biodisponibilidade no organismo, a sua distribuição nos diversos compartimentos fisiológicos, o seu transporte, o metabolismo e excreção pelo corpo e, finalmente, sua interação com o alvo. Partindo-se destas informações, um modelo de atividade é construído

e usado para a seleção de compostos de um banco de dados. Por melhor que seja o poder de previsão do modelo proposto, os resultados devem ser sempre considerados apenas como previsões de atividade, ou seja é necessária a validação experimental dos resultados [Feng *et al.*, 2005], principalmente dos resultados positivos.

Além de um modelo preditivo de *virtual screening*, o sucesso desta estratégia depende também da preparação de um banco de dados de compostos. Considerações sobre o número de compostos e sobre a diversidade estrutural de bancos de dados para serem usados em *virtual screening* serão apresentadas e discutidas mais a diante

A proposição de modelos de *virtual screening* válidos é dificultada pela complexidade inerente às interações ligante-receptor ao nível molecular além de outros fatores tais como: múltiplos modos de ligação de um ligante,[Mattos & Ringe, 1993] múltiplas conformações do ligante e do receptor,[Klebe, 2006] ligação com proteínas plasmáticas bem como estabilidade metabólica, absorção, distribuição, excreção e toxicidade do ligante.[Hodgson, 2001] Tais dificuldades são dificilmente superadas por um único modelo computacional e este, mesmo que bem-sucedido, não poderia ser extrapolado para classes de compostos desconhecidas, ainda não consideradas pelo modelo.[Oprea, 2002] Uma alternativa sugerida envolve a aplicação seqüencial de diferentes filtros de seleção.[Klebe, 2006] Cada filtro proposto é específico para determinada propriedade ou característica estrutural. Assim, por exemplo, determinado filtro pode selecionar compostos capazes de serem absorvidos por via oral, enquanto que outro selecionaria compostos que contém determinado grupo funcional necessário para a interação com o receptor e, por fim, um terceiro avaliaria a afinidade entre o ligante e o receptor. Não é sugerida uma hierarquia para a aplicação de cada um destes filtros, ou seja, os filtros podem ser aplicados em qualquer seqüência sem, em princípio, alterar o resultado final da seleção. Entretanto, com o objetivo de se otimizar o uso dos recursos computacionais, em geral, os filtros são aplicados em ordem crescente de complexidade do cálculo. Assim, para o banco de dados original e inteiro se aplicam primeiramente filtros que necessitam baixos recursos computacionais, deixando os filtros mais complexos e elaborados para serem aplicados nas etapas finais do processo, quando o banco de dados já foi reduzido significativamente. Uma das desvantagens da aplicação hierárquica de filtros de seleção é que os filtros mais rápidos são também os mais grosseiros, ou seja, podem eliminar precocemente quantidades

significativas da diversidade estrutural e de informação química relevante, podendo, assim, descartar potenciais bons candidatos. Quando isto acontece, estes resultados são denominados resultados falso-negativos.

Assim como no caso dos testes de *HTS*, os testes de *virtual screening* são permeados de resultados falso-positivos e falso-negativos causados por limitações intrínsecas de cada método. Novamente, esta aparente fragilidade não diminui a importância do *virtual screening* como ferramenta para o descobrimento de fármacos, uma vez que o seu sucesso deve ser avaliado considerando-se o descobrimento de pelo menos *alguns* compostos candidatos e não somente para a descoberta de *todos* os compostos candidatos presentes no banco de dados.[Shoichet, 2004] Adicionalmente, o *virtual screening* apresenta outras características semelhantes às do *HTS* como: o uso de grandes bibliotecas de compostos; a grande velocidade dos testes em detrimento à precisão dos resultados e a necessidade de validação posterior dos resultados.

Doman e colaboradores,[Doman *et al.*, 2002] descrevem um estudo comparativo dos procedimentos de *virtual screening* e de *HTS* aplicados na busca de novos inibidores da tirosina fosfatase-PTP1B (um possível alvo para o tratamento da diabetes tipo II) utilizando bibliotecas com 400.000 e 235.000 compostos, respectivamente, para cada estratégia. Através do procedimento de *HTS* foram encontrados 85 (0,021%) compostos com valores de IC_{50} menores do que 100 μM enquanto que o procedimento de *virtual screening* selecionou 365 compostos para serem testados experimentalmente, dos quais 127 (34,8%) apresentaram valores de IC_{50} menores do que 100 μM. Como esperado, o procedimento de *virtual screening* enriqueceu o banco de dados 1700 vezes em relação à simples busca sistemática do *HTS*. Entretanto, interessantemente, cada uma das técnicas encontrou classes de compostos diferentes, sugerindo-se que os dois procedimentos podem ser complementares. A **Tabela 1** apresenta um resumo dos dados apresentados no trabalho.[Doman *et al.*, 2002]

Tabela 1. Resultados da busca por inibidores da tirosina fosfatase-PTP1B utilizando as estratégias de *HTS* e *Docking*. Número de compostos testados, número de compostos ativos identificados e porcentagem de compostos ativos [Doman *et al.*, 2002]

Estratégia	Compostos testados	Compostos com IC_{50}<100μM	Porcentagem de compostos ativos
HTS	400.000	85	0,021
Docking	365	127	34,8

Conforme mencionado anteriormente, a redução do número de compostos de um banco de dados utiliza diversos tipos filtros, específicos para determinadas propriedades ou características estruturais. De modo geral, os filtros podem ser agrupados em três categorias distintas: (*i*) filtros uni- e bidimensionais (1D e 2D), que se baseiam nas propriedades físico-químicas e na conectividade dos compostos (filtros *drug-like*); (*ii*) filtros tridimensionais (3D), que se baseiam na distribuição espacial de características específicas (filtros farmacofórico) e, (*iii*) filtros de interação molecular, que se baseiam em medidas de afinidade entre o composto e seu alvo biológico (filtros de *docking*). A seguir, cada um destes tipos de filtros será discutido em detalhes .

1.4.1. Filtros *drug-like*

Os filtros de compostos *drug-like* são baseados no conceito de que fármacos, em geral, apresentam determinadas propriedades que os distinguem dos demais compostos orgânicos. Estas propriedades podem ser de natureza físico-química, como a massa molecular ou o coeficiente de partição do composto; ou relativas à conectividade molecular, como a ausência de grupos funcionais sabidamente tóxicos ou a presença de um grupo funcional, que pode promover uma determinada atividade biológica. [Oprea, 2000]

Uma das principais vantagens dos filtros *drug-like* é que estes podem ser calculados muito rapidamente, sendo possível o processamento de milhares de compostos por segundo, em um computador pessoal.[Charifson & Walters, 2002] Outra vantagem é que estes filtros são capazes de reduzir significativamente o número de compostos em bancos de dados comercialmente disponíveis. Estes dois aspectos tornam os filtros *drug-like* os primeiros a serem aplicados em bancos de dados para a eliminação precoce de compostos indesejados. Este procedimento pode ser em parte especulativo e subjetivo, uma vez que estes filtros são sabidamente imprecisos e que os limites que separam compostos desejados, de indesejados, são definidos arbitrariamente. Assim, para reduzir o número de resultados falso-positivos ou falso-negativos, é importante examinar periodicamente os compostos eliminados pelo filtro, para evitar que compostos interessantes sejam eliminados inadvertidamente. Compostos com certas propriedades indesejadas podem ser otimizados em etapas subseqüentes.[Charifson & Walters, 2002]

Um dos exemplos mais conhecidos de aplicação de filtros físico-químicos para a seleção de compostos *drug-like* é a "regra dos 5" ou "*Rule of 5*", publicada por Lipinski e colaboradores.[Lipinski *et al.*, 2001] Partindo do banco de dados *World Drug Index - WDI* (um banco de dados com mais de 80.000 medicamentos comercialmente disponíveis), os autores selecionaram 2.245 compostos apresentando biodisponibilidade oral e com características de fármacos. Para tanto, os autores selecionaram do banco de dados original, apenas os compostos que haviam passado por ensaios clínicos fase II e que receberam alguma indicação de uso terapêutico. Adicionalmente, os autores excluíram os compostos poliméricos, peptídicos e sais quaternários de amônio. A seguir, foram calculadas as seguintes propriedades dos compostos selecionados: massa molecular; coeficiente de partição (logP, calculado pelo programa CLOGP[Biobyte CORP., 1995]); número de grupamentos doadores de ligação de hidrogênio (representado pela soma de grupamentos –OH e –NH); e número de grupamentos aceptores de ligação de hidrogênio (representado pelo número de átomos de oxigênio e de nitrogênio). Com base nestas propriedades, os autores observaram que em mais de 90% dos compostos *drug-like* selecionados, o valor de massa molecular era $\leq$500, o valor do logP era $\leq$5, o número doadores de ligação de hidrogênio era $\leq$5 e, o número de aceptores de ligação de hidrogênio era $\leq$10. Assim, foi criada a regra mnemônica "regra dos 5", uma vez que todos os valores são múltiplos de 5. A regra afirma que baixa absorção ou, baixa permeabilidade, é esperada para compostos que violem duas ou mais das condições citadas acima. As exceções à regra ocorreram em compostos derivados de produtos naturais, incluindo antibióticos, antifúngicos, vitaminas e glicosídeos cardiotônicos. Os autores sugerem que, apesar de violarem a "regra dos 5", estas exceções são biodisponíveis, por via oral, porque podem ser substratos de transporte ativo.

O trabalho apresentado por Lipinski, em 1997,[Lipinski *et al.*, 2001] foi citado até o presente momento mais de 1290 vezes [segundo consulta feita no *Web of Science* em 11/2007] e representa a consolidação da idéia de que fármacos ocupam uma região distinta dentro do espaço químico. A partir deste trabalho, diversos programas de síntese ou de aquisição de compostos passaram a incluir filtros para a "regra dos 5" de Lipinski. Este procedimento pode ser considerado apropriado quando o objetivo do programa for a seleção de fármacos biodisponíveis por via oral. Entretanto, se o objetivo do trabalho for a

descoberta de estruturas líder, os critérios de seleção estabelecidos por Lipinski deixam de ser adequados, uma vez que estes foram propostos para encontrar fármacos biodisponíveis e não estruturas líder.

A diferença entre um fármaco e uma estrutura líder foi estudada por Oprea e colaboradores.[Oprea *et al.*, 2001] Por definição, uma estrutura líder é um composto com afinidade pelo alvo biológico, com propriedades próximas às ótimas, nas quais se introduzem modificações para se tornar um fármaco. Oprea e colaboradores realizaram estudo comparativo sistemático entre duas séries de compostos, uma formada por 75 fármacos com diferentes indicações terapêuticas e outra formada pelas 62 estruturas líder que originaram estes fármacos (em alguns casos, uma mesma estrutura líder originou mais de um fármaco).

Para as duas séries, foram calculados os seguintes parâmetros físico químicos: massa molecular; lipofilicidade (logP, obtido por CLOGP[Biobyte CORP., 1995]); número de grupamentos doadores de ligação de hidrogênio; número de grupamentos aceptores de ligação de hidrogênio; número de ligações com rotação livre; e número de anéis.

A análise comparativa da evolução das propriedades das estruturas líder, com aquelas correspondentes ao(s) fármaco(s) resultante(s), mostrou que: (*i*) a massa molecular aumenta (79 daltons); (*ii*) a lipofilicidade aumenta (1,25 unidades de logP); (*iii*) duas ligações com rotação livre são adicionadas; (*iv*) 0,45 aceptores de ligação de hidrogênio são adicionados; (*v*) nenhuma mudança no número de doadores de ligação de hidrogênio; e (*vi*) 0,55 anéis são adicionados. Estes valores são calculados em relação à média e não são estatisticamente significativos. Entretanto, mostram que existe clara tendência de aumento da complexidade (por exemplo: aumento da massa molecular, aumento do número de anéis e de ligações com rotação livre) e da lipofilicidade de uma estrutura líder, quando esta é transformada em um fármaco. A tendência de aumento da massa molecular e da lipofilicidade do composto, quando se evolui de uma estrutura líder para um fármaco foi confirmada em outros estudos.[Hann *et al.*, 2001; Lipinski, 2000] Finalmente, analogamente ao trabalho de Lipinski, Oprea sugere valores de corte para as propriedades que devem ser usados em filtros para a seleção de estruturas líder. Assim, segundo os autores, a busca por estruturas líder deve selecionar compostos com massa molecular de $\leq$450; CLOGP entre -3,5 e 4,5; não mais que 4 anéis; não mais que 10 grupamentos

metileno terminais; número de doadores de ligação de hidrogênio ≤5 e; número de aceptores de ligação de hidrogênio ≤8. [Oprea *et al.*, 2001]

Um dos problemas ao se atribuir valores menores de massa molecular e de lipofilicidade para a seleção de potenciais compostos-líder é que compostos menores e menos lipofílicos tendem a ser também menos potentes,[Charifson & Walters, 2002] reduzindo, assim, as chances de se encontrar um composto ativo. Esta tendência pode ser explicada considerando o princípio utilizado por químicos medicinais para aumentar a potência *in vitro* de candidatos, segundo o qual a inclusão de grupamentos lipofílicos torna o composto candidato mais potente.[Lipinski *et al.*, 2001] [Hann *et al.*, 2001] Em contraste, a adição de grupamentos polares somente contribui para o aumento da potência do ligante, se estes forem adequadamente posicionados no receptor, de modo que contribuição negativa da dessolvatação do grupamento polar seja sobrepujada por uma contribuição positiva da interação deste com um grupamento polar do receptor. O efeito do posicionamento e o valor da contribuição energética de diferentes grupos funcionais para a energia de interação ligante-receptor foi estudado por Andrews[Andrews *et al.*, 1984] e será apresentada em detalhes mais adiante nesta tese (Item 1.5.).

Para se evitar a necessidade de uma atribuição subjetiva de valores de corte na seleção de compostos *drug-like*, uma alternativa é a utilização de modelos de inteligência artificial, nos quais programas de computador são treinados para reconhecer padrões apresentados à eles. Por exemplo, Sadowski e Kubinyi[Sadowski & Kubinyi, 1998] treinaram uma rede neural para diferenciar fármacos, de não-fármacos. Redes neurais são modelos de redes altamente interconectadas, capazes de classificar objetos com base em uma série de treinamento. Neste caso, os autores prepararam uma série de treinamento com 5000 fármacos, extraídos do *WDI* e 5000 não-fármacos, obtidos do *Available Chemical Directory – ACD* (banco de dados com mais de 160.000 compostos comercialmente disponíveis). Aos fármacos foi atribuído o valor 1 (um) e aos não-fármacos foi atribuído o valor 0 (zero). Estes compostos foram submetidos à rede neural e esta foi treinada para reconhecer cada um deles, considerando-se descritores topológicos. Após o treinamento da rede neural, o modelo foi utilizado para a previsão dos demais compostos nos dois bancos de dados *WDI* e *ACD,* respectivamente. A **Figura 8** apresenta a distribuição dos valores preditos para os dois bancos de dados. As linhas sólidas correspondem a cada uma das

séries de treinamento com 5000 compostos do *ACD* e 5000 compostos do *WDI*, respectivamente. As linhas pontilhadas correspondem à predição de 169.331 compostos do *ACD* e 38.416 compostos do *WDI*, respectivamente.

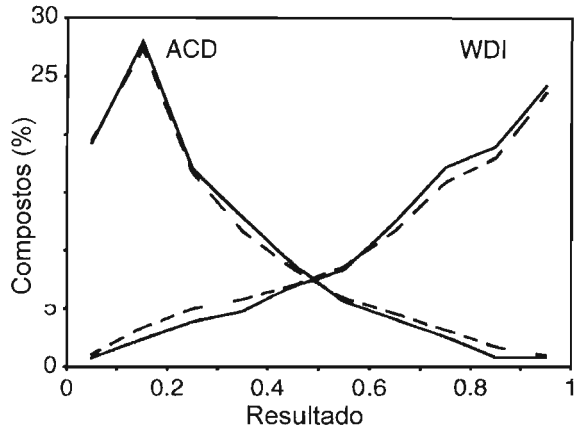


Figura 8. Resultados do modelo de redes-neurais utilizado para distinguir fármacos (*ACD*) de não-fármacos (*WDI*). As linhas sólidas correspondem às séries de treinamento e as linhas pontilhadas correspondem à aplicação do modelo sobre os dois bancos de dados, respectivamente.[Sadowski & Kubinyi, 1998]

Conforme pode ser observado na **Figura 8**, existe clara distinção entre os valores preditos para os fármacos do *WDI* dos não-fármacos do *ACD*. Entretanto, ainda se observam muitos fármacos do *WDI* com valores próximos de 0 e muitos não-fármacos do *ACD* com valores próximos de 1. Os autores consideram que esta imprecisão do modelo ocorre devido a erros de atribuição de valores na série de treinamento, principalmente para os compostos do *ACD*, que foram considerados não-fármacos, mas que na verdade podem ser fármacos ou fármaco-similares. Segundo os autores, o modelo proposto somente deve ser usado com o objetivo de enriquecer bancos de dados e não para avaliar compostos individualmente, uma vez que não se observa um valor que separe os dois conjuntos, ou seja, fármacos, de não-fármacos.

Em um trabalho semelhante, Ajay e colaboradores também foram capazes de separar fármacos de não-fármacos, utilizando algoritmo de rede neural.[Ajay *et al.*, 1998] Neste trabalho, 3.500 fármacos do banco de dados *Comprehensive Medicinal Chemistry - CMC* e 3.500 não-fármacos do *ACD* foram utilizados para o treinamento do modelo, empregando os seguintes descritores moleculares: logP; massa molecular; número de grupamentos doadores de ligação de hidrogênio; número de grupamentos aceptores de

ligação de hidrogênio; número de ligações com rotação livre; aromaticidade; ramificação e, índices topológicos (descritores ISIS). O modelo resultante foi utilizado para prever o caráter *drug-like* dos compostos do banco de dados *MDDR* (*MACCS Drug Data Report*). Aproximadamente, 80% dos compostos do *MDDR* foram corretamente classificados como *drug-like*.

Uma das principais críticas para este tipo de estratégia é que os modelos de aprendizado são treinados apenas para reconhecer fármacos que já são conhecidos. Ou seja, novas classes de compostos não seriam reconhecidas pelos modelos, a menos que suas propriedades fossem semelhantes às encontradas na série de treinamento. Um outro problema associado aos sistemas de inteligência artificial é que não é possível interpretar, e/ou, extrair regras gerais de redes neurais, uma vez que a contribuição relativa de cada propriedade para o modelo é desconhecida.

Todos os exemplos de classificação de fármacos por suas propriedades, citados nesta tese, mostram que os fármacos parecem se agrupar em regiões definidas do espaço químico, assim como galáxias espalhadas no universo (**Figura 9**), ou como ilhas e arquipélagos no oceano. [Kubinyi, 2005].

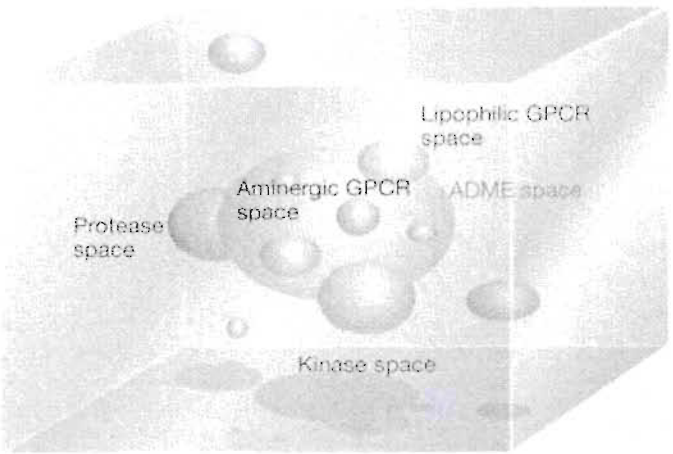


Figura 9. Representação esquemática da distribuição de compostos de diferentes classes terapêuticas, no universo de todos os compostos químicos. [Lipinski & Hopkins, 2004]

Em uma tentativa de realizar um mapeamento global do espaço químico, Oprea e Gottfries [Oprea & Gottfries, 2001] propuseram o *ChemGPS,* ou seja, *Chemical Global Positioning System* (Sistema de Posicionamento Global Químico). Analogamente a um

sistema de coordenadas cartesiano, o *ChemGPS* descreve compostos orgânicos a partir de "coordenadas" extraídas de uma análise de componentes principais (*PCA*). Os autores utilizaram 423 compostos, cuidadosamente selecionados, para cobrir o máximo da diversidade do espaço químico de fármacos. Para tanto, foram incluídos, além de compostos com propriedades tipicamente *drug-like*, compostos com valores extremos em uma, ou mais, destas propriedades, mas que ainda continham fragmentos *drug-like*. Estes compostos foram chamados de compostos satélites, pois marcam os limites do espaço químico que pode ser ocupado por um fármaco. Foram utilizados 72 descritores (moleculares e topológicos) representando as seguintes propriedades: tamanho lipofilicidade, polarizabilidade, carga, flexibilidade, rigidez e ligação de hidrogênio. Uma vez calculadas as "coordenadas" dos compostos, três componentes principais extraídas da análise são plotadas nos eixos cartesianos *X*, *Y* e *Z*, criando um espaço tridimensional, em que a posição relativa dos compostos pode ser uma representação destes no espaço químico. Segundo os autores, a utilização do sistema *ChemGPS* permite a visualização de regiões de interesse do espaço químico, como a região ocupada por fármacos e a região ocupada por estruturas líder. A definição destas regiões reduz o espaço químico que deve ser explorado para a procura de novos compostos de interesse. O *ChemGPS* permite tanto a comparação visual de bibliotecas de compostos quanto a percepção de que partes do espaço químico já foram exploradas. Um exemplo da aplicação do *ChemGPS* é a exploração do espaço químico ocupado por agentes antineoplásicos usados na terapia do câncer.[Lloyd *et al.*, 2006] A **Figura 10** mostra a distribuição de compostos anticâncer, inibidores de quinases, no espaço químico de compostos medicinais. As esferas azuis são os compostos ativos inibidores de quinases, os pontos em amarelo são compostos genéricos que passaram pela "regra dos 5" e em vermelho são compostos representando o espaço de compostos medicinais (ativos e inativos). Neste exemplo, pode-se constatar que compostos anticâncer, inibidores de quinases, ocupam regiões tanto dentro como fora do espaço definido pela "regra dos 5" e que, portanto, estas regras de seleção não se aplicam para esta classe de compostos.

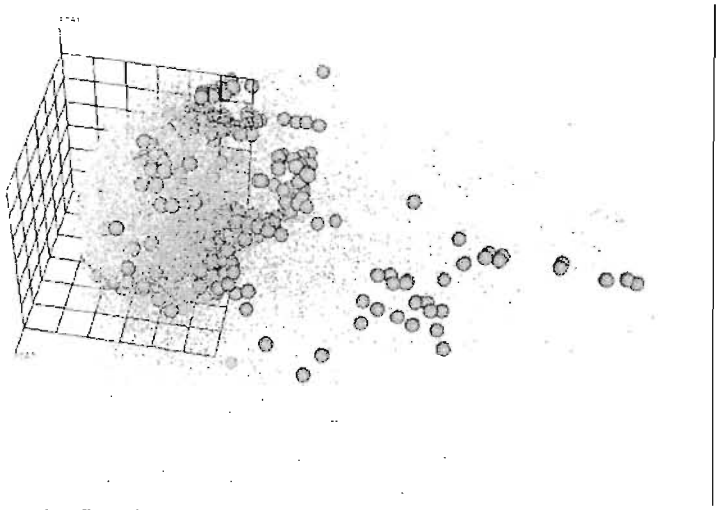


Figura 10. Gráfico das três componentes principais obtidas pelo modelo *ChemGPS* (*Chemical Global Positioning System*) mostrando a distribuição de compostos inibidores de quinase (esferas azuis), em relação a compostos que passaram pela "regra dos 5" (pontos amarelos) e outros compostos medicinais (pontos vermelhos).[Lloyd *et al.*, 2006]

Análises como esta mostram que existem diferenças entre regras globais e locais, para a seleção de composto *drug-like*. Por exemplo, como era de se esperar, inibidores de quinase têm propriedades diferentes de inibidores de protease, que são por sua vez diferentes das dos anestésicos locais. Ou seja, existem limites de precisão para a construção de regras globais que se aplicam a todas as situações. Uma alternativa para superar esta limitação é a construção de modelos que se aplicam somente a determinados subgrupos de fármacos com mesma indicação terapêutica. [Walters & Murcko, 2002]

Pelo exposto, pode-se concluir que nem todo composto biologicamente ativo possui as propriedades físico-químicas necessárias para se tornar um fármaco e, igualmente, nem todo composto com propriedades *drug-like* será um candidato a fármaco. Além de propriedades moleculares específicas, um fármaco deve também apresentar interações específicas entre o composto e seu alvo biológico. Os filtros *drug-like*, são filtros uni- ou bidimensionais, incapazes de considerar a distribuição espacial das propriedades *drug-like* em três dimensões. Os filtros que consideram a distribuição espacial de propriedades três dimensões são conhecidos como filtros farmacofóricos.

1.4.2. Filtros farmacofóricos

Filtros farmacofóricos são filtros que selecionam compostos a partir de um modelo farmacofórico. Por definição,[Wermuth *et al.*, 1998] um farmacóforo é o arranjo espacial

das propriedades estéricas e eletrônicas necessárias para garantir interações supramoleculares (não-covalentes) ótimas entre um ligante e seu alvo biológico específico, e, assim, ativar (ou inibir) uma resposta biológica. Um farmacóforo não é uma molécula real, ou uma associação de grupos funcionais, mas um conceito abstrato que incorpora as habilidades de interação molecular comuns em uma série de compostos, em relação a seu alvo biológico. O farmacóforo pode ser considerado o maior denominador comum, compartilhado por uma série de moléculas ativas, que são reconhecidas por um mesmo alvo.

Desta forma, um farmacóforo deve ser capaz de indicar quais grupos funcionais de um composto ativo são os responsáveis pelas interações com o alvo, e qual o posicionamento e as distâncias ideais entre estes grupos. Uma vez definido um farmacóforo, ele deve ser capaz de explicar tanto a atividade de compostos ativos bem como a falta de atividade de análogos inativos. Adicionalmente, através de um farmacóforo deve-se poder prever a atividade de novos compostos ainda não testados e/ou propor novos compostos ativos. Até que um farmacóforo tenha sido validado e apresente alto poder de previsão, qualquer farmacóforo proposto é chamado de "hipótese" ou de "modelo farmacofórico"

Modelos farmacofóricos podem ser obtidos a partir de série de compostos ativos (*i.e.* baseado na estrutura do ligante - *ligand based*), a partir da estrutura do receptor (*structure based)* ou, a partir de ambos. A principal vantagem de modelos farmacofóricos desenvolvidos a partir de séries de compostos ativos é que estes podem ser desenvolvidos independentemente de se conhecer, ou não, o alvo biológico. Neste caso, toda a informação necessária para a construção do modelo farmacofórico é extraída dos dados de atividade de uma série de compostos ativos. Entretanto, a qualidade do modelo proposto será limitada pela qualidade da série de compostos utilizada.[Triballeau *et al.*, 2006] Uma série ideal deve conter tanto compostos com estruturas diversas, incluindo compostos rígidos e flexíveis bem como compostos com estruturas simples e estruturas complexas, que explorem todo o espaço químico ao seu redor. Adicionalmente, a medida da atividade dos compostos deve ser específica e quantitativa e a série deve incluir compostos muito potentes, pouco potentes e inativos. Por fim, deve haver evidências de que todos os compostos se ligam ao mesmo sítio de ligação do alvo biológico. Na ausência de uma, ou

mais, destas condições, os modelos farmacofóricos gerados serão limitados às informações disponíveis e sua aplicabilidade também será limitada.

O maior problema associado à obtenção de modelos farmacofóricos baseados no ligante é a definição da conformação bioativa. O número de conformações possíveis de um composto flexível cresce exponencialmente com o número de ligações com rotação livre em sua estrutura. Na ausência de análogos rígidos (e ativos) ou de outras evidências experimentais, a definição da conformação bioativa de um composto flexível passa a ser um exercício de adivinhação. Normalmente, são usadas conformações de mínimo de energia do análogo mais potente da série (*Active Analog Approach*[Wolber & Langer, 2005; Pires *et al.*, 2001]), entretanto o uso destas conformações não garante que o modelo será válido.

Quando o modelo farmacofórico é baseado na estrutura do complexo ligante-receptor, a conformação bioativa é conhecida e, assim, a determinação do arranjo espacial dos pontos de interação é simplificada. Entretanto, uma vez que somente compostos com alta afinidade pelo receptor (ou fragmentos/ligantes com alta eficiência[Hopkins et al., 2004, Hartshorn et al., 2005]) podem ser co-cristalizados com o receptor, a falta de informações sobre o modo de ligação de compostos pouco ativos, ou inativos, impede a avaliação da contribuição relativa de cada ponto de interação entre o ligante e o receptor, dificultando a determinação de quais pontos são realmente fundamentais para a construção do modelo farmacofórico.[Triballeau *et al.*, 2006]

Uma vez definida a conformação bioativa, um outro obstáculo para a construção do modelo farmacofórico é o alinhamento dos demais de compostos sobre o composto de referência. Este alinhamento pode ser rígido, semiflexível ou flexível.[Poptodorov *et al.*, 2006] Os alinhamentos rígidos são os mais simples e rápidos, porém partem do princípio de que todos os compostos já se encontram em sua conformação bioativa. Esta aproximação quase nunca é válida e, portanto, alinhamentos rígidos são raramente usados. Métodos semiflexíveis são aqueles em que um determinado número de conformações são pré-estabelecidas para cada composto e estas são usadas na etapa de alinhamento rígido. As conformações podem ser obtidas por procura sistemática, por métodos de amostragem do espaço conformacional, como algoritmo genético, Monte Carlo ou dinâmica molecular.[Carlson *et al.*, 2000] Esta etapa de análise conformacional deve garantir

amostragem representativa do espaço ocupado pelos compostos, para que sejam maiores as chances de inclusão da conformação bioativa no conjunto de conformações geradas. Nos métodos de alinhamento flexível, a análise conformacional é realizada simultaneamente (*on-the-fly*) à etapa de alinhamento e, por isso, requerem grandes recursos computacionais.

Além dos problemas de definição da conformação bioativa e do alinhamento dos compostos, para que se obtenha um modelo farmacofórico é preciso identificar as propriedades estéricas e eletrônicas que compõem o farmacóforo. A percepção destas propriedades depende da forma com que os compostos são apresentados. Diferentes formas tautoméricas, diferentes estados de ionização e de dissociação, a atribuição correta de propriedades de ligação de hidrogênio e compostos quirais são algumas das dificuldades encontradas para a identificação correta do modelo farmacofórico.[Kubinyi, 2005]

Para exemplificar os problemas de flexibilidade e de identificação de propriedades físico-químicas ao se definir um modelo farmacofórico, a **Figura 11** apresenta série de diferentes modelos que podem ser propostos a partir da molécula da histamina (uma amina endógena envolvida na resposta imune). A **Figura 11** apresenta os seis diferentes modelos farmacofóricos resultantes da variação no estado de protonação do anel imidazólico, de uma rotação de 180° da cadeia lateral. O exemplo mostra como mesmo uma molécula relativamente simples - com uma única ligação com rotação livre e três heteroátomos - pode apresentar um problema desafiador ao se buscar um modelo farmacofórico.[Kubinyi, 2005]

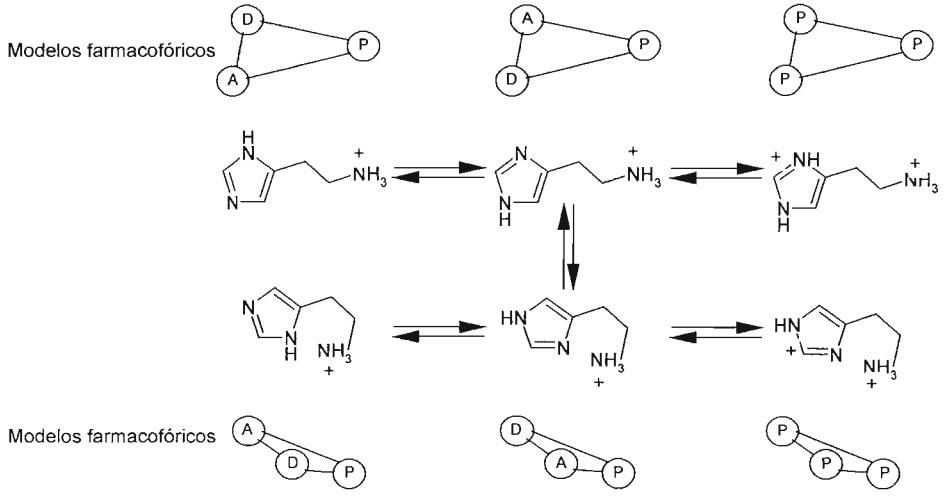


Figura 11. Representação esquemática dos 6 diferentes modelos farmacofóricos possíveis da histamina, a partir da mudança da protonação do anel imidazólico e de uma rotação de 180° da cadeia lateral.[Kubinyi, 2005]

O modo mais comum para se descrever um farmacóforo é através de restrições geométricas (*i.e.* distâncias, ângulos e ângulos diedros) e de propriedades físico-químicas, como ligações de hidrogênio e interações hidrofóbicas. Não existe um número ideal de pontos de interação para a proposição de um modelo farmacofórico. Um modelo farmacofórico pode ser formado tanto por apenas dois pontos de interação, como observado no exemplo da **Figura 12-A** (modelo farmacofórico de agonistas da serotonina, um neurotransmissor endógeno, formado por um anel aromático e um nitrogênio básico separados por 5,6Å [Hibert *et al.*, 1988]), bem como por múltiplos pontos de interação, como no exemplo da **Figura 12-B** (modelo farmacofórico de antifúngicos formado pelo arranjo espacial de duas regiões de interação hidrofóbica[verde], um anel aromático[vermelho], um nitrogênio aromático[azul] e duas regiões de exclusão de volume[preto] [Di Santo *et al.*, 2005]).

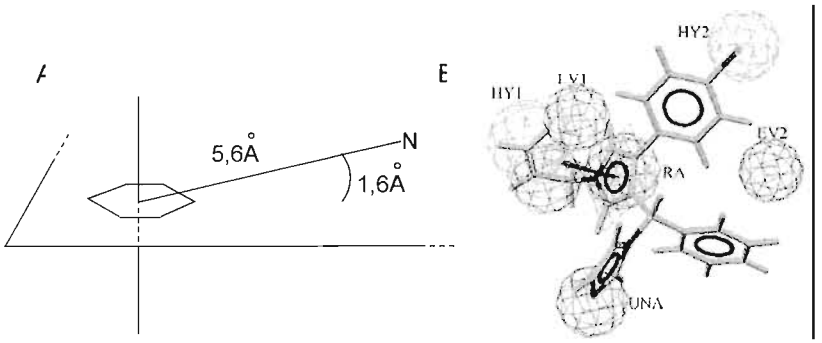


Figura 12. (A) Modelo farmacofórico de antagonistas de 5HT;[Hibert *et al.*, 1988] (B) Modelo farmacofórico de antifúngicos, posicionado sobre a estrutura do 1-[(Aril)(4-aril-1*H*-pirrol-3-il)metil]-1*H*-imidazol.[Di Santo *et al.*, 2005] Grades em verde representam regiões de interação hidrofóbica; grades em vermelho representam posições para um anel aromático, grade em azul representa posição de um nitrogênio aromático, grades em preto representam regiões de exclusão de volume.

Existem diversos programas, comercialmente disponíveis, que possibilitam reconhecer e gerar automaticamente modelos farmacofóricos. Exemplos destes programas são: *CATALYST*; [Catalyst, 2005] *DISCO*; [Sybyl, 2004] *GASP*; [Sybyl, 2004] e *LIGANDSCOUT* [Wolber & Langer, 2005]. Todos estes podem gerar modelos farmacofóricos quantitativos para estudos da relação estrutura atividade em três dimensões (QSAR-3D). Por exemplo, Bhattacharjee e colaboradores,[Bhattacharjee *et al.*, 2004]

geraram um modelo farmacofórico quantitativo, utilizando o programa *CATALYST,*[Catalyst, 2005] para identificar potenciais inibidores de uma quinase do *Plasmodium falciparum* (Pfmrk). Usando uma série de treinamento, com 15 inibidores estruturalmente diversos e com valores de atividade variando de 0,13-1100 μM, o melhor modelo proposto foi formado por duas regiões aceptoras de ligação de hidrogênio, por um anel aromático e por uma região hidrofóbica alifática (**Figura 13**). Este modelo foi submetido a validação externa, com uma série de teste com 15 inibidores apresentando os seguintes valores para os parâmetros estatísticos: coeficiente de correlação da série de treinamento r=0,9 e coeficiente de correlação da série de teste r=0,7. A seguir, este modelo foi usado para a seleção virtual em um banco de dados com, aproximadamente, 290.000 compostos. Foram selecionados 16 compostos com valor de atividade calculado inferior a 52 μM. Os 16 compostos foram testados experimentalmente e todos se mostraram ser inibidores da Pfmrk com valores de atividade *in vitro* abaixo de 100 μM.

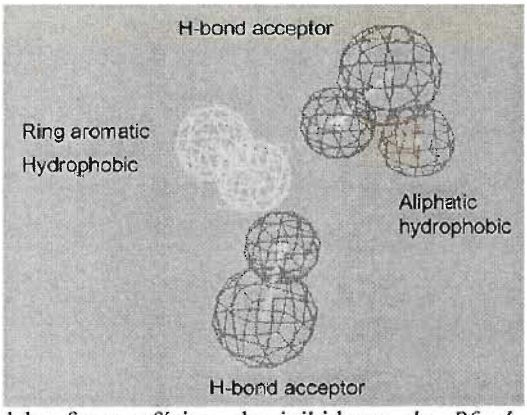
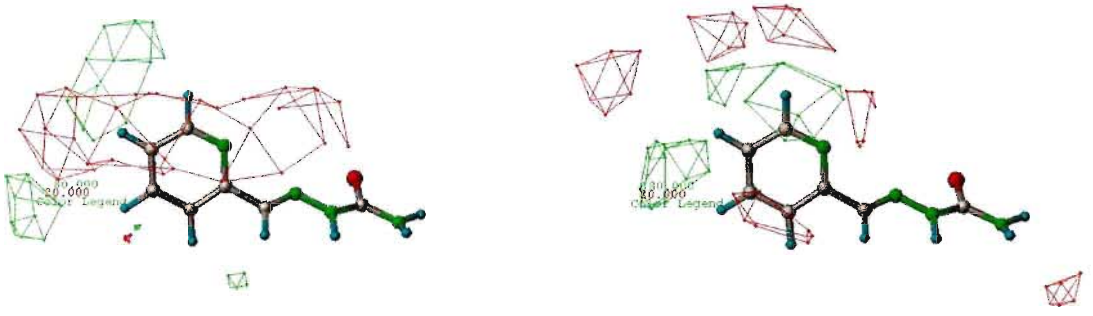


Figura 13. Modelo farmacofórico de inibidores da *Pfmrk.* Grades em verde representam regiões aceptoras de ligação de hidrogênio, grades em bege representam a posição de um anel aromático e a grade em azul representa uma região hidrofóbica alifática [Bhattacharjee *et al.*, 2004]

Além da representação através de propriedades físico-químicas, os modelos farmacofóricos também podem ser definidos por campos de interação molecular. Um dos métodos que descreve relações estrutura-atividade em três dimensões através de campos de interação molecular é a metodologia *CoMFA* (do inglês *Comparative Molecular Field Analysis*).[Cramer *et al.*, 1988] Na metodologia *CoMFA* tradicional, uma série de compostos ativos é alinhada e posicionada individualmente em uma grade regularmente espaçada. Átomos, ou grupos de teste (*probes*), são dispostos alternadamente nos vértices

da grade para a avaliação da energia de interação entre o *probe* e cada um dos compostos da série de teste. Os resultado da avaliação dos campos de interação estérica e eletrostática são analisados por *PLS* (*Partial Least Squares*), que não é sensível a co-linearidade dos descritores de campo e é capaz de fornecer um modelo estatístico significativo correlacionando os campos moleculares com a variação da atividade biológica dos compostos. Os resultados da análise PLS podem ser apresentados na forma de mapas de contorno tridimensionais, que indicam as regiões do espaço importantes para a atividade biológica. A **Figura 14** apresenta um exemplo dos campos moleculares estéricos e eletrostáticos calculados para uma série de α-(*N*)-piridina carboxaldeído tiossemicarbazonas substituídas. Na abordagem *CoMSIA* (do inglês, *Comparative Molecular Similarity Indices Analysis*), outros campos moleculares são incorporados na análise comparativa, como por exemplo o campo hidrofóbico e o potencial de ligação de hidrogênio, e adicionalmente, as funções de energia de Lennard-Jones e de Coulomb foram modificadas para contornar uma das limitações do CoMFA, que era a definição arbitrária dos valores de corte para os campos utilizados.



A B

Figura 14. Representação esquemática dos mapas de contorno CoMFA, obtidos a partir de uma série de α-(N)-piridina carboxaldeído tiossemicarbazonas substituídas. **A**) Mapas de contorno eletrostático: O poliedro verde representa regiões onde cargas positivas aumentam a atividade, enquanto que o poliedro vermelho representa regiões onde cargas negativas são mais favoráveis para a atividade. **B**) Mapas de contorno estérico: poliedro verde representa regiões onde volume estérico aumenta a atividade biológica. enquanto que o poliedro vermelho representa regiões onde volume estérico diminui a atividade biológica.

Apesar de descreverem relações de estrutura atividade em três dimensões, os modelos de CoMFA (ou CoMSIA) fogem da definição de um farmacóforo,[Wermuth *et al.*, 1998] pois eles são sensíveis apenas a variações estruturais em uma série de compostos, ou

seja, se todos os compostos da série de treinamento apresentam uma mesma característica estrutural esta não será incluída no modelo, mesmo que esta seja fundamental para a atividade.

Uma outra aplicação dos campos de interação molecular para a construção de modelos farmacofóricos é a abordagem disponível no programa *GRID*.[Goodford, 1985] Nesta abordagem, a proteína alvo é colocada em uma caixa com uma grade regularmente espaçada e, então, sondas moleculares (representando átomos ou moléculas de prova) são posicionadas nos vértices desta grade para a definição de pontos de interação favoráveis ("*hot spots*" [Gohlke *et al.*, 2000b]) ao redor da estrutura do receptor biológico. Estes pontos de interação favoráveis representam um modelo farmacofórico identificado a partir da estrutura do alvo biológico.

Existem diversas estratégias para a construção de modelos farmacofóricos independentes da conformação e do alinhamento dos ligantes, como por exemplo: métodos binários (*Fingerprints* [McGregor & Muskal, 1999]), a partir de árvores de propriedades (*feature trees* [Rarey & Dixon, 1998]) e a partir de vetores entre pontos de interação favoráveis extraídos de campos de interação molecular (*Almond* [Patel *et al.*, 2002]). Descritores binários ou *Fingerprints* [McGregor & Muskal, 1999] são conjuntos de farmacóforos gerados ao se listar todas as possíveis combinações de três, ou quatro, propriedades farmacofóricas, em relação a um conjunto de distâncias pré-definidas. O *fingerprint* de uma molécula será a ocorrência, ou não (*0* ou *1*; descritor binário), de cada um dos farmacóforos pré-estabelecidos, considerando-se todas as possíveis conformações da molécula. Uma vez computados os *fingerprints* de todas as moléculas em um banco de dados, estes podem ser usados para realizar buscas muito rápidas (por serem descritores binários) em programas de *virtual screening*.[Karnachi & Kulkarni, 2006]

Árvores de propriedades, ou *feature trees*,[Rarey & Dixon, 1998] também são modelos farmacofóricos independentes da conformação. Neste, os compostos são divididos em unidades menores, em que cada unidade representa uma propriedade diferente (como por exemplo: grupo hidrofóbico, grupo doador/aceptor de ligação de hidrogênio). Estas unidades são conectadas seqüencialmente como em uma árvore de propriedades, obtendo assim uma representação simplificada da estrutura do composto. Estas árvores podem ser alinhadas por similaridade de propriedades. A vantagem desta abordagem é que as árvores

podem ser divididas em porções menores, de três ou quatro propriedades conectadas, e estas unidades podem ser usadas para uma busca em bancos de dados por compostos que apresentem as mesmas seqüências de propriedades. [Rarey *et al.*, 2006]

Por fim, o programa *Almond* [Pastor *et al.*, 2000] [Afzelius *et al.*, 2002] [Patel *et al.*, 2002] também apresenta descritores 3D independentes do alinhamento. Estes descritores são distâncias entre pontos de interação favoráveis extraídos de campos de interação molecular.[Goodford, 1985] Estes descritores podem ser usados para a obtenção de modelos de QSAR-3D que podem ser usados para prever e selecionar compostos ativos.

A obtenção de modelos farmacofóricos ou de QSAR-3D independentemente de se conhecer a estrutura do receptor é uma estratégia de sucesso no caminho para a descoberta de novos compostos ativos. Entretanto, com a crescente disponibilidade de estruturas de receptores biológicos (obtidos por cristalografia de raio-X; por RMN de proteínas e/ou por modelagem por homologia) o uso de informações do alvo para a avaliação da afinidade entre o ligante e a proteína tem se tornado uma estratégia cada vez mais utilizada para a seleção de compostos ativos, apesar das dificuldades ainda existentes, inerentes aos métodos que consideram a estrutura do alvo biológico. A seguir será apresentado o terceiro tipo de filtro aplicado em *virtual screening* que são os filtros de *docking*.

1.4.3. Filtros de *Docking*

Filtros de *docking* são filtros que selecionam compostos baseando-se na estimativa da afinidade entre um ligante e seu receptor biológico. Esta estimativa de afinidade é obtida após o *docking*, ou em português "ancoragem", do composto na correspondente macromolécula. Esta metodologia fornece uma previsão do modo de ligação mais favorável entre um composto e uma determinada proteína alvo, bem como apresenta avaliação quantitativa desta interação. A aplicação desta estratégia para bibliotecas de compostos, permite a seleção de compostos com maior afinidade pela proteína alvo.

O procedimento de *docking* pode ser dividido em duas etapas distintas e que podem ser executadas concomitante ou em separado. Numa primeira etapa, os possíveis modos de ligação (poses) entre o ligante e a proteína são gerados, variando-se tanto a conformação do ligante quanto sua orientação (rotação e translação) em relação à proteína. A segunda etapa é a avaliação quantitativa de cada uma destas possíveis poses, para a seleção da mais

favorável. Na maioria dos casos, estas duas etapas são executadas interativamente, de modo que a avaliação da qualidade da pose orienta a geração da próxima, fazendo com que a procura seja ativamente conduzida na direção da pose mais favorável.

Existem diversos programas de *docking* disponíveis comercialmente ou de acesso livre. Cada programa utiliza estratégias diferentes para a busca e para a avaliação das poses, durante o procedimento de *docking*. Muitos trabalhos foram publicados, nos quais o desempenho destes programas é comparado entre si.[Bissantz *et al.*, 2000] [Kellenberger *et al.*, 2004] [Kontoyianni *et al.*, 2005] [Kontoyianni *et al.*, 2004] [Perola *et al.*, 2004] O que se observa, entretanto, é que, embora alguns programas apresentem melhores resultados que outros em determinadas situações específicas, nenhum deles apresenta desempenho muito superior aos outros quando comparados em situações diversas.[Rester, 2006].

A maior parte dos programas de *docking* considera apenas a flexibilidade do ligante, enquanto que a estrutura da proteína é mantida rígida. As estratégias de amostragem do espaço conformacional do ligante utilizadas por estes programas podem ser agrupadas em quatro categorias:

- Métodos baseados em campos de força. Estes métodos utilizam simulações de dinâmica molecular ou de Monte Carlo para explorar o espaço conformacional do ligante, buscando-se mínimos de energia, a partir de protocolos de *simulated annealing* (arrefecimento simulado). Um dos problemas destes métodos é que os ligantes podem ficar confinados a mínimos locais, caso sejam utilizados tempos de simulação muito curtos. Exemplos de programas que utilizam simulações de Monte Carlo são o *AUTODOCK*[Morris *et al.*, 1998] e o *GLIDE*.[Schulz-Gasch & Stahl, 2003]

- Métodos evolutivos (algoritmo genético). Estes métodos reduzem as probabilidades do ligante ficar confinado em mínimos locais, entretanto, não permitem um ajuste fino da pose. Normalmente, os programas que adotam estes métodos têm melhores desempenhos quando associados à uma etapa adicional de minimização para a localização do mínimo global. Exemplos de programas que utilizam algoritmo genético são o *AUTODOCK*[Morris *et al.*, 1998] e o *GOLD*.[Verdonk *et al.*, 2003]

- Métodos de ancoragem seqüencial. Nestes métodos o ligante é dividido em fragmentos menores que são seqüencialmente posicionadas e minimizados no sítio de

ligação da proteína. Um exemplo de programa que utiliza ancoragem seqüencial é o *FlexX*.[Rarey *et al.*, 1997]

• Métodos de complementaridade de forma. Estes métodos procuram por uma complementaridade estrutural entre as formas do ligante e da correspondente cavidade da proteína. Os programas que se utilizam deste método são o *DOCK*[Ewing & Kuntz, 1997] e o *FRED*.[Schulz-Gasch & Stahl, 2003]

O desempenho dos programas de *docking* é, normalmente, avaliado observando-se o quanto os resultados obtidos reproduzem dados experimentais de cristalografia de raios-X. É consenso que, dentro do erro experimental definido pela resolução do cristal, a pose de um ligante encontrada em um complexo de raios-X é a conformação mais próxima da conformação bioativa, ou seja, a pose mais correta.[Klebe, 2006] Corroboram para esta afirmação evidências experimentais de que certas enzimas permanecem ativas e funcionais mesmo quando na forma cristalina,[Bohm *et al.*, 1996] Entretanto, outros dados experimentais mostram que dependendo das condições utilizadas na cristalização, os ligantes podem assumir modos de ligação diferentes.[Steuber *et al.*, 2006] Em geral, todos os programas de *docking*, empregando qualquer uma das estratégias de busca no espaço conformacional mencionadas acima, são capazes de reproduzir satisfatoriamente (com tolerâncias de até 2Å nos valores de RMSD) as poses dos ligantes presentes nos complexos cristalinos usados nas séries de validação e avaliação de desempenho.[Kellenberger *et al.*, 2004]

Um dos principais fatores limitantes para que a pose correta seja encontrada, é o tempo necessário para encontrá-la. Uma limitação aleatória do tempo gasto na procura da pose mais favorável pode acarretar uma amostragem insuficiente do espaço conformacional do ligante. De fato, na maioria dos casos o desempenho dos programas de *docking* é pior quanto maior for o tamanho da cavidade da proteína ou quanto maior for o número de ligações com rotação livre do ligante. Isto significa dizer que os programas de *docking* têm maiores dificuldades em encontrar a pose correta quanto maior for o número de graus de liberdade disponíveis ao ligante.[Kellenberger *et al.*, 2004] Os programas gastam, em média, entre alguns segundos a alguns minutos por *docking*, por composto, dependendo da complexidade do problema a ser resolvido. Este tempo, apesar de curto, pode tornar inviável o *docking* de bancos de dados muito grandes. Por exemplo, se um procedimento de

BIBLIOTECA
INSTITUTO DE QUÍMICA
Universidade de São Paulo

docking gasta, em média, hum minuto por composto, o processamento de um banco de dados com 100.000 compostos (o que representa um banco de dados relativamente pequeno) levaria mais de dois meses para terminar. Esta escala de tempo é muito grande e, por essa razão, os filtros de *docking* são normalmente os últimos a serem aplicados em programas de *virtual screening*, quando então, o banco de dados já foi reduzido por outros filtros mais rápidos. [Klebe, 2006]

Uma vez encontrada a pose correta do ligante dentro da cavidade da proteína, o programa de *docking* deve ser capaz de identificar esta pose como a mais correta, entre todas as outras poses que foram geradas. As funções utilizadas para avaliar as poses são chamadas de "funções de ranqueamento" (do inglês, *scoring functions*). As funções de ranqueamento dos principais programas de *docking* podem ser agrupadas em três tipos:

- Funções de ranqueamento baseadas em campos de força. Estas funções utilizam a mecânica molecular clássica para apresentar uma estimativa do valor da energia livre de ligação entre o ligante e a proteína. A energia do complexo é calculada a partir de interações eletrostáticas e de *Van der Waals*. A principal limitação desta estratégia é a incapacidade de se considerar, explicitamente, os fatores entrópicos envolvidos na formação do complexo, como por exemplo, a perda de graus de liberdade e a solvatação/dessolvatação do ligante e da proteína, respectivamente. O mesmo problema é encontrado quando são aplicados cálculos semi-empíricos como função de ranqueamento de resultados de *docking*.[Vasilyev & Bliznyuk, 2004] Um exemplo de função de ranqueamento baseadas em campos de força é a *Glidescore*.[Friesner *et al.*, 2004]

- Funções de ranqueamento empíricas. Estas funções calculam o valor da energia livre de ligação a partir da somatória de termos derivados de parâmetros estruturais. A contribuição de cada parâmetro estrutural é definida através do ajuste dos termos da equação a valores de constante de ligação (K_i) de complexos ligante-proteína, obtidos experimentalmente, em uma série de treinamento. O maior problema das funções de ranqueamento empíricas é que estas não são treinadas para prever os valores da energia de ligação, de complexos estruturalmente diferentes daqueles usados na série de treinamento. Exemplos de programas que utilizam funções de ranqueamento empíricas são: *SCORE*[Wang *et al.*, 1998] e *FRED*.[Schulz-Gasch & Stahl, 2003]

• Funções de ranqueamento baseadas em dados de cristalografia. Estas funções, conhecidas como *knowledge-based*,[Gohlke *et al.*, 2000a] consideram o valor da afinidade de ligação como sendo a somatória das interações, átomo-a-átomo, entre o ligante e a proteína. Os valores do potencial de interação, átomo-a-átomo, para diferentes pares de átomos são obtidos a partir de dados de cristalografia de raio-X (extraídos do *PDB* ou do CCDC), sendo que estes potenciais são convertidos em valores de energia livre, dependentes da distância. Neste caso, a definição arbitrária dos tipos de átomos envolvidos na interação é um fator que limita a aplicabilidade destas funções de ranqueamento. Exemplos de programas que utilizam funções de ranqueamento *knowledge-based* são: *GOLDSCORE*[Verdonk *et al.*, 2003] e *SURFLEX*.[Jain, 2007]

Em geral as funções de ranqueamento apresentam uma tendência de classificar melhor tanto os compostos com maior peso molecular como os com menos ligações com rotação livre.[Kubinyi, 2006; Kitchen *et al.*, 2004] Compostos grandes podem formar muitas interações possíveis com o receptor e, portanto apresentam, em média, maiores valores de afinidade preditos do que os observados para compostos pequenos. Por outro lado, os termos de penalidade entrópica, adicionados para compensar a perda de mobilidade de ligações com rotação livre do ligante ao formar o complexo, são maiores quanto mais ligações com rotação livre o composto apresenta. Como resultado, compostos flexíveis tendem a ser mais penalizados do que compostos rígidos e, portanto, são pior classificados.

Estudos comparativos mostram que todas as funções de ranqueamento têm dificuldades para identificar a pose mais correta, dentre todas as poses geradas. Kellenberger[Kellenberger *et al.*, 2004] preparou uma série de teste com 100 complexos proteína-ligante e comparou o desempenho de oito programas de *docking*, respectivamente, *DOCK*, *FLEXX*, *FRED*,*GLIDE*, *GOLD*, *SLIDE*, *SURFLEX* e *QXP*. Considerando apenas a pose melhor ranqueada para cada um dos 100 complexos, e considerando corretas apenas as posses com valores de RMSD $\leq$2Å em relação a pose no cristal, os programas tiveram taxas de acerto variando entre 20 e 55%. Ou seja, o melhor programa de *docking* acertou a pose de apenas um, em cada dois complexos, da série testada. Quando consideradas todas as poses geradas pelos programas (e não apenas a melhor ranqueada), para cada um dos complexos da série de teste, as taxas de acerto subiram para 50 a 90%.

Kellenberger[Kellenberger *et al.*, 2004] verificou também, a utilidade destes programas em um ensaio de *virtual screening*. A estrutura de raio-x da Timidina quinase (TK) do vírus *Herpes simplex* foi usada como alvo para o *docking* de um banco de dados com 990 compostos *drug-like* e com 10 inibidores conhecidos da TK. Considerando-se apenas os 25 compostos melhores ranqueados por cada programa, o número de inibidores da TK identificados variou entre 0 e 4, para os oito programas testados. Já entre os 100 compostos melhores ranqueados por cada programa, o número de inibidores da TK identificados variou entre 0 a 10 (apenas um dos programas testados não foi capaz de encontrar nenhum dos inibidores de TK presentes no banco de dados). Estes resultados mostram que, em geral, os programas de *docking* são sensíveis e capazes de identificar compostos ativos para determinado alvo biológico, entretanto apresentam uma baixa seletividade e tem dificuldade de separar compostos realmente ativos de resultados falso-positivos.[Parker & Bajorath, 2006]

Uma estratégia utilizada para reduzir do número de resultados falso-positivos é a utilização de mais de uma função de ranqueamento para avaliar a mesma pose.[Parker & Bajorath, 2006] Esta estratégia, chamada de *consensus docking,* se baseia no princípio de que qualquer função de ranqueamento pode ser combinada com qualquer estratégia de busca.[Rester, 2006] Muitos trabalhos publicados já mostraram a superioridade do *consensus docking*, quando comparado com o *docking* com uma única função de ranqueamento,[Charifson *et al.*, 1999] [Clark *et al.*, 2002] [Paul & Rognan, 2002] [Wang & Wang, 2001] [Jacobsson *et al.*, 2003] Esta estratégia, entretanto, só será realmente efetiva quando todas as funções de ranqueamento, usadas no consenso, apresentarem desempenhos similares. Considerando-se o exemplo de Kellenberger[Kellenberger *et al.*, 2004] descrito anteriormente, o consenso entre os resultados da função de ranqueamento que encontrou dez entre os dez inibidores de TK (100%) e os resultados da função que não encontrou nenhum (0%), é pior do que usar apenas os resultados da função que encontrou todos os inibidores.

Em um outro estudo comparativo, Wang e seus colaboradores [Wang *et al.*, 2004b] compilaram um banco de dados (*PDB*bind [Wang *et al.*, 2004a]) com 800 complexos ligante-proteína, extraídos do *PDB*, e com valores de K_i conhecidos. Os autores utilizaram 14 diferentes funções de ranqueamento, para avaliar os complexos e correlacionaram os

dados previstos de afinidade, extraídos das funções de ranqueamento, com os correspondentes dados experimentais de K_i. O que foi observado é que nenhuma das funções de ranqueamento foi capaz de prever corretamente valores de K_i para a séries de teste. Os valores do desvio padrão de todas as funções variaram entre 1,8 M e 2,2 M unidades de K_i (nota-se que o K_i é apresentado em uma escala log). Uma das explicações para este resultado é que a precisão do valor previsto nunca pode ser melhor do que a precisão com que os valores experimentais são determinados. Segundo Klebe, [Klebe, 2006] coleções de dados experimentais determinados por diferentes grupos de pesquisa, como o banco de dados *PDB*bind,[Wang *et al.*, 2004a] utilizado no teste, podem apresentar desvios de 0,7 a 1,5 unidades de log nos valores de K_i. Esta faixa de desvio é apenas um pouco melhor do que os valores de desvio encontrados entre as melhores funções de ranqueamento, usadas no teste realizado por Wang.[Wang *et al.*, 2004b] Uma outra explicação para a falta de precisão das funções de ranqueamento é o desconhecimento de todas as forças envolvidas na interação ligante-receptor ou, ainda, a incapacidade dos descritores usados descreverem completamente esta interação. Diferentes aspectos sobre a natureza das forças envolvidas na interação ligante-receptor, serão discutidos mais adiante nesta tese (Item 1.5.).

Apesar de todos os problemas e limitações associados às funções de ranqueamento, o desempenho destas funções ainda é considerada satisfatória para a busca de estruturas líder, em bancos de dados de compostos.[Klebe, 2006; Kitchen *et al.*, 2004] Isto porque estas funções são usadas para enriquecer bancos de dados de compostos e não para o cálculo preciso dos correspondentes valores das constantes de ligação K_i. O sucesso na aplicação de um filtro de *docking,* independe da posição relativa em que um composto ativo é classificado pela função de ranqueamento, desde que o composto ativo esteja entre os compostos mais bem ranqueados e, portanto, entre os escolhidos para serem testados experimentalmente.[Kitchen *et al.*, 2004] O número de compostos que devem ser experimentalmente testados em um projeto envolvendo a estratégia de *docking* é uma decisão subjetiva e arbitrária, que depende de fatores externos como a disponibilidade dos compostos e a disponibilidade dos recursos materiais e de tempo para a realização dos testes. Em média, o número de estruturas líder identificadas (com valores de atividade na faixa de micromolar e submicromolar) em projetos de *virtual screening* , que utilizam

procedimentos de *docking*, variam entre 1 a 10% dos compostos identificados pelo modelo.[Rester, 2006] Um exemplo típico de sucesso na aplicação do *docking* é observado na busca de inibidores da Tirosina Kinase (TK), uma enzima envolvida no tratamento da leucemia.[Peng *et al.*, 2003] Um banco de dados de 200.000 compostos foi docado sobre o sítio ativo da TK, usando-se o programa *DOCK*.[Ewing & Kuntz, 1997] Os 1000 compostos melhor ranqueados, foram agrupados por similaridade molecular e filtrados segundo a "regra dos 5" de Lipinski.[Lipinski *et al.*, 1997] Quinze (15) compostos foram selecionados para o teste de atividade biológica, dentre os quais oito inibiram o crescimento de células tumorais, com valores de IC_{50} entre 10 e 200μM.

Além de um método de busca no espaço conformacional e de uma função de ranqueamento, um outro pré-requisito para o procedimento de *docking* é a identificação do local em que o ligante vai interagir com o receptor.[Hendlich *et al.*, 1997] Estes sítios de ligação são, normalmente, localizados em depressões ou em cavidades na superfície da proteína. A afinidade e a especificidade do ligante, em relação ao receptor, são dependentes das características desta cavidade e, por isso, a identificação destas cavidades é uma etapa fundamental para qualquer estratégia baseada na estrutura do receptor.[Sotriffer & Klebe, 2002] Uma das melhores maneiras de se identificar o sítio de ligação do ligante é localizar no respectivo complexo cristalino, a cavidade em que ele está posicionado.

Quando não se dispõem de uma estrutura cristalina (ou por RMN) do ligante no complexo e o sitio de ligação é desconhecido, é possível identificar cavidades de ligação, a partir do reconhecimento da geometria da superfície do receptor. Existem programas capazes de identificar automaticamente possíveis cavidades de ligação na superfície de proteínas. Um desses programas é o *LIGSITE*.[Hendlich *et al.*, 1997] Neste programa, o reconhecimento das cavidades é feito posicionando-se a proteína em uma grade regularmente espaçada. A seguir, procurando-se ao longo dos três eixos cartesianos e das quatro diagonais cúbicas, são identificados todos os pontos da grade que se encontram entre dois átomos da proteína. Os pontos da grade, que estão localizados entre dois átomos da proteína, em pelo menos dois eixos ou em duas diagonais, são considerados como pertencentes a uma cavidade. Os pontos identificados como cavidades, que são adjacentes a outros pontos em cavidades, são unidos para formar uma única cavidade. As cavidades que não apresentam um número mínimo de pontos são descartadas.

Uma vez localizadas as cavidades em um determinado receptor é preciso, a seguir, avaliar se estas são capazes de acomodar uma molécula pequena (*drug-like*).[Hajduk *et al.*, 2005b] Adicionalmente, em alguns casos, o receptor pode estar envolvido em interações proteína-proteína e o seu sítio de ligação é muito raso, ou muito grande, para acomodar com alta afinidade e com especificidade, uma molécula pequena. Nestes casos, ou seja, quando o receptor não pode ter sua atividade modulada por um ligante *drug-like*, o alvo é considerado não-*druggable* e não serve para ser usado em procedimentos de *docking*, nem tampouco para qualquer outra estratégia de desenvolvimento de ligantes que possam modular sua atividade.[Hopkins & Groom, 2002]

Para que uma cavidade seja *druggable*, ou em outras palavras, capaz de acomodar uma molécula pequena, ela deve que apresentar valores de superfície de contato polar menores do que 75Å^2, além de índices de rugosidade, complexidade e compactação adequados.[Hajduk *et al.*, 2005a] Entretanto, estes critérios ainda são, ainda, rudimentares o que reflete a limitada capacidade de se prever quando uma cavidade é capaz de acomodar um ligante.[Klebe, 2006]

Uma alternativa para se avaliar se uma cavidade é *druggable*, considera comparativamente cavidades de proteínas similares ou que apresentam a mesma função e que apresentem um ligante complexado. A análise comparativa de cavidades de proteínas pode ser feita pelo método tradicional, baseando-se na similaridade da seqüência de aminoácidos e no padrão de empacotamento da proteína. Entretanto, a comparação de cavidades também pode ser feita independentemente da seqüência e do padrão de empacotamento, comparando-se apenas as propriedades físico-químicas das cavidades. O programa que possibilita a aplicação desta abordagem é o *CAVBASE*.[Schmitt *et al.*, 2001] Este programa utiliza um banco de dados com todas as cavidades de todas as estruturas depositadas no *PDB* (as cavidades são identificadas pelo programa *LIGSITE*). As propriedades físico-químicas dos resíduos que compõem a cavidade e que são expostas na superfície da cavidade, são transformadas em "pseudocentros". Estes pseudocentros são classificados de acordo com cinco propriedades de reconhecimento molecular, a saber: doador de ligação de hidrogênio; aceptor de ligação de hidrogênio; doador/aceptor de ligação de hidrogênio, hidrofóbico aromático e hidrofóbico alifático. O algoritmo de busca do programa é capaz de procurar cavidades similares, em seu banco de dados, com base na

similaridade dos pseudocentros e no formato da cada cavidade. O reconhecimento de cavidades similares pode contribuir para identificação de sítios de ligação em proteínas que ainda não tiveram sua função ou sítio ativo definido.

Um outro ponto crítico em um procedimento de *docking* é a escolha da conformação da proteína e, conseqüentemente, a conformação do sítio de ligação. Proteínas são macromoléculas flexíveis com muitos graus de liberdade e que podem adotar muitos conformações isoenergéticas.[Verli & Barreiro, 2005] Adicionalmente, está demonstrado que os sítios de ligação podem se "abrir" ou se "fechar" para o ligante, ou, ainda, podem sofrer modificações estruturais induzidas pela própria acomodação do ligante (*induced fit* [Murray *et al.*, 1999] [Sotriffer *et al.*, 2004] [Davis & Teague, 1999] [Verli & Barreiro, 2005]). Um exemplo deste ajuste induzido é a variação da estrutura da HIV protease (**Figura 15**). Esta, quando cristalizada com um ligante, se apresenta em sua conformação "fechada" (*PDB* 1HVR) enquanto que sem o ligante se apresenta em sua conformação "aberta" (*PDB* 1HHP).

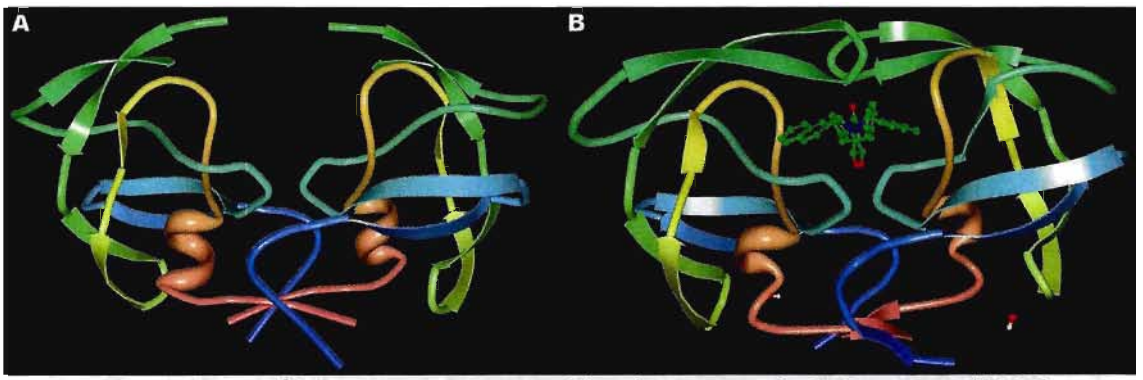


Figura 15. (A) Representação esquemática da estrutura da HIV protease (1HHP) cristalizada sem o ligante em sua conformação "aberta"; (B) Representação esquemática da estrutura da HIV protease (1HVR) cristalizada com um ligante em sua conformação "fechada"

Os modelos de *docking* que consideram a proteína rígida não são capazes de considerar qualquer flexibilidade ou mudança conformacional da proteína induzida pelo ligante. Alguns programas de *docking* são capazes de considerar alguma flexibilidade na proteína, como movimentos da cadeia lateral de resíduos de aminoácidos posicionados no sítio ativo[Sherman *et al.*, 2006] ou a rotação de grupamentos hidroxila envolvidos em ligações de hidrogênio. [Verdonk *et al.*, 2003] Entretanto, nenhum programa de *docking* é, ainda, capaz de considerar grandes mudanças conformacionais na cadeia principal da

proteína. Uma alternativa para esta limitação é a busca por um conjunto de conformações possíveis da proteína. Estas diferentes conformações podem ser obtidas através de buscas no espaço conformacional da proteína ou, simplesmente, analisando-se múltiplas estruturas cristalinas com diferentes ligantes complexados.[Murray *et al.*, 1999] [Sotriffer *et al.*, 2004] A seguir, uma ou mais destas conformações podem ser escolhidas para o *docking* com a proteína rígida. A escolha de qual, ou quais, conformações devem ser usadas no procedimento de *docking* depende do papel de cada conformação na função da proteína. [Klebe, 2006] Por exemplo, se a variabilidade conformacional no sítio de ligação da proteína ocorre para que esta possa acomodar uma variedade de ligantes com estruturas diversas, diferentes conformações devem, então, ser consideradas no procedimento de *docking*. Entretanto, se a mudança conformacional da proteína for um reflexo de sua condição "ativa" ou "inativa", então, apenas uma das conformações deve ser usada.

Uma outra maneira de se considerar a flexibilidade da proteína é feita através da utilização de protocolos de Dinâmica Molecular ou de Monte Carlo (DM/MC) para analisar o complexo ligante-receptor. [Hansson *et al.*, 2002] [Sotriffer, 2005] [Karplus & PETSKO, 1990] A DM/MC considera a flexibilidade total, tanto do ligante quanto da proteína e, portanto, é sensível ao ajuste induzido provocado pela formação do complexo. Entretanto, os protocolos de DM/MC são muito demorados e não permitem que esta metodologia seja aplicada em bancos de dados com muitas moléculas. Em média, uma simulação de DM/MC pode levar mais de uma semana por composto, mesmo usando *clusters* de computadores para acelerar o cálculo.[Hansson *et al.*, 2002]

A presença de moléculas de água no sítio ativo é um outro aspecto que deve ser considerado durante a etapa de preparação da proteína para o *docking*. Uma análise das estruturas cristalinas do banco de dados *PDB* revelou que, em aproximadamente dois terços dos complexos examinados, uma molécula de água está envolvida na ligação, freqüentemente mediando o contato entre a proteína e o ligante. [Klebe, 2006] Portanto, qualquer abordagem baseada na previsão do modo de ligação de moléculas candidatas deve levar em consideração, explicitamente, as moléculas de água que participam efetivamente nesta ligação. Uma das estratégias utilizadas, considera as moléculas de água importantes para a ligação como parte da estrutura do receptor. A definição da posição de moléculas de água conservadas pode ser feita através da análise de dados de cristalografia de raios-X de

cavidades similares. Na ausência de dados experimentais relativos à posição de moléculas de água conservadas, utiliza-se, alternativamente, programas capazes de prever a posição de moléculas de água no sítio ativo. Um destes programas é o já mencionado programa *GRID*.[Goodford, 1985] Neste caso, a proteína é colocada em uma grade regularmente espaçada e uma molécula de água é posicionada, em cada ponto da grade, para a avaliação da energia de interação. Os pontos de mínimo de energia serão os pontos mais favoráveis para uma molécula de água e estas podem ser incluídas na estrutura do receptor.

Um outro problema associado à preparação do sítio de ligação para o procedimento de *docking* é a resolução dos complexos cristalinos utilizados na determinação estrutural. Em média, os valores da resolução de uma estrutura cristalina depositada no *PDB* é de 2,18 ± 1,3Å [*PDB* web em 30-Nov-07] Com este nível de resolução não podem ser observados átomos de hidrogênio e, portanto, não é possível determinar nem a orientação de hidrogênios de hidroxilas envolvidos em ligações de pontes de hidrogênio, nem definir o estado de protonação de resíduos ionizáveis. Adicionalmente, não se pode diferenciar entre átomos de oxigênio, de nitrogênio e de carbono e, portanto, não é possível determinar a orientação correta da cadeia lateral de resíduos como a treonina, a asparagina, a glutamina e a histidina. Todas estas limitações geram incertezas quanto a definição da estrutura do sítio ativo e juntamente com os demais problemas citados anteriormente nesta tese (como por exemplo a definição da cavidade, a flexibilidade da proteína, o ajuste induzido e a presença de moléculas de água), impedem que o procedimento de *docking* seja totalmente automatizado, fazendo-se sempre necessário a intervenção do operador para avaliar criticamente a qualidade, a validade química e a utilidade do modelo estrutural do receptor biológico.[Kubinyi, 2006]

Apesar do *docking* ser uma estratégia baseada no uso do computador (*Computer Aided Drug Design - CADD*), a intervenção do operador é, ainda, fundamental para o seu sucesso, tanto na etapa de preparação do *docking*, quanto na interpretação dos resultados. A inspeção visual dos resultados é uma etapa crítica no procedimento de *docking* uma vez que as funções de ranqueamento ainda são rudimentares e não superam o conhecimento e experiência química de um pesquisador envolvido com os aspectos multifatoriais da interação entre um ligante e um alvo biológico.[Klebe, 2006; Sotriffer *et al.*, 2004; Kitchen *et al.*, 2004; Gruneberg *et al.*, 2001] Em um exemplo de sucesso da aplicação do

procedimento de *docking* Klebe e colaboradores[Gruneberg *et al.*, 2002; Gruneberg *et al.*, 2001] partiram de um banco de dados *Maybridge*[Maybridge Chemicals Ltd.,] com 90.000 compostos visando a procura de inibidores da anidrase carbônica humana (uma metalo enzima que catalisa a conversão de CO_2 em HCO_3^-). Inicialmente, os autores aplicaram filtros físico-químicos e farmacofóricos para reduzir o banco de dados inicial para aproximadamente 3.000 (3%). Estes compostos foram submetidos ao *docking* flexível (programa FlexX) e os 100 melhores resultados ranqueados pelo programa foram visualmente inspecionados. Treze compostos foram escolhidos para o teste experimental de inibição da anidrase carbônica. Destes treze compostos, três foram identificados como inibidores picomolares, um nanomolar e sete micromolares. A pose de ligação de dois destes compostos foi confirmada experimentalmente por cristalografia de raios-X.

O procedimento de *virtual screening* utilizando filtros de *docking*, associado com outros filtros, já foi aplicado com sucesso a diversos sistemas. Em uma extensa revisão feita em 2006,[Kubinyi, 2006] Kubinyi identificou mais de 50 diferentes alvos moleculares em que o procedimento de *docking* foi aplicado e nos quais estruturas líder com valores de atividade micromolar e submicromolar foram identificadas. A **Tabela 2** apresenta uma lista completa dos alvos moleculares agrupadas por tipo.

Tabela 2. Alvos moleculares submetidos ao procedimento de *docking*[Kubinyi, 2006]

Tipos	Alvos
Receptores acoplados a Proteína-G	Receptor a1A, receptor de dopamina D3, receptor de endotelina A (ETA), receptor de hormônio concentrador de melanina tipo 1, receptor muscarínico M3, receptor de neuroquinina-1, receptor de neuropeptídeo Y tipo 5, receptor purinérgico A2A, receptor de urotensina II (GPR14)
Receptores nucleares	Receptor de ácido retinóico, receptor de hormônio da tireóide
Quiases	Akt 1, tirosina quinase Bcr-abl, kinase 1 de ponto de controle, quinase 2 dependente de ciclina, quinase 4 dependente de ciclina, quinase glicogênio sintetase, tirosina quinase p56 de célula-T linfóide, proteína quinase CQU2, quinase de receptor fator de crescimento b
Proteases	Catepsina D, falcipaína-2, HIV protease, plasmepsina II, proteinase de síndrome respiratória aguda severa CoV tipo-3C, trombina
Outras hidrolases	Acetilcolinesterase, adenilato ciclase, b-lactamase AmpC, fosfodiesterase 4, fosfatase 1B de tirosina
Oxidases e redutases	Aldose redutase, dihidrofolato redutase, inibidores de inosina 5-monofosfato desidrogenase
Enzimas diversas	Ribonucleotídeo 5-aminoimidazol-4-carboxamida transformilase, anidrase carbônica II, DNA girase, dTDP-6-desoxi-D-xilo-4-hexulose 3,5-epimerase, farnesil transferase, guanina fosforibosil transferase, HIV-1 integrase, tRNA-guanina transglicosilase
Canal iônico	Canal seletivo de Ca^{2+} tipo-T, canal de potássio Kv1.5, canal de potássio shaker
Interfaces Proteína–	Interação proteína–proteína dos complexos Bcl-2, ciclofilina A, proteína de

proteína,	ligação-FK506, proliferação celular mesangial, interação proteína–proteína Rac1, VLA-4
Interações proteína–RNA	Elemento responsivo de transativação de HIV-1 RNA

Outras aplicações do procedimento de *docking* incluem o alinhamento de séries de compostos para o uso em estratégias baseadas na estrutura do ligante (*CoMFA, CoMSIA*);[Moro *et al.*, 2005; Zhou & Madura, 2004] a identificação de resultados falso positivos em estudos de *HTS*[Jenkins *et al.*, 2003] e a análise de produtos metabólicos de substratos da *P450*[Venhorst *et al.*, 2003].

1.5. Descritores de interação molecular

Fundamentos Teóricos

Para que se possa descrever e prever a interação entre um ligante e seu receptor é preciso conhecer a grandeza e natureza das forças que determinam esta interação. Na formação de um complexo não-covalente [RL], assumindo-se que as condições de equilíbrio termodinâmico sejam alcançadas, a afinidade entre um ligante e seu receptor pode ser descrita pela constante de equilíbrio (K_i):

$$R + L \rightleftharpoons RL$$

onde,
R é o receptor
L é o ligante
RL é o complexo

Sendo que

$$K_i = [RL] / [R]\,[L]$$

onde,
K_i é o valor da constante de equilíbrio (ou constante de afinidade)
$[RL]$ é o valor da concentração do Complexo
$[R]$ é o valor da concentração do receptor (sítios ativos)
$[L]$ é o valor da concentração do ligante

Considerando-se que o valor do logaritmo da constante de afinidade ($\log K_i$) se correlaciona linearmente com o valor da variação da energia livre de *Gibbs* do processo de formação do complexo *RL*, pode-se escrever

$$\Delta G = -2{,}3RT \log K$$

onde, ΔG representa a variação do valor da energia livre da formação do complexo ligante-receptor;

R é o valor da constante universal dos gases
T é o valor da temperatura (em graus Kelvin)
logK é o valor da constante de equilíbrio da interação entre o ligante e o receptor

Os valores experimentais de K_i variam tipicamente entre 10^{-2} a 10^{-12} M, correspondendo à valores de energia de *Gibbs* entre -10 e -70 kJ/mol, a 25°C.[Bohm, 2003]

Por outro lado, a energia livre de *Gibbs,* envolvida no processo de formação do complexo [RL], pode ser decomposta em duas componentes, representando as contribuições, respectivamente, entrópicas (ΔS) e entálpicas (ΔH) podendo, então, ser expressa por:

$$\Delta G = \Delta H - T\Delta S$$

Sendo assim, a formação do complexo ligante-receptor pode ser explicada por variações entrópicas e entálpicas. Os valores da contribuição relativa de cada componente da energia livre de *Gibbs* podem ser obtidos diretamente por experimentos de microcalorimetria.[Weber *et al.*, 1992] A análise dos resultados destes experimentos mostra que o mecanismo de formação do complexo ligante receptor pode ser ora determinado por efeitos entrópicos ora por efeitos entálpicos.[Bohm & Klebe, 1996] Por exemplo, na ligação da biotina (**Figura 16**) com a streptavidina (uma proteína da bactéria *Streptomyces avidinii*) o valor da componente entálpica é maior que o observado para a componente entrópica, enquanto que na ligação do HABA (Ácido 2,4-hidroxifenilazo benzóico) com a mesma proteína (**Figura 16**) o valor da componente entrópica é maior do que o observado para a componente entálpica. A **Figura 16** mostra uma representação esquemática das principais interações entre os ligantes e a proteína, enquanto que na **Tabela 3** são apresentados os valores experimentais dos parâmetros termodinâmicos, respectivamente, K_i, ΔG, ΔH e $T\Delta S$ determinados por microcalorimetria, para a formação do complexo com a strptavidina-ligante.[Weber *et al.*, 1992]

Tabela 3. Valores dos parâmetros termodinâmicos K_b, ΔG^0, ΔH^0 e $T\Delta S^0$ para a formação do complexo com a streptavidina-ligante, determinados por microcalorimetria[Weber *et al.*, 1992]

Ligante	K_b (M^{-1})	ΔG^0 (kcal/mol)	ΔH^0 (kcal/mol)	$T\Delta S^0$ (kcal/mol)
HABA	7295 (±554)	-5,27 (±,05)	1,70 (±,011)	6,97 (±,04)
d-biotina	$2,5 \times 10^{13}$	-18,3	-32,0	-13,7

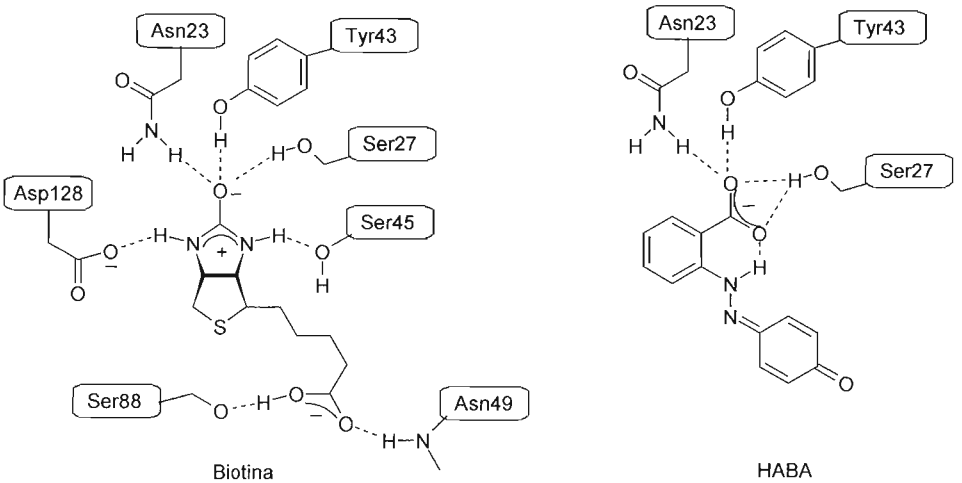


Figura 16. Representação esquemática da estrutura da Biotina e do HABA. Estão indicadas as interações com os resíduos do sítio de ligação da streptavidina[Weber *et al.*, 1992]

Durante a formação do complexo *RL* as contribuições entrópicas e entálpicas podem ser compensadas mutuamente, isto é, o aumento da contribuição de um dos termos pode provocar a diminuição da contribuição do outro. Um exemplo deste efeito foi descrito por Klebe[Klebe, 2006] e está esquematicamente representado na **Figura 17**. Para dois inibidores da trombina (uma serino protease humana envolvida na cascata de coagulação), apresentando diferenças apenas no tamanho do anel terminal (grupamentos ciclopentila ou ciclohexila), foram observados valores equivalentes de ΔG. Entretanto, observa-se que as contribuições relativas dos termos, respectivamente, ΔH e $-T\Delta S$ são diferentes, para cada inibidor. Uma explicação para esta diferença é que o ligante contendo o grupo ciclohexila terminal não se ajusta tão bem à cavidade do sítio ativo quanto o ligante contendo o ciclopentila, provocando uma diminuição da contribuição entálpica. Entretanto, este pior ajuste do ciclohexila proporciona uma maior mobilidade do ligante dentro da cavidade e, conseqüentemente, uma menor penalidade entrópica quando comparada ao ciclopentila, que se encontra firmemente ligado ao sitio ativo. Estes resultados foram validados por dados de cristalografia de raios-X.[Klebe, 2006]

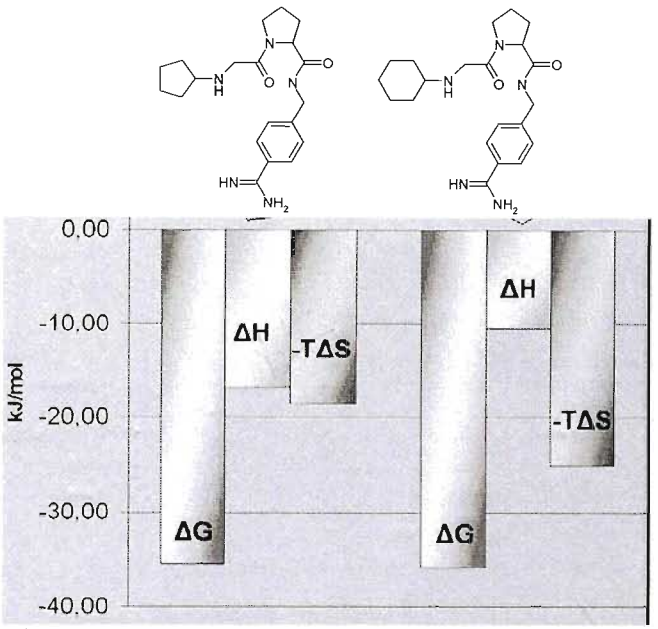


Figura 17. Representação esquemática da estrutura de dois inibidores da trombina; e gráfico dos valores dos parâmetros termodinâmicos ΔG, ΔH e $-T\Delta S$ para a formação do complexo com a trombina-ligante, determinados por microcalorimetria[Klebe, 2006]

Os valores dos termos, respectivamente, ΔH e $-T\Delta S$ observados nos experimentos de microcalorimetria são o resultado líquido de diversas interações, de diferentes naturezas, que ocorrem durante o processo de formação do complexo ligante-receptor. A **Figura 18** apresenta um esquema das diferentes contribuições respectivamente, entrópicas e entálpicas que podem ocorrer durante o processo de formação do complexo [RL].[Andrews *et al.*, 1984]

Na **Figura 18** observa-se que, antes da formação do complexo, o ligante e o receptor encontram-se completamente solvatados. As moléculas de água do solvente promovem ligações de hidrogênio com os grupamentos polares tanto do ligante quanto do receptor, enquanto que as superfícies hidrofóbicas de ambos são cobertas por moléculas de água ordenadas. Ao se formar o complexo, o ligante perde parte de sua esfera de solvatação vindo a ocupar as posições das moléculas de água que estavam solvatando o receptor. O processo de desolvatação envolve contribuições de naturezas entálpicas ($\Delta H_{L\text{-}A}$; $\Delta H_{R\text{-}A}$; $\Delta H_{L\text{-}R}$) e entrópicas (ΔS_A), respectivamente.

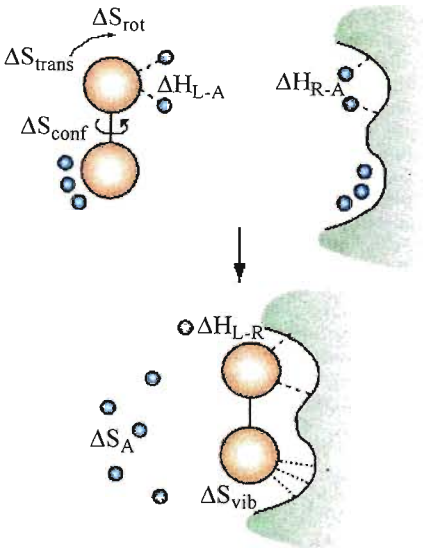
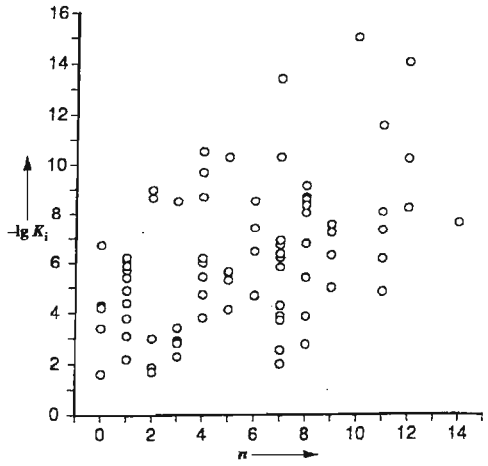


Figura 18. Representação esquemática das contribuições entrópicas e entálpicas que ocorrem durante o processo de formação do complexo [RL][Andrews *et al.*, 1984] Moléculas de água estão representadas por círculos azuis, o ligante está representado pelas esferas beges e o receptor está representado pela superfície verde. $\Delta H_{L\text{-}A}$; $\Delta H_{R\text{-}A}$; representam as entalpias de desolvatação do ligante e do receptor, respectivamente; $\Delta H_{L\text{-}R}$ representa a contribuição entálpica das interações ligante-receptor; ΔS_A representa a variação da entropia resultante da desolvatação das superfícies hidrofóbicas; ΔS_{rot}, ΔS_{trans} e ΔS_{conf} correspondem a variação da entropia rotacional, translacional e conformacional, respectivamente

As ligações de hidrogênio entre grupamentos polares do ligante, e/ou do receptor, com as moléculas de água do solvente são quebradas e novas ligações entre o ligante e o receptor são formadas. Para que este processo seja favorável o termo entálpico relativo às ligações de hidrogênio entre o ligante e o receptor deve ser maior que o termo entálpico relativo às ligações de hidrogênio com as moléculas de água antes da formação do complexo (*i.e.* $\Delta H > 0$). Quando isto não ocorre, a contribuição da adição de ligações de hidrogênio para a afinidade do complexo vai ser próxima de zero. De fato, Böhm e Klebe[Bohm & Klebe, 1996] mostraram que existe correlação não-significativa n=80, r^2=0,21 entre o número de ligações de hidrogênio e valores da constante de afinidade, em uma série de 80 complexos ligante-receptor obtidos por cristalografia de raio-X. (**Figura 19**)

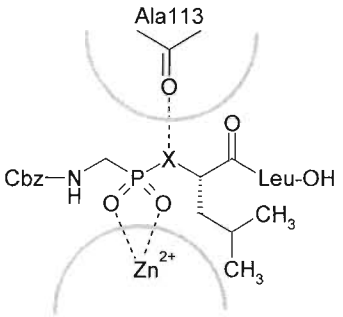


n=80, r^2=0,21

Figura 19. Gráfico da afinidade do complexo ligante-receptor (*-lnK$_i$*) em função do número (*n*) de ligações de hidrogênio entre o ligante e o receptor. Dados extraídos de 80 complexos [RL] depositados no *PDB*.[Bohm & Klebe, 1996]

Por outro lado, a perda de ligações de hidrogênio, sem o correspondente pareamento de grupamentos polares no complexo, é entalpicamente desfavorável para a ligação. Um exemplo deste efeito pode ser constatado observando-se os valores de K_i de inibidores da termolisina (uma metaloprotease da bactéria *Bacillus thermoproteolyticus*) contendo fosfonamidas, fosfonatos e fosfinatos. A **Figura 20** apresenta os valores de K_i e o modo de ligação dos inibidores da termolisina. No complexo com a termolisina, o -NH da fosfonamida forma uma ligação de hidrogênio com a carbonila da Ala 113. Entretanto, quando em solução, este mesmo -NH forma também uma ligação de hidrogênio com uma molécula do solvente, ou seja, não existe variação no número de ligações de hidrogênio formadas pelo ligante. A mesma situação ocorre quando o -NH é substituído por um grupo $-CH_2-$. O grupo metileno do fosfinato não forma ligações de hidrogênio nem com o solvente nem com o receptor, mantendo-se assim, como no caso da fosfonamida, o mesmo número de ligações de hidrogênio formadas pelo ligante. Uma vez que todas as outras interações são preservadas, os valores de K_i da fosfonamida e do fosfinato são equivalentes. Entretanto, no caso do fosfonato, o grupamento -O- forma uma ligação de hidrogênio com uma molécula do solvente, mas não com o receptor, ou seja, o número de ligações de hidrogênio formadas pelo ligante diminui ao se formar o complexo. Como resultado, constata-se a diminuição da afinidade do fosfonato em 1000 vezes, como observado na **Figura 20**. Conclui-se, então, que a formação ou não de ligações de hidrogênio não tem

correlação significativa com a afinidade entre o ligante e o receptor, mas a perda de uma ligação de hidrogênio ao se formar o complexo é, entretanto, altamente desfavorável e pode impedir o reconhecimento molecular entre o ligante e o receptor.



X	K_i
-NH-	0,01
$-CH_2-$	0,01
-O-	9

Figura 20. Representação esquemática dos ligantes da termolisina. Estão indicadas as interações entre os ligantes e a proteína e os valores de K_i.

A desolvatação das superfícies hidrofóbicas do ligante e do receptor também contribui para a formação do complexo. As moléculas de água ordenadas sobre as superfícies hidrofóbicas ao serem liberadas para o solvente promovem um aumento da entropia do sistema, favorecendo a formação do complexo. Este efeito é conhecido como efeito hidrofóbico e é, reconhecidamente, um dos principais responsáveis pelo aumento da afinidade entre o ligante e o receptor. Böhm e Klebe[Bohm & Klebe, 1996] demonstraram que quanto maior o contato entre as superfícies hidrofóbicas do ligante e do receptor, maior a afinidade do complexo. Tipicamente, a afinidade de ligação varia em relação a área de contato hidrofóbico com valores entre 80-200 J/(mol $Å^2$).[Bohm, 2003] Além do aumento da entropia, o efeito hidrofóbico também apresenta uma componente entálpica resultante das forças de dispersão (*Van der Waals*) do contato entre as superfícies apolares do ligante e do receptor.

Assim, estudos de inibição da *MUP* (*Mouse Urine Protein*- uma proteína abundante encontrada na urina do rato) com dois ferormônios realizados através de titulação microcalorimétrica [Homans, 2007] mostraram haver uma contribuição entálpica maior que a contribuição entrópica do contato hidrofóbico entre o ligante e a proteína. O autor atribui este fato inesperado a uma hidratação sub-ótima da cavidade de ligação da *MUP*, devido a sua natureza hidrofóbica e sua forma tubular.

Na **Figura 18**, os termos, respectivamente, ΔS_{rot}, ΔS_{trans} e ΔS_{conf} correspondem aos graus de liberdade rotacional, translacional e conformacional, respectivamente, do ligante

livre no solvente. Ao formar o complexo com o receptor, o ligante perde parte destes graus de liberdade resultando em uma diminuição da entropia do sistema.

Se dois ligantes, um flexível e o outro rígido se ligam ao receptor com o mesmo padrão de interações polares e hidrofóbicas, o ligante rígido irá apresentar maior afinidade pelo receptor do que o ligante flexível. Isto ocorre pois o ligante rígido não irá sofrer a mesma perda de liberdade conformacional que sofre o ligante flexível. Existem muitos exemplos onde este efeito pode ser diretamente observado, como por exemplo a introdução de uma amina cíclica ao inibidor NAPAP da trombina (**Figura 21**) aumenta em 60 vezes a afinidade pelo receptor.[Mack *et al.*, 1995] Entretanto, existem exemplos em que o ligante mantém parte de sua liberdade rotacional, translacional e conformacional mesmo estando ligado ao receptor. Um exemplo extremo desta situação, é a ligação da cânfora, do adamanteno e da tiocânfora ao citocromo $P450_{CAM}$ (uma monoxigenase da *Pseudomonas putida*). Todos estes ligantes podem girar livremente dentro do sitio ativo da $P450_{CAM}$, resultando em uma hidroxilação sem qualquer stereoseletividade definida.[Raag & Poulos, 1991]

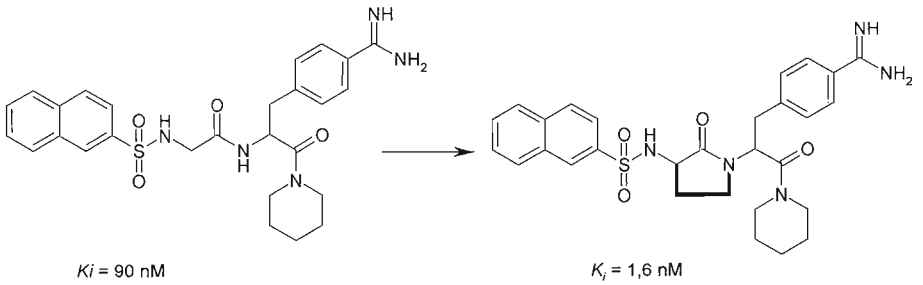


Figura 21. Representação esquemática dos inibidores da trombina, NAPAP e NAPAP-cíclico, respectivamente. Estão indicados os valores de K_i, determinados para cada um dos inibidores. [MACK *et al.*, 1995]

O receptor, por outro lado, pode também sofrer perda de graus de liberdade conformacional ao se formar o complexo com o ligante [RL], resultando em uma penalidade entrópica desfavorável à ligação. Esta perda de mobilidade do receptor, causada pela formação do complexo, é também conhecida como seleção conformacional (do inglês "*conformational selection*")[Höfliger & Beck-Sickinger, 2003] e é usada para explicar o mecanismo de ação de sítios alostéricos.[Luque & Freire, 2000] Um exemplo de seleção conformacional pelo ligante é a ação de agonistas ou de antagonistas de receptores de membrana acoplados à proteína-G.[Höfliger & Beck-Sickinger, 2003]

Para compensar a penalidade entrópica resultante da perda de graus de liberdade do receptor, algumas proteínas podem aumentar a mobilidade de resíduos distantes do sítio de ligação.[Homans, 2007; Zidek *et al.*, 1999] Este efeito foi descrito como um mecanismo de compensação "entropia-entropia" que visa anular o custo entrópico contrário à formação do complexo.[Homans, 2007]

O valor da constante de afinidade (K_i) do complexo [RL] também pode também ser obtido por cálculo.[Bohm, 2003] Para que se possa calcular com precisão a afinidade entre um receptor e um ligante é preciso conhecer a função de estado de todas as espécies envolvidas no processo de formação do complexo (*i.e.* o ligante solvatado, o receptor solvatado e o complexo solvatado). Uma função de estado é uma propriedade que depende apenas da estrutura de um sistema em equilíbrio. Por exemplo, a entropia e a entalpia são funções de estado que descrevem a energia do ligante ou do receptor solvatados. Rigorosamente, o cálculo de funções de estado deve ser feito a partir do ensemble médio do sistema, ou seja, considerando todos os possíveis microestados com igual probabilidade de ocorrerem. Os métodos mais precisos para a determinação de funções de estado são os cálculos de mecânica estatística, como os métodos Monte Carlo ou de Dinâmica Molecular. Estes métodos são capazes de apresentar uma amostragem significativa do ensemble do sistema e permitem uma estimativa fiel da entropia do sistema uma vez que consideram explicitamente as todas as moléculas do sistema, *i.e.* receptor, ligante e solvente. Estudos de mecânica estatística envolvendo protocolos de perturbação de energia livre (*FEP – Free Energy Perturbation*), são usados para determinar a variação da energia livre (ΔG) entre duas funções de estado, por exemplo, entre o ligante livre no solvente e o ligante complexado com o receptor. A partir da energia livre calculada por estas abordagens é possível determinar com precisão os valores de K_i de um complexo [*RL*].[Miyamoto & Kollman, 1993; Rao *et al.*, 1992; Reddy *et al.*, 1991] Entretanto, estes métodos apresentam limitações devido às aproximações usadas nos campos de força, como por exemplo, a impossibilidade de formar ou romper ligações covalentes. Deste modo, estes métodos são adequados apenas para séries pequenas de compostos, uma vez que requerem grandes recursos computacionais e não são confiáveis para o cálculo de diferenças de energia entre sistemas muito diferentes.[Hansson *et al.*, 2002; Bohm & Klebe, 1996]

Uma alternativa que permite o cálculo rápido e aproximado da energia de ligação é o uso de funções de ranqueamento. As funções de ranqueamento calculam o valor da energia de ligação considerando-se apenas uma única estrutura do complexo receptor-ligante mas não consideram as funções de estado do ligante e do receptor antes de formarem o complexo. Adicionalmente, as funções de ranqueamento consideram a energia do sistema como uma somatória de termos, independentes entre si, e cuja contribuição para a energia final é aditiva. Um exemplo do princípio da aditividade na interação receptor-ligante é o trabalho de Andrews[Andrews *et al.*, 1984], que calculou a contribuição de diferentes grupos funcionais para a energia de ligação, a partir de uma série com 200 complexos com valores de K_i conhecidos. A **Tabela 4** apresenta os valores da contribuição média dos diferentes grupos, utilizados por Andrews, para caracterizar os ligantes da serie escolhida.

Tabela 4. Valores[a] da contribuição média de cada grupo para a afinidade do complexo *RL* [Andrews *et al.*, 1984]

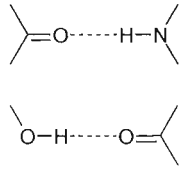
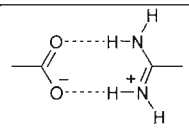
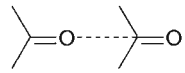
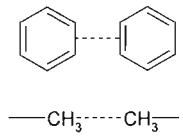
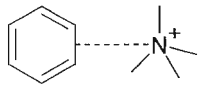
Grupo	Energia (Kcal/mol)
$C(sp^2)$	0,7
$C(sp^3)$	0,8
N^+	11,5
N	1,2
CO_2^-	8,2
OPO_3^{2-}	10,0
OH	2,5
C=O	3,4
O,S	1,1
Halogênio	1,3

a. Calculados a partir de uma série com 200 complexos com valores de K_i conhecidos.

Entretanto, o modelo aditivo não é capaz de descrever efeitos de cooperatividade, ou de anti-cooperatividade, que são observados no processo de reconhecimento molecular entre um ligante e seu receptor. [Williams & Bardsley, 1999; Bohm & Klebe, 1996] Por exemplo, a pré-organização e a estabilização do ligante em sua conformação bioativa antes da formação do complexo, que pode ocorrer com certos ligantes devido à interações hidrofóbicas intra-moleculares (processo conhecido como colapso hidrofóbico) é um fenômeno não aditivo, que resulta em desvios significativos no cálculo dos valores da energia de ligação, utilizando funções de ranqueamento.[Bohm & Klebe, 1996]

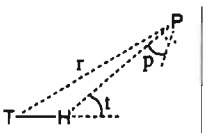
A grande maioria das funções de ranqueamento utilizadas nos programas de *docking* e de *virtual screening* decompõem a energia livre de ligação em termos capazes de descrever as ligações não covalentes estabelecidas entre o ligante e o receptor, como por exemplo: ligações de hidrogênio, ligações iônicas, ligações hidrofóbicas, interações π-π, interações de transferência de carga e complexos com metais. A **Tabela 5** apresenta as ligações não covalentes tipicamente encontradas em complexos receptor-ligante e as respectivas faixas de valores de energia envolvidas.

Tabela 5. Principais tipos de interações entre um composto e o sistema biológico, um exemplo e as respectivas faixas de energias envolvidas.

Tipo	Energia (kJ/mol)	Exemplo
ligação de hidrogênio	4-17	
Iônica	40	
dipolo-dipolo	4-17	
Hidrofóbica	4	
Π – cátion	17	

Estas ligações receptor-ligante podem ser descritas diretamente por mecânica molecular, baseando-se na mesma parametrização utilizadas nos campos de força conhecidos (*e.g.* CHARM, MM2 ou AMBER) ou, ainda, podem ser descritas por termos específicos, especialmente parametrizados para descrever interações receptor-ligante. Um exemplo destes termos especiais, utilizados em funções de ranqueamento, utiliza a função de energia desenvolvida por Goodford e colaboradores[Wade *et al.*, 1993; Boobbyer *et al.*, 1989], para descrever ligações de hidrogênio intermoleculares. O valor da energia de ligações de hidrogênio é dada pela equação:

$$E_H = \Sigma E_r \times \Sigma E_t \times \Sigma E_p$$



Onde E_r é a função de energia relativa à distância r;
E_t é a função de energia relativa ao ângulo t;
E_p é a função de energia relativa ao ângulo p;

Cada uma das funções de energia, respectivamente, E_r, E_t e E_p é parametrizada para descrever a orientação de ligações de hidrogênio observadas experimentalmente em estruturas de raios-X de moléculas orgânicas. Esta função de energia está incorporada no programa *GRID*,[Goodford, 1985] e é usada para identificar regiões favoráveis onde ligantes, representados por átomos de prova, podem se ligar ao receptor através de ligações de hidrogênio. Funções de energia como esta são desenvolvidas para cada um dos diferentes tipos de interação entre o ligante e o receptor e são baseadas em princípios físicos que determinam a afinidade.

Uma estratégia diferente para descrever as interações ligante-receptor se baseia na freqüência com que determinadas interações ocorrem em função da distância do par de átomos envolvidos na interação.[Gohlke *et al.*, 2000a] As distribuições de freqüência, para cada tipo de par de átomos, são extraídas de dados experimentais de cristalografia de raios-X, como os bancos de dados *PDB* e *CSD*. A partir dos gráficos das distribuições de freqüência, são obtidos potenciais de força média (do inglês *potential of mean force*) que são utilizados para o cálculo do valor da energia de interação ligante-receptor. A **Figura 22** apresenta três exemplos de potenciais de força média em função da distância para os pares de átomos, respectivamente, O.3-O.3; O.3-O.co2 e O.co2-N.pl3 (nomenclatura para tipos de átomos utilizadas no programa *SYBYL*[Sybyl, 2004]), derivados de 6026 complexos proteína-ligante depositados no *PDB*. Os valores de mínimo indicam as distâncias com maior probabilidade estatística de ocorrerem e correspondem às distâncias ideais de, respectivamente, uma ligação de hidrogênio normal(O.3-O.3), uma ligação de hidrogênio assistida por uma carga (O.3-O.co2) e, uma ligação iônica (O.co2-N.pl3).[Gohlke *et al.*, 2000a]

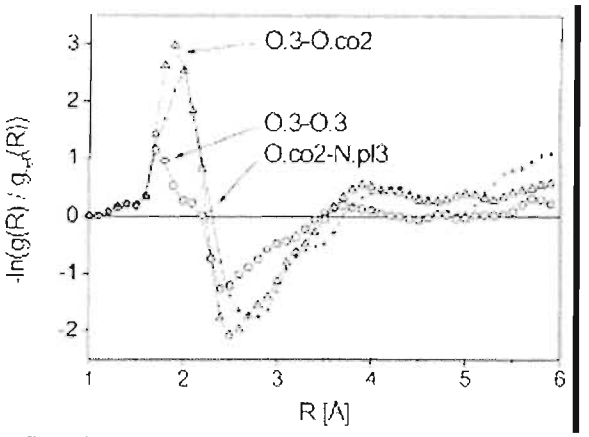


Figura 22. Gráfico do potencial de força média (-ln(g(R)/$g_{i,j}$(R))) em função da distância (R[Å]) para três pares de átomos, respectivamente, O.3-O.3; O.3-O.co2 e O.co2-N.pl3 (nomenclatura para tipos de átomos utilizada no *SYBYL*[Sybyl, 2004]), derivados de 6026 complexos proteína-ligante depositados no *PDB*. $g_{i,j}$(R) é a função de distribuição radial dos átomos tipo *i*, e *j*; e g(r) é a função de distribuição radial de dois átomos de referência. [Gohlke *et al.*, 2000a]

Além do cálculo dos valores de energia de interações intermoleculares, as funções de ranqueamento também calculam o valor da energia interna conformacional do ligante. O objetivo é penalizar os ligantes que assumem conformações altamente energéticas. A correspondente energia interna do receptor não é avaliada pois este é considerado rígido e em uma conformação de mínimo de energia. Considerações sobre variações de entropia conformacional, rotacional e translacional do ligante podem ser explicitamente incluídas em termos específicos, como por exemplo, a adição de um valor de contribuição constante para cada ligação com rotação livre, ou podem ser consideradas implícitas nos demais termos.[Gohlke *et al.*, 2000a; Ajay & Murcko, 1995]

Nenhum dos métodos de cálculo do valor da constante de afinidade, citados até o momento nesta tese, são capazes de prever tanto a formação ou, a quebra de ligações covalentes, bem como considerar efeitos quânticos, como por exemplo a estabilização adicional de sistemas com conjugação estendida. Para que tais efeitos pudessem ser considerados seria necessário realizar cálculos semi-empíricos ou, *ab-initio*, ainda muito complexos para sistemas com tantos átomos como são os como complexos receptor-ligante solvatados. Mesmo a aplicação de métodos híbridos como a Mecânica-Quântica/Mecânica-Molecular (MQ-MM) para o cálculo de constantes de afinidade são, ainda, muito custosos e limitados a casos excepcionais.[Reddy & Erion, 2007]

Devido a complexidade das interações envolvidas na formação do complexo, a previsão da afinidade entre um ligante e seu receptor biológico ainda será limitada a cálculos aproximados e relativos, como o caso das funções de ranqueamento utilizadas em programas de *docking*, ou restrita a poucos compostos, quando se deseja uma precisão maior, como por exemplo utilizando cálculos de *FEP*. Qualquer que seja a estratégia escolhida, no entanto, existem exemplos de sucesso e desafios para serem superados, tornando o desenvolvimento de novos fármacos, com o auxílio do computador – *CADD* - uma área de pesquisa ao mesmo tempo recompensadora e desafiadora para aqueles que escolhem trilhar seus caminhos.

1.6. Doença de Chagas

1.6.1. Epidemiologia, Transmissão e Sintomas

A doença de Chagas é endêmica no Brasil e estima-se que na América Latina entre 16 e 18 milhões de pessoas estejam infectadas pelo parasita e que ocorram aproximadamente 50.000 mortes causadas pela doença de Chagas por ano.[Gelb & Hol, 2002; Coura & de Castro, 2002; WHO, 2002; Urbina, 2001; Pecoul *et al.*, 1999] O agente etiológico da doença de Chagas é o protozoário *Trypanosoma cruzi*, que tem como vetor algumas espécies de insetos triatomíneos, sendo o principal o *Triatoma infestans*.

O parasita invade o homem pela ferida causada pela picada do inseto, por transfusão de sangue contaminado ou por via oral, através da ingestão de alimentos contaminados com o inseto macerado. Uma vez na corrente sanguínea do vertebrado, o protozoário invade as células do hospedeiro, onde passa a se reproduzir até a rompimento da membrana celular.[24] Durante a fase aguda da doença, os sintomas são semelhantes aos de uma infecção de gravidade variável, dependendo da virulência do parasita. Durante a fase crônica, o hospedeiro pode desenvolver sintomas como miocardite, megaesôfago ou megacólon. Em ambas as fases da doença, os sintomas podem levar ao óbito.[Filardi & Brener, 1987]

1.6.2. Terapias em uso

BIBLIOTECA
INSTITUTO DE QUÍMICA
Universidade de São Paulo

Desde que foi descrita pelo médico sanitarista Carlos Chagas, em 1909, muitos compostos foram testados sem sucesso para a quimioterapia da doença de Chagas. Dentre os compostos quimioterápicos, testados neste período, podem ser destacados derivados quinoléicos, arsenicais, fenantridinas, sais de ouro, sais de bismuto, sais de cobre, sais de estanho, iodeto de sódio, hidrazidas, anti-histamínicos, sulfonamidas, cortisona e anfotericina B, além de 30 antibióticos diferentes e alguns nitrofuranos.[Coura & de Castro, 2002]

No final da década de 1960, foram introduzidos no mercado dois medicamentos comprovadamente eficazes para o tratamento da fase aguda da doença; o nifurtimox [3-metil-4-(5-nitrofurfurilidenoamino)tetradro-1,4-tiazina-1,1-dióxido] e o benznidazol [*N*-benzil-2-nitroimidazol acetamida] - introduzidos no Brasil com os nomes Lampit® (Bayer) e Rochagan® (Roche), respectivamente (**Figura 23**).

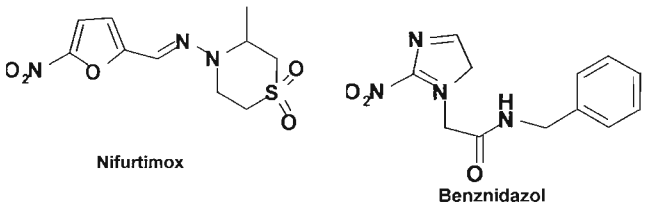


Figura 23. Representação esquemática da estrutura dos compostos antichagásicos nifurtimox e benznidazol, respectivamente.

O mecanismo de ação citotóxica proposto para compostos nitroaromáticos, como o nifurtimox e o benznidazol, envolve em uma primeira etapa a redução do grupo nitro, para o correspondente ânion nitro-radical. Numa segunda etapa, esse ânion nitro-radical reage com O_2 gerando espécies quimicamente reativas, como o ânion superóxido e o peróxido de hidrogênio. Estas espécies radicais livres são bastante danosas para as células e, uma vez geradas, podem interagir com vários componentes celulares, como os ácidos nucléicos, e propagar seus efeitos indesejáveis. Isso é particularmente verdadeiro para o *T. cruzi* devido à sua deficiência das enzimas antioxidantes como catalase e a glutationa peroxidase, o que torna o parasita extremamente vulnerável ao peróxido de hidrogênio. As células dos mamíferos, ao contrário, têm várias enzimas, como catalase, peroxidase e superóxido dismutase, prevenindo o efeito letal dos radicais livres.[Docampo, 1990; Docampo & Moreno, 1984; Moreno *et al.*, 1983]

Na década de 1980, o nifurtimox teve sua comercialização interrompida no Brasil e recentemente o benznidazol foi vendido pela Roche para o Laboratório Farmacêutico do Estado de Pernambuco (Lafepe). O desinteresse da indústria farmacêutica na produção de medicamentos para a doença de Chagas, considerados como "medicamentos sociais", está ligado à baixa demanda, restrito somente a alguns países na América Latina, para população de baixa renda e, portanto, com pequena margem de lucro.[Jannin & Villa, 2007]

Os tripanomicidas nifurtimox e benznidazol causam efeitos adversos graves e não são eficazes na fase crônica da doença. Além disso, algumas cepas de *T. cruzi* já apresentaram resistência à terapia com estes medicamentos.[Coura & de Castro, 2002; Urbina, 2001; Pecoul *et al.*, 1999; Filardi & Brener, 1987] A deficiência na quimioterapia, bem como o limitado interesse na pesquisa de novos antichagásicos - restrito a países em desenvolvimento – motivam a busca de novos compostos antichagásicos.

1.6.3. Planejamento de novos compostos antichagásicos

Atualmente, são descritos diversos estudos que visam o planejamento racional de novos compostos antichagásicos. Em comum, estes estudos procuram alvos moleculares capazes de explorar as diferenças bioquímicas e morfológicas entre o agente invasor e o hospedeiro, para que o composto desenvolvido possa apresentar a seletividade requerida.[Du *et al.*, 2002; Gelb & Hol, 2002; Coura & de Castro, 2002; Urbina, 2001; Werbovetz, 2000; Bohm & Klebe, 1996]

Entre os alvos de interesse usados no planejamento de compostos antichagásicos encontram-se enzimas essenciais para o desenvolvimento dos parasitos no hospedeiro, como por exemplo: proteases,[Cazzulo *et al.*, 2001; Lima *et al.*, 2001; Yong *et al.*, 2000; Mcgrath *et al.*, 1995] tripanotiona redutase,[Lopez *et al.*, 2000; Jockersscherubl *et al.*, 1989; Fairlamb & CERAMI, 1985; Fairlamb *et al.*, 1985] enzimas do metabolismo de esteróis,[Elhalem *et al.*, 2002; Urbina, 2001] glicoconjugados de superfície,[Colli & Alves, 1999] gliceraldeído-3-fosfato desidrogenase,[Pavao *et al.*, 2002; Souza *et al.*, 1998] transferases,[Coura & de Castro, 2002; Werbovetz, 2000; Freymann *et al.*, 2000] DNA-topoisomerases,[Coura & de Castro, 2002; Jonckers *et al.*, 2002; Werbovetz, 2000; Kubinyi, 1998a] diidrofolato redutase,[Gilbert, 2002; Li *et al.*, 2000] enzimas do

metabolismo do inositol,[Urbina, 2001] e farnesilpirofosfato sintase.[Coura & de Castro, 2002]

As proteases de protozoários são alvos interessantes devido ao seu papel importante em inúmeros processos bioquímicos como nutrição, replicação, metabolismo, sobrevivência e patologia do parasita.[Powers *et al.*, 2002] Adicionalmente, as proteases podem ser inibidas especificamente por inibidores peptídicos, peptidomiméticos ou não-peptídicos, estes últimos, com características tipicamente *drug-like.* As diferenças estruturais entre as proteases de parasitas e as correspondentes humanas já foram exploradas anteriormente com sucesso, para o desenvolvimento de terapias para diversas doenças infecto-contagiosas.[Leung *et al.*, 2000; Klebe, 2000] Dentre as proteases de interesse para a doença de Chagas, a cruzipaína[Eakin *et al.*, 1992; Cazzulo *et al.*, 1990] é reconhecida como um dos principais alvos para o desenvolvimento novos compostos antichagásicos.[Cazzulo *et al.*, 2001; McKerrow, 1999]

A cruzipaína é a protease mais importante do *T.cruzi,* expressa nos três estágios de vida do parasita (*i.e.*, tripomastigota, epimastigota e amastigota).[Cazzulo *et al.*, 1997] A enzima é encontrada tanto no lisossoma do parasita, quanto presa na membrana plasmática e pode ser excretada para o meio extracelular.[Cazzulo *et al.*, 2001] Suas funções incluem nutrição do parasita, diferenciação, invasão e alteração da resposta imune do hospedeiro, sendo o antígeno imuno-dominante reconhecido durante a infecção humana.[Cazzulo *et al.*, 2001; DelNery *et al.*, 1997a; Meirelles *et al.*, 1992; Scharfstein *et al.*, 1986]. A cruzaína é codificada por mais de 100 cópias de genes polimórficos localizados em vários cromossomos podendo ser expressa em diversas isoformas. Ainda assim é uma das enzimas mais bem caracterizadas do parasita causador da doença de Chagas.[Cazzulo, 1999]

A primeira demonstração do efeito de inibidores de cisteíno-proteases sobre o *T. cruzi* foi o trabalho de Ashall e colaboradores,[Ashall *et al.*, 1990] que mostraram a lise de formas tripomastigotas do parasita, em meio de cultura com células de mamífero, na presença de um inibidor fluormetilcetona. Meirelles e colaboradores[Meirelles *et al.*, 1992] demonstraram que as formas tripomastigotas perdiam a capacidade de penetrar em células cardíacas na presença de inibidores diazometanos e que estes inibidores eram capazes de bloquear a replicação intracelular de formas amastigotas e sua diferenciação em tripomastigotas. Estes resultados foram confirmados por Harth e colaboradores,[Harth *et*

al., 1993] que também mostraram que a replicação de amastigotas, e sua diferenciação a tripomastigotas eram bloqueadas por inibidores fluormetilcetona, enquanto que Bonaldo e colaboradores[Bonaldo *et al.*, 1991] mostraram o efeito de inibidores diazometanos no bloqueio da diferenciação de epimastigotas a tripomastigotas. Decazzulo e colaboradores[Decazzulo *et al.*, 1994] testaram três inibidores (uma aciloximetilcetona e duas diazometilcetonas) e observaram que os inibidores de cisteíno-protease eram capazes de bloquear completamente a diferenciação de epimastigotas a tripomastigotas. Usando um modelo animal, o tratamento de dez camundongos com uma vinilsulfona resultou em 100% de cura da infecção, enquanto que os seis animais controle, mantidos sem tratamento morreram em até 22 dias após a inoculação do parasita.[Engel *et al.*, 1998] Mais recentemente,[Doyle *et al.*, 2007] camundongos imunossuprimidos infectados com *T.cruzi*, foram tratados e curados com um inibidor de cisteíno-protease vinilsulfona, indicando que estes compostos podem ser usados para o tratamento de pacientes imunodeprimidos, como são os portadores de HIV. Todos estes resultados corroboram entre si para que a cruzaína seja considerada um alvo de interesse para a quimioterapia da doença de Chagas.

A cruzipaína foi isolada e purificada a partir do *T.cruzi* inicialmente por Bontempi e colaboradores,[Bontempi *et al.*, 1984] e sua forma recombinante, contendo apenas a unidade catalítica (denominada cruzaína), foi expressa em *Escherichia coli* por Eakin e colaboradores.[Eakin *et al.*, 1992] Estudos comparativos das formas selvagem e recombinante da enzima não identificaram diferenças significativas em relação à especificidade, à estabilidade ou aos correspondentes parâmetros de ativação.[Judice *et al.*, 2001] A cruzaína foi identificada como sendo uma cisteíno-protease da família da papaína. Esta classe de proteases é caracterizada por conter uma tríade catalítica (*Cys25*, *His159*, e *Asn175*) e uma cavidade para acomodar um oxiânion (*Cys25* e *Gln19*).[Otto & Schirmeister, 1997] O enxofre da *Cys25* e o anel imidazólico da *His159* formam um par iônico. O sítio de ligação é ainda formado por uma seqüência de cavidades alinhadas em sentidos opostos, de modo a acomodar os aminoácidos do substrato. Segundo a terminologia usual as cavidades direcionadas no sentido *N*-terminal do substrato são denominadas *S1*, *S2*, *S3*, ..., S_n, enquanto que as cavidades direcionadas no sentido C-terminal do substrato são denominadas *S1'*, *S2'*, *S3'*, ..., S_n'. Analogamente, os resíduos do

substrato que ocupam as cavidades correspondentes são denominados $P_n, P_2, P_1, P_1', P_2', P_n'$, respectivamente, conforme apresentado na **Figura 24.**

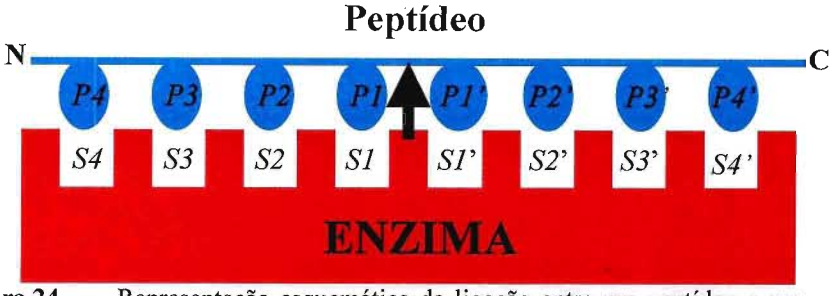


Figura 24. Representação esquemática da ligação entre um peptídeo e uma enzima protease. Estão indicadas a terminologia usual para as cavidades da enzima e os correspondentes resíduos do substrato. A seta indica o ponto em que ocorre a hidrólise do peptídeo.

No mecanismo de ação proposto para esta classe de enzimas, o nucleófilo (ânion tiolato) ataca o carbono carbonílico da ligação peptídica (**Figura 25-A**) gerando um intermediário tetraédrico (**Figura 25-C**). O intermediário tetraédrico é estabilizado por ligações de hidrogênio na cavidade oxiânion. Este intermediário tetraédrico perde a porção amina deixando a enzima acilada (**Figura 25-E**). A enzima acilada reage com uma molécula de água (**Figura 25-F**) liberando o produto ácido e restaurando a enzima livre (**Figura 25-A**). Estudos de efeito isotópico mostraram que a etapa limitante do mecanismo catalítico é a quebra da ligação C-N durante a decomposição do intermediário tetraédrico (**Figura 25-D**).[Babine & Bender, 1997] Estes resultados foram corroborados por estudos de mecânica quântica, que propuseram um mecanismo concertado, em que o intermediário tetraédrico não apresenta um papel essencial.[Harrison *et al.*, 1997] Adicionalmente, a substituição de um enxofre, no lugar do oxigênio carbonílico da ligação hidrolisável, em substratos de cisteíno-proteases, não altera a reatividade da enzima, corroborando a hipótese de que a estabilização do intermediário tetraédrico não é essencial para o mecanismo catalítico destas enzimas.[Asboth *et al.*, 1985; Asboth & Polgar, 1983]

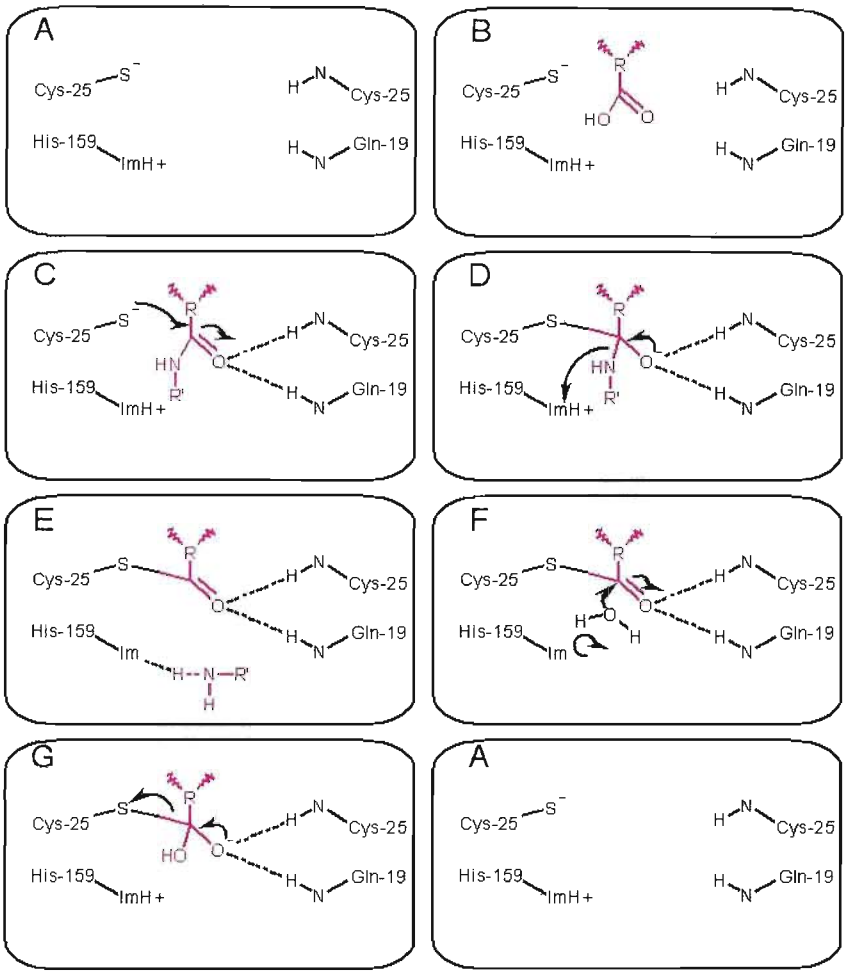


Figura 25. Representação esquemática do mecanismo de proteólise de cisteíno-proteases: (**A**) Forma ativa da enzima com o par iônico *SG⁻- Cys25/ImH⁺-His159*; (**B**) aproximação de um ligante; (**C**) formação do complexo não covalente com ligações de hidrogênio entre o oxigênio carbonílico do peptídeo e os hidrogênios da *Cys25* (cadeia principal) e da *Gln19* (cadeia lateral); (**D**) primeiro intermediário tetraédrico, formado pelo ataque nucleofílico do *SG-Cys25* ao carbono carbonílico do peptídeo, e estabilizado pelas ligações de hidrogênio; (**E**) esterificação da enzima, e saída da porção amina após catalise ácida do imidazol da *His159*; (**F**) ataque nucleofílico de uma molécula de água ao carbono acil da enzima; (**G**) Segundo intermediário tetraédrico; liberação do produto ácido e reconstituição da enzima livre (**A**). [Otto & Schirmeister, 1997]

A especificidade da cruzaína foi amplamente estudada por diversos grupos. Tipicamente, a hidrólise de pequenos peptídeos pela cruzaína requer a presença dos resíduos *Arg* ou *Lys* na posição *P1*,[Alves *et al.*, 2001b; DelNery *et al.*, 1997b] e a presença de pelo menos um segundo resíduo hidrofóbico, ou positivo, na posição *P2*, sendo preferidos os resíduos de *Arg*, de *Leu*, de *Phe* ou de *Val*.[Lecaille *et al.*, 2001; Alves *et al.*, 2001a; Harris *et al.*, 2000; Gillmor *et al.*, 1997] As posições *P3* e *P4* podem acomodar grande diversidade de resíduos, sendo que a posição *P3* apresenta uma preferência para

resíduos básicos,[Harris *et al.*, 2000; Serveau *et al.*, 1999] enquanto que a posição *P4* tem pouca afinidade para resíduos hidrofóbicos.[DelNery *et al.*, 1997b] A posição *P1'* tem baixa especificidade e pode acomodar diversos tipos de resíduos,[Brinen *et al.*, 2000; DelNery *et al.*, 1997b] enquanto que a posição P2' mostrou preferência por um resíduo de prolina.[Serveau *et al.*, 1996]

A maior parte dos inibidores de cruzaína conhecidos são formados por uma porção peptídica, responsável pelo reconhecimento pela enzima e uma porção contendo um grupo funcional capaz de reagir reversivelmente ou irreversivelmente com a cisteína catalítica do sítio ativo. Os inibidores reversíveis conhecidos da cruzaína podem ser nitrilas, trifluormetilcetonas e aldeídos.[Lecaille *et al.*, 2002; Powers *et al.*, 2002; Scheidt *et al.*, 1998; Otto & Schirmeister, 1997; Babine & Bender, 1997] As nitrilas e a trifluormetilcetonas são inibidores que simulam o estado de transição catalítico, enquanto que os aldeídos reagem com a enzima para formar um intermediário tetraédrico hemicetal, que não apresenta a funcionalidade necessária para prosseguir com a reação até a formação dos produtos. Os inibidores irreversíveis da cruzaína podem ser α-clorometilcetonas,[Wood *et al.*, 2003] α-fluormetilcetonas,[Lecaille *et al.*, 2001; Gillmor *et al.*, 1997; Mcgrath *et al.*, 1995; Eakin *et al.*, 1993; Harth *et al.*, 1993] α-hidroximetilcetonas,[Huang *et al.*, 2003] α-cetoácidos,[Choe *et al.*, 2005] diazocetonas,[Serveau *et al.*, 1996; Meirelles *et al.*, 1992] aziridinas,[Vicik *et al.*, 2006b] semicarbazonas,[Otto & Schirmeister, 1997] dicetonas,[Otto & Schirmeister, 1997] cetonas α,β-insaturadas,[Otto & Schirmeister, 1997] epóxidos,[Gonzalez *et al.*, 2007; Roush *et al.*, 2000; Roush *et al.*, 1998a; Cazzulo *et al.*, 1990] vinil sulfonas[Jaishankar *et al.*, 2007; Brinen *et al.*, 2000; Jacobsen *et al.*, 2000; Roush *et al.*, 1998b; Palmer *et al.*, 1995] e mercaptometilcetonas.[Freitas *et al.*, 2007; Huang *et al.*, 2002] Todos estes inibidores apresentam, em comum, um átomo eletrofílico capaz de sofrer um ataque do tiolato da *Cys25*, resultando na alquilação irreversível da enzima. Alguns destes inibidores, como as α-clorometilcetonas e as diazocetonas são compostos muito reativos e, portanto, não são bons candidatos para o desenvolvimento de uma terapia contra a doença de Chagas. Adicionalmente, inibidores peptídicos apresentam baixa biodisponibilidade oral e dificilmente são convertidos em bons fármacos.

A cruzaína foi cristalizada[Eakin *et al.*, 1993] e sua estrutura foi resolvida por cristalografia de raios-X.[Mcgrath *et al.*, 1995] (ver **Tabela 6** – *Resultados e Discussão*)

A partir da disponibilidade da estrutura da cruzaína, algumas técnicas de *CADD* foram aplicadas para o desenvolvimento de inibidores da cruzaína. Por exemplo, Li e colaboradores[Ring *et al.*, 1993] realizaram buscas por ligantes da cruzaína, não-peptídicos, em bancos de dados de compostos *drug-like* (*Fine Chemicals Directory*), utilizando o programa de *docking, DOCK*.[Ewing & Kuntz, 1997] Estas buscas resultaram na descoberta da *bis*-2-hidroxi-benzaldeído-(carboxi)-hidrazona (**Figura 26**), que foi experimentalmente confirmado como um inibidor da cruzaína (IC_{50} = 20 μM).

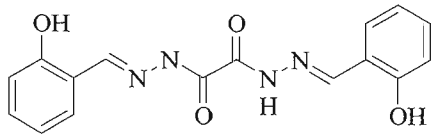


Figura 26. Representação esquemática da estrutura do inibidor da cruzaína bis-2-hidroxi-benzaldeído-(carboxi)-hidrazona

A partir desta primeira hidrazida, diversas séries de derivados de hidrazidas[Leite *et al.*, 2007; Millet *et al.*, 2002; Caffrey *et al.*, 2002; Ifa *et al.*, 2000; Troeberg *et al.*, 2000; Selzer *et al.*, 1997] e chalconas[Chibale & Musonda, 2003; Liu *et al.*, 2003; Lunardi *et al.*, 2003; Troeberg *et al.*, 2000; Nielsen *et al.*, 1998; Li *et al.*, 1996; Li *et al.*, 1995] foram preparadas para o desenvolvimento de compostos mais potentes e seletivos, e para o desenvolvimento de relações quantitativas entre estrutura e atividade. O mecanismo de ação proposto para esta série de compostos sugere que o ataque nucleófilo do tiolato da *Cys25* pode ocorrer tanto no carbono carbonílico da hidrazona/chalcona, como no carbono vinílico (para uma adição de Michael).[Rodrigues *et al.*, 2002] Em um outro trabalho de *virtual screening,* utilizando o programa *DOCK*[Ewing & Kuntz, 1997] para a busca no banco de dados *ACD*, foi identificada uma série de uréias, de tiouréias e de tiossemicarbazonas como potentes inibidores da cruzaína.[Du *et al.*, 2000; Eakin *et al.*, 1992] Assim como no exemplo anterior, diversas séries destes compostos foram preparadas para a obtenção de compostos mais ativos e para a elaboração de modelos de *QSAR*.[Leite *et al.*, 2006; Fujii *et al.*, 2005; Greenbaum *et al.*, 2004; Chiyanzu *et al.*, 2003; Du *et al.*, 2002] As tiossemicarbazonas foram descritas como inibidores reversíveis da cruzaína, que simulam e estabilizam o estado de transição da enzima, conforme apresentado na **Figura 27**.

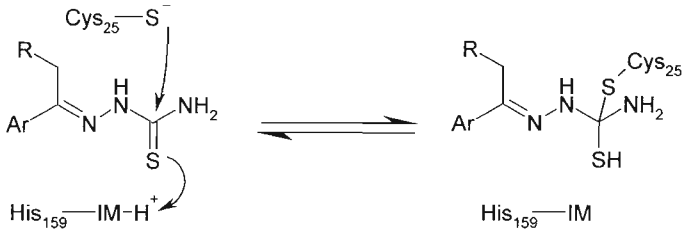


Figura 27. Representação esquemática do mecanismo de ação proposto para a inibição da cruzaína por tiossemicarbazonas. [EAKIN *et al.*, 1992]

Alguns dos compostos identificados como inibidores da cruzaína foram caracterizados como inibidores promíscuos,[Malvezzi *et al.*, 2008; McGovern *et al.*, 2002] ou seja, compostos capazes de formar agregados moleculares que nas condições dos testes de atividade inibem enzimas de modo inespecífico. Como descrito anteriormente nesta tese, inibidores promíscuos são classificados como falso-positivos e são indesejados para o desenvolvimento de fármacos.

Apesar da extensa literatura dedicada à busca de inibidores da cruzaína, até o presente momento, não existem relatos conhecidos de compostos candidatos que estejam em etapas clínicas de desenvolvimento. Isto significa que o desenvolvimento pré-clínico de compostos ativos ainda não encontrou o candidato ideal, que reúna as condições de especificidade, de potência, de toxicidade e de biodisponibilidade necessárias para o início dos testes em seres humanos. O interesse na descoberta de novas classes de compostos *drug-like,* inibidores de cisteíno-proteases, não se limita ao tratamento da doença de Chagas. A hidrólise desregulada ou incontrolada causada por cisteíno-proteases pode levar a diversas doenças incluindo enfisema, infarto, infecções, malária, câncer, Alzheimer, inflamação, catarata, artrite, osteoporose entre outras.[Vicik *et al.*, 2006a; Powers *et al.*, 2002] Assim, a aplicação de estratégias modernas para a descoberta de novos inibidores da cruzaína pode beneficiar não somente o desenvolvimento de um quimioterapia para a doença de Chagas, mas também pode contribuir para a terapia de doenças envolvendo outras cisteíno-proteases.

2. Materiais e Métodos

2.1. Reagentes e Solventes

- Cruzaína: *EC 3.4.22.51* (63 nM ou 1,8 μM solução aquosa; cortesia de Prof. Dr. L. Juliano Depto. Biofísica, UNIFESP-SP)
- DMSO: *Dimetilsulfóxido* (P.A., Merck)
- DTT: *Ditiotreitol* (99%, Sigma)
- EDTA: *Ácido Etilenediaminotetracético* (P.A., Fluka)
- Na_2HPO_4: *Fosfato de dissódico* (P.A., Synth);
- Z-FR-MCA: *Benzoil-fenilalanina-arginina-7-amido-4-metil-cumarina* (Sigma)
- **E-64**: *Trans-epoxisuccinil-l-leucilamido-(4-guanidino)butano* (2,8 mM solução aquosa; cortesia de Prof. Dr. L. Juliano, Depto. Biofísica, UNIFESP-SP)
- **ZINC03326243**: *2-cloro-N-[2-[(3-ciano-4,5,6,7-tetraidrobenzotiofenil)carbamoil] -1-fenil-etil]benzamida,* Enamine Ltd (T5306814)
- **ZINC03192044**: *Tiofeno-2-carboxil[2-(1H-indol-3-il)-1-(pirrolidin-2-ilideno hidrazinocabonil)-etil]amida*; Enamine Ltd (T0503-8879)
- **ZINC02663001**: *N-[1-(1H-Benzoimidazol-2-il)-2-fenil-etil]-3-tiofen-2-il-acrilamida*; Enamine Ltd (T5254688)
- **ZINC01936854**: *N,N'-bis(2-cianofenil)oxamida*; Asinex Limited (BAS 01936854)
- **ZINC02682879**: *(E)-3-(2-clorofenil)-N-(3-cianotiofen-2-il)prop-2-enamida;* Life Chemicals (F1385-0068)
- **ZINC02470662**: *N-(1-ciano-cicloexil)-2-oxo-cromeno-3-carboxamida;* Specs (AP-845/41115584)
- **ZINC00874200**: *Ácido (Z)-3-(3-cloro-4,5-dimetóxi-fenil)-2-ciano-acrilico N-[2-(1H-indol-3-il)-etil]-hidrazida;* IBScreen (STOCKS2S-49269)
- **ZINC00890529**: *(Z)-2-Ciano-N-ciclopentil-3-[1-({[(R)-1-(tetraidro-furan-2-il)metil]-carbamoil}-metil)-1H-indol-3-il]-acrilamida*; Asinex (BAS02985564)
- **ZINC01414485:** *Ácido 5-[N'-(benzilcarbamoil-ciano-metileno)-hidrazino]-3H-imidazol-4-carboxílico etil éster*; Toslab (600063)
- **ZINC01758885:** *(Z)-2-Ciano-N-(2,4-dimetil-fenil)-3-(1-{[(furan-2-ilmetil)-carbamoil]-metil}-1H-indol-3-il)-acrilamida*; IBScreen (STOCK3S-57280)
- **ZINC01794384:** *Ácido 4-[N'-(3-cloro-benzo[b]tiofeno-2-carbonil)-hidrazino]-4-oxo-but-2-enóico etil éster*; Life Chemicals (F3145-4620)
- **ZINC01794422:** *Ácido 3-{N'-[2-(1-bromo-naftalen-2-iloxi)-acetil]-hidrazinocarbonil}-acrílico*; Life Chemicals (F3145-4606)

- **ZINC01794844:** *Ácido 3-{N'-[2-(2-isopropil-5-metil-fenoxi)-acetil]-hidrazinocarbonil}- acrílico*; Life Chemicals (F3145-3436)
- **ZINC01806509:** *N-(4-Metóxi-fenil)-2-[(4-metóxi-fenilcarbamoil)-metildisulfanil]-acetamida*; Life Chemicals (F3047-0312
- **ZINC02188675:** *Ácido 2-((R)-2-sulfossulfanil-butirilamino)-4,5,6,7-tetraidro-benzo[b]tiofeno-3-carboxílico etil éster*; Specs (AG-205/37047002
- **ZINC02198693:** *Ácido 3-[N'-(2,5-dimetoxi-fenilaminoxalil)-hidrazino]-but-3-enóico (4-cloro-fenil)-amida*; IBScreen (STOCK2S-31029
- **ZINC02209285:** *Ácido (Z)-2-ciano-3-(5-piperidin-1-il-furan-2-il)-acrílico N-[2-(1H-indol-3-il)-etil]-hidrazida*; IBScreen (STOCK2S-31935
- **ZINC02342142:** *Ácido (S)-2-{2-ciano-3-[5-(4-fluoro-fenil)-furan-2-il]-acriloilamino}-6-metil-4,5,6,7-tetraidro-benzo[b]tiofeno-3-carboxílico amida*; IBScreen (STOCK2S-99288
- **ZINC02399860:** *N-[2-(1H-Benzoimidazol-2-il)-etil]-2-(3-furan-2-il-acriloilamino)-3-(4-isopropoxi-fenil)-acrilamida*; IBScreen (STOCK4S-32328)
- **ZINC02400672:** *Ácido (Z)-3-[N'-(7,7-dimetil-2-fenil-7,8-diidro-5H-pirano[4,3-d]pirimidin-5-il)-hidrazinocarbonil]- acrílico*; IBScreen (STOCK4S-33348)
- **ZINC02411240:** *Ácido 2-{[1-o-tolil-met-(E)-ilideno-hidrazinooxalil]-amino}-5,6-diidro-4H-ciclopenta[b]tiofeno-3-carboxílico metil éster*; IBScreen (STOCK4S-44363)
- **ZINC02768688:** *(Z)-3-(4-Carbamoilmetoxi-3-ethoxi-fenil)-2-ciano-N-fenetil-acrilamida*; Vitas-M (STK008910)
- **ZINC04964356:** *Ácido (2,5-dimetil-fenoxi)-acático [2-{[2-(2,5-dimetil-fenoxi)-acetil]-hidrazono}-et-(Z)-ilideno]-hidrazida*; ChemBridge (5230379)
- **ZINC03243155:** *Ácido but-2-enodióico bis-[(2-metil-benzotiazol-6-il)-amida]*; Enamine (T0507-1260)
- **ZINC03306238:** *Ácido carbônico mono-[4-((Z)-2-ciano-2-fenetilcarbamoil-vinil)-2-metóxi-fenil] éster*; Enamine (T0515-7498)

2.2. Equipamentos

2.2.1. Fluorímetro

As medidas de inibição da cruzaína foram realizados nos seguintes aparelhos:

- Fluorímetro HITACHI, modelo F-2000, equipado com agitação magnética e banho termostatizado, pertencente ao laboratório do Prof. Dr. Luiz Juliano, Departamento de Biofísica, Universidade Federal de São Paulo-SP.

- Fluorímetro HITACHI, modelo F-4500, equipado com agitação magnética e banho termostatizado, pertencente ao laboratório do Prof. Dr. Willem Baader do Departamento de Química Fundamental, Instituto de Química da Universidade de São Paulo - SP.
- Fluorímetro, HITACHI, modelo F-4500, equipado com agitação magnética e banho termostatizado, pertencente ao laboratório da Profª. Drª. Shirley Schreier do Departamento de Bioquímica do Instituiu de Química da Universidade de São Paulo - SP.

2.2.2. pH-âmetro

Todos os valores de pH das soluções tampão utilizados na determinação da inibição da cruzaína foram determinados em um pH-âmetro marca Digicrom, modelo DM-20, utilizando-se eletrodos de vidro Ag/AgCl, $KCl_{(sat.)}$, calibrados com solução tampão-titrisol, nos valores de pH = 3,0 e 7,0 respectivamente.

2.2.3. Estações de Trabalho

Os cálculos computacionais foram realizados nas seguintes estações de trabalho:

- PC com processador Athlon XP 2600+, com 512 de memória RAN. Sistema operacional *Linux Red Hat 7.3;*
- Estação de trabalho AMD Opteron, com 4 processadores e 4GB de memória RAM. Sistema operacional *Linux Fedora Core 4*;
- Estação de trabalho SGI Power Challenge 10000, com 4 processadores de 194MHz, e 512MB de memória RAM. Sistema operacional *IRIX 6.5.* Somente para os cálculos com o programa *SYBYL*;[Sybyl, 2004]
- Estação de trabalho SGI Fuel, com um processador MIPS R16000 800MHz e 1GB de memória RAM. Sistema operacional *IRIX 6.5*. Somente para os cálculos com o programa *CATALYST*.[Catalyst, 2005]

2.3. Bancos de dados de moléculas virtuais

Para o desenvolvimento, validação e aplicação do modelos de *virtual screening* foram utilizados os seguintes bancos de dados de moléculas virtuais:

- *ZINC*[Irwin & Shoichet, 2005] versão 5 e versão 7, disponíveis para *download* no endereço eletrônico *http://blaster.docking.org/zinc/*.

- *Tocris*[Tocris, 2005] disponível através do CD: "*Tocris Structure Database Files*";
- *NCI*[NCI-NIH, 2005] disponível gratuitamente no endereço eletrônico http://cactus.nci.nih.gov/ncidb2/download.html; e
- *Cambridge Structural Database* (*CSD*)[Taylor, 2002] disponível no endereço eletrônico http://www.ccdc.cam.ac.uk/.
- *Protein Data Bank* (*PDB*)[Westbrook *et al.*, 2003], disponível para consulta no endereço eletrônico http://www.rcsb.org

2.4. Programas computacionais

2.4.1. *CATALYST*

O modelo farmacofórico[Leite *et al.*, 2006] foi construído utilizando-se o programa *CATALYST*,[Catalyst, 2005] versão 4.11, comercializado pela *Accelrys*.

2.4.2. *CAVBASE*

A procura por cavidades similares à cruzaína foi realizada com o módulo *CAVBASE*,[Schmitt *et al.*, 2002] presente junto do programa *RELIBASE+*[Gohlke *et al.*, 2000a] comercializado pela *CCDC*.[CCDC, 2006]

2.4.3. *CORINA*

As estruturas 3D de todas as moléculas do banco de dados *ZINC* e a conversão entre os diferentes formatos de arquivos de moléculas (*.smiles*; *.mol2*; *.sdf*; *.PDB*, respectivamente) foram realizadas pelo programa *Corina*[Sadowski & Gasteiger, 1993] versão 3.20, comercializado pela *Molecular Networks* (www.molecular-networks.com).

2.4.4. *DBFILTER*

A aplicação dos filtros para propriedades *drug-like* foi feita com o programa *DBFilter*,[Wang, 2003] versão 2.2.7., distribuído gratuitamente no endereço eletrônico: http://home.pchome.com.tw/team/gentamicin/mol/mol.htm (em Julho de 2003)

2.4.5. *DEEPVIEW*

O alinhamento e a visualização das estruturas dos complexos da cruzaína, bem como a seleção dos resíduos que compõem o sítio ativo foram realizadas no programa *DeepView*,[Guex & Peitsch, 1997] versão 3.7, distribuída gratuitamente no endereço eletrônico http://au.expasy.org/spdbv/

2.4.6. *GOLD*

Os procedimentos de *docking* foram executados pelo programa, *GOLD*,[Verdonk *et al.*, 2003] versão 3.0, comercializado pela *CCDC*.[CCDC, 2006]

2.4.7. *GRID*

Os valores dos campos de interação moleculares utilizados para a geração do modelo farmacofórico foram calculados com o pacote de programas *GRID*,[Goodford, 1985] versão 22a, distribuído gratuitamente no endereço eletrônico http://www.moldiscovery.com/.

2.4.8. *GROMACS*

Todos os cálculos de dinâmica molecular foram executados com o pacote de programas *Gromacs*[Lindahl *et al.*, 2001; Berendsen *et al.*, 1995] versão 3.2.1., distribuído gratuitamente no endereço eletrônico http://www.gromacs.org/

2.4.9. *SYBYL*

Todas as minimizações de energia dos ligantes extraídos dos complexos PDB, assim como o alinhamento dos complexos de cisteíno-proteases (com cavidades similares a cruzaína) foram realizados utilizando-se o programa *SYBYL*[SYBYL, 2004] versão 7.0.

2.4.10. *XLOGP*

Os valores de coeficiente de partição calculados para as moléculas do banco de dados *ZINC* foram calculados com o programa *XLOGP*,[Wang *et al.*, 1997] distribuído gratuitamente no endereço eletrônico (ftp://ftp2.ipc.pku.edu.cn/).

2.5. Procedimentos computacionais

2.5.1. Manipulação e alinhamento das estruturas extraídas do *PDB*

Cada uma das treze estruturas da cruzaína resolvidas por cristalografia de raios-X foi obtida diretamente do banco de dados *PDB*, no formato "*.pdb*". As etapas descritas a seguir foram realizadas no programa *DeepView3.7*.[Guex & Peitsch, 1997] As estruturas da cruzaína foram alinhadas átomo-a-átomo, utilizando-se as estruturas originais diretamente como depositadas no *PDB*, sem qualquer procedimento de minimização ou de relaxação da estrutura. O complexo 1ME4 foi usado como referência para o alinhamento das demais estruturas, pois é o complexo de melhor resolução do cristal (1,2 Å). O alinhamento foi feito utilizando-se o comando "*Magic-fit*" e, em seguida, "*Improve-fit*" do programa que busca a minimização dos valores de *RMSD* (*Root Mean Square Distance*), considerando todos os 1592 átomos da estrutura da cruzaína. Não foram considerados para o alinhamento os átomos dos inibidores e as moléculas de água presentes nos complexos. O resultado do alinhamento dos 13 complexos da cruzaína está apresentado na **Figura 28**.

Para o cálculo dos valores de *RMSD,* ou seja, os valores da raiz quadrada da média dos quadrados das distâncias, entre átomos correspondentes foi utilizada a **Equação 1**:

$$RMSD = ((\sum\Delta^2)/\eta)^{1/2}$$ **Equação 1**

onde Δ é o valor da distância entre um par de átomos correspondentes alinhados e η é o número de átomos na molécula.

O resultado do cálculo dos valores de *RMSD*, para cada uma das estruturas da cruzaína depositadas no *PDB* está apresentado na **Tabela 7**.

As estruturas dos inibidores da cruzaína presentes em cada um dos complexos foram extraídas dos respectivos complexos, alinhados como descrito na etapa anterior. Desta forma, os inibidores foram visualizados alinhados dentro do sítio ativo da enzima. O alinhamento dos inibidores da cruzaína foi, posteriormente, utilizado para a preparação do

modelo farmacofórico. O resultado do alinhamento dos inibidores da cruzaína extraídos dos complexos depositados no *PDB* está apresentado na **Figura 30**.

Cada um dos 20 complexos de cisteíno-proteases com cavidades similares à cruzaína foram igualmente obtidos diretamente do banco de dados *PDB*, no formato "*.pdb*". O alinhamento destas estruturas foi realizado no programa *SYBYL*,[Sybyl, 2004] utilizando-se a função "*Analyse>Fit_Atoms*". O alinhamento foi feito, átomo-a-átomo, com a estrutura do complexo 1ME4 como referência, buscando-se a minimização dos correspondentes valores de *RMSD* considerando-se, no entanto, apenas 4 átomos do sítio ativo das cisteíno-proteases, a saber: *CA-Cys25*; *CA-Gly66*; *CA-Asp158*; *CA-His159,* respectivamente. As estruturas dos inibidores alinhados foram extraídas diretamente das estruturas dos correspondentes complexos alinhados, utilizando-se o programa *SYBYL*. O resultado do alinhamento dos 13 complexos da cruzaína está apresentado na **Figura 34**.

A estrutura dos inibidores, extraídos dos complexos da cruzaína, foi modificada para representar os inibidores antes de sofrerem o ataque nucleofílico do enxofre catalítico da *Cys25*. Estas modificações foram feitas no programa *SYBYL*,[Sybyl, 2004] utilizando-se a função "*sketch molecule*" e, em seguida, foi realizada a minimização, de energia apenas na porção da molécula que foi alterada. Esta minimização foi feita com o campo de força *Tripos* e utilizando-se todos os parâmetros *default* do programa. O resultado do alinhamento dos inibidores da cruzaína extraídos dos complexos depositados no *PDB* está apresentado na **Figura 35**.

2.5.2. Obtenção dos campos de interação molecular

Conforme apresentado na *Introdução*, itens **1.4.2** e **1.4.3.**, o reconhecimento das propriedades do sitio de ligação de um alvo biológico pode ser feito a partir dos campos de interação molecular (*molecular interaction fields – MIF*),[Goodford, 1985] determinados ao redor da estrutura do alvo biológico.

A obtenção dos valores dos campos de interação molecular foi feita com o programa *GRID.*[Harrison *et al.*, 1997] A estrutura da cruzaína, como depositada no complexo 1ME4 no formato "*.pdb*", é o arquivo de entrada para o programa. Em seguida, são selecionados os átomos (ou moléculas) que serão utilizados como sondas moleculares para determinação dos campos de interação molecular. Foram selecionadas as seguintes sondas: "H_2O";

"*Carbono aromático*", "*Oxigênio carbonílico*" e "*Nitrogênio de amida*". Utilizando-se um espaçamento entre as linhas da grade no valor de 0,5 Å (*Keyword* "*NPLA* = 2") e uma caixa com as seguintes coordenadas *X*; (87; 113); *Y* (10; 31); *Z* (-26; 3). Os valores dos campos de interação molecular, computados para cada sonda, são visualizados no programa *SYBYL*, que permite a escolha do valor do nível do contorno isoenergético que se deseja visualizar os resultados. Os resultados obtidos correspondentes aos campos de interação molecular no sítio ativo da cruzaína estão apresentados na **Figura 31**.

2.5.3. Procura por cavidades similares à cruzaína

Conforme apresentado na *Introdução*, item **1.4.3.**, as propriedades de reconhecimento molecular da cavidade do sitio de ligação podem ser analisadas através de pseudocentros.[Schmitt *et al.*, 2001] Estes são pontos na superfície da cavidade, que representam as propriedades físico-químicas dos resíduos que compõem esta cavidade e que são expostas na superfície.

A procura por cavidades similares à cruzaína foi realizada utilizando-se o programa *CAVBASE*,[Schmitt *et al.*, 2002] que é um módulo do programa *RELIBASE+*.[Gohlke *et al.*, 2000a] O programa *CAVBASE* contém um banco de dados, pré-calculado, com todas as cavidades de todas as proteínas depositadas no *PDB*. As propriedades de reconhecimento molecular expostas na superfície de cada uma destas cavidades são descritas através de pseudocentros. A partir dos pseudocentros identificados na cavidade do sítio ativo da cruzaína, como depositada no complexo 1ME4, foi realizada uma busca no banco de dados do programa *CAVBASE* por cavidades similares à cruzaína e que contenham um inibidor em seu interior. O resultado da busca é classificado pelo programa e apresentado em ordem decrescente de similaridade. As estruturas dos complexos com cavidades similares à cruzaína podem ser visualizadas no próprio programa *CAVBASE*.

Os resultados, representando todos os pseudocentros identificados no sítio ativo da cruzaína, e os pseudocentros utilizados na busca por cavidades similares estão apresentados nas **Figura 32** e **Figura 33**, respectivamente. Uma lista dos complexos com cavidades similares à cruzaína identificados pelo programa *CAVBASE* está apresentada na Tabela 8.

2.5.4. Dinâmica molecular

Conforme discutido na *Introdução* desta tese, item **1.4.3.**, uma das estratégias para se considerar tanto a flexibilidade da proteína, quanto a possibilidade da ocorrência de ajuste induzido na conformação da proteína devido à presença de um ligante é uma das aplicações de protocolos de dinâmica molecular.

Os procedimentos de dinâmica molecular foram realizados com o pacote de programas *Gromacs3.2.1*,[Lindahl *et al.*, 2001; Berendsen *et al.*, 1995] utilizando-se o campo de força G43a1, distribuído em conjunto com o programa. As coordenadas da cruzaína, extraídas do complexo 1ME4 (resolução do cristal = 1,2Å), foram utilizadas como ponto de partida para as simulações de dinâmica molecular. Foram realizadas as simulações da cruzaína sem o ligante e da cruzaína com o ligante do complexo 1ME4. Seguindo procedimentos descritos na literatura,[Sotriffer *et al.*, 2004] a estrutura inicial da proteína foi submetida a minimização da respectiva energia, no vácuo, tendo sido realizados 20 passos de minimização pelo método *steepest descent*, seguido de 180 passos de minimização pelo método *conjugate gradient*. A seguir, o sistema foi posicionado no centro de uma caixa cúbica, com lados de 80 Å e, a seguir, esta caixa foi solvatada com moléculas de água. Para o sistema da proteína sem ligante, foram adicionadas, automaticamente, pelo programa 16.110 moléculas de água, enquanto que para o sistema com a proteína complexada com o ligante, foram adicionadas 16.075 moléculas de água. Após a etapa de solvatação foram adicionados, nos dois sistemas respectivamente, 12 íons de sódio com carga +1, para garantir a neutralidade de cargas do sistema (a cruzaína apresenta uma carga de –12 devido ao maior número de resíduos carregados negativamente). Subseqüentemente, as moléculas de água e os íons sódio foram minimizados quanto às suas energias, com 20 passos pelo método *steepest descent*, seguido de 180 passos pelo método *conjugate gradient*. A seguir, a dinâmica molecular teve início com o aquecimento do solvente, por um período de 20 ps, até à temperatura de 300 K, seguida de um resfriamento, por um período de 5 p até à temperatura de 100 K, mantendo-se fixos os átomos da proteína.

Depois desta etapa, todo o sistema (solvente e proteína) foi gradualmente aquecido até a temperatura de 300 K, por um período de 25 ps, e mantido nesta temperatura por mais 50 ps para garantir que o equilíbrio do sistema seja atingido. A simulação foi, então,

conduzida por 3000 ps (3 ns), à temperatura e pressão constantes (NPT – usando algoritmo de *Berendsen*[Berendsen *et al.*, 1984]), aplicando-se condições periódicas de contorno. As simulações das ligações de hidrogênios covalentes foram delimitadas pelo algoritmo *Shake*.[Ryckaert *et al.*, 1977] O intervalo de tempo usado foi de 2 fs. Interações do tipo *Van der Waals* foram limitadas à distância de 9 Å, enquanto que, as interações eletrostáticas foram tratadas pelo método *PME* (*Particle Mesh Ewald*).[MacKerell, 2001] Os valores das coordenadas atômicas para toda a trajetória, utilizados na análise dos resultados, foram registrados a cada 0,5 ps. Os gráficos dos valores de *RMSD* em função do tempo e em função das distâncias interatômicas selecionadas em função do tempo, foram gerados a partir da trajetória de 3 ns. Os resultados dos procedimentos de dinâmica molecular estão apresentados de modo gráfico nas **Figura 36** e **Figura 37** e estão resumidos nas **Tabela 9** e **Tabela 10**.

2.5.5. Preparação e aplicação do modelo farmacofórico (Modelo I e II)

Conforme discutido na *Introdução* desta tese, item **1.4.**, a estratégia de *virtual screening*, utilizada na descoberta de compostos bioativos, consiste na aplicação seqüencial de diferentes filtros virtuais em bibliotecas de compostos. Uma discussão detalhada sobre modelos farmacofóricos e suas aplicações como filtros para a seleção de compostos bioativos está apresentada no item **1.4.2.**.

Todas as etapas de preparação e de aplicação do modelo farmacofórico foram realizadas com o programa *CATALYST*.[Catalyst, 2005]

Com o objetivo de poder aplicar o modelo farmacofórico sobre o banco de dados *ZINC*, primeiramente realizou-se a conversão dos compostos presentes neste banco de dados para o formato "*.bdb*" utilizado pelo programa *CATALYST*. Para tanto, os 3.294.741 compostos do banco de dados *ZINC* foram inicialmente convertidos do formato unidimensional, "*.smiles*", para o correspondente formato tridimensional "*.sdf*", em uma conformação de mínimo de energia, utilizando-se o programa *CORINA*. Em seguida, o programa *CATALYST* gerou, aleatoriamente, 50 conformações diferentes para cada composto do banco de dados *ZINC* criando-se, então, um outro banco de dados com 3.288.893 compostos (alguns compostos foram rejeitados pelo programa *CATALYST*).

A construção do modelo farmacofórico foi feita a partir da estrutura do inibidor da cruzaína, extraído do complexo 1ME4. Inicialmente, o programa reconheceu as propriedades físico-químicas da molécula, como por exemplo grupamentos doadores/aceptores de ligação de hidrogênio, grupamentos hidrofóbicos ou grupamentos com cargas. A seguir, foram indicadas pelo usuário quais destas propriedades compõem o modelo farmacofórico e, para cada propriedade selecionada, é atribuído um valor de raio de tolerância para sua localização. Este modelo farmacofórico putativo é, a seguir, testado sobre os demais inibidores da cruzaína (em suas respectivas conformações extraídas dos complexos alinhados) e sobre os compostos do banco de dados *ZINC*. Este procedimento é repetido, interativamente, modificando-se, gradualmente, as propriedades fisico-químicas e as restrições espaciais do modelo farmacofórico com o objetivo de se maximizar tanto o número de compostos de referência aceitos pelo modelo, bem como o número de compostos do banco de dados *ZINC* rejeitados pelo modelo.

Representações esquemáticas dos dois modelos farmacofóricos construídos estão apresentados, respectivamente, na **Figura 39** (**Modelo I**) e nas **Figura 45** e **Figura 47** (**Modelo II**)

2.5.6. Filtros de propriedades físico-quimicas (Modelo I e II)

Uma discussão detalhada sobre filtros virtuais de compostos, baseados em propriedades físico-químicas está apresentada no item **1.4.1.** da *Introdução* desta tese.

A aplicação dos filtros de propriedades físico-químicas foi feita utilizando-se o programa *DBFilter*.[Wang, 2003] Para esta etapa, o banco de dados *ZINC* foi, inicialmente, convertido do formato unidimensional "*.smiles*" para o correspondente formato tridimensional "*.mol2*". Todas as moléculas, no formato *MOL2,* foram transferidas para um único arquivo, que serviu de entrada para o programa *DBFilter*. Os valores do coeficiente de partição (logP) de todas as moléculas do banco de dados *ZINC* foram calculados pelo programa *XLOGP*.[Wang *et al.*, 1997] Este programa calcula os valores do logP a partir da contribuição atômica e de fatores de correção para interações intramoleculares e não considera a contribuição da ionização para a partição. A seguir, os valores de logP foram registrados em um único arquivo, que serviu de entrada para o programa *DBFilter*. Para a aplicação dos filtros de propriedades físico-químicas, os valores de máximo e de mínimo

tolerados são registrados em um arquivo de entrada para o programa *DBFilter*, que, ao ser executado, separa, automaticamente, as moléculas em "rejeitadas" ou "aceitas", de acordo com os critérios indicados. Os valores de máximo e de mínimo, utilizados como critérios para a seleção dos compostos nos **Modelos I** e **II**, estão respectivamente apresentados nas **Tabela 11** e **Tabela 15**.

2.5.7. *Docking* (Modelo I e II)

Uma discussão detalhada sobre filtros de *docking* aplicados para a seleção de compostos bioativos está apresentada no item **1.4.3.** da *Introdução* desta tese.

Todos os procedimentos de *docking* foram executados com o programa *GOLD*.[Verdonk *et al.*, 2003] O programa de *docking GOLD* utiliza um algoritmo genético para realizar buscas no espaço conformacional/rotacional/translacional de um determinado ligante, com o objetivo de encontrar as poses de mínimo de energia entre este ligante e uma determinada macromolécula biológica. A função de energia *GOLDSCORE*, utilizada pelo programa para avaliar a afinidade entre o ligante e seu alvo, é baseada em mecânica molecular e atribui um valor de *Fitness* para cada pose encontrada. Valores maiores de *Fitness* indicam maior afinidade entre o ligante e a proteína. Esta função de energia é composta de quatro termos, para descrever, respectivamente, as ligações de hidrogênio intramoleculares, as ligações de hidrogênio intermoleculares, as interações de *Van der Waals* e, finalmente, a energia interna do ligante.

Para o procedimento de *docking* utilizando-se o programa *GOLD*, as estruturas tanto da enzima alvo quanto as dos ligantes a serem docados são preparadas no formato *MOL2*, considerando-se todos os hidrogênios (polares e apolares), mas sem considerar os valores de carga parcial atômica. Os átomos constituintes do sítio ativo da cruzaína foram definidos considerando-se todos os átomos em um raio de 20 Å, a partir do átomo de enxofre da *Cys25*. Os parâmetros do algoritmo genético são ajustados, automaticamente, pelo programa de acordo com o tamanho do ligante, para se obter o máximo de eficiência do *docking*. A seguir, cada ligante é, consecutivamente, docado 10 vezes sobre a estrutura da enzima. Para a análise dos resultados do *docking* de cada ligante, considera-se, apenas, a estrutura da pose com o maior valor de *Fitness* entre os 10 *dockings* realizados. As poses resultantes dos procedimentos de *docking* podem ser analisadas em qualquer programa de

visualização de estruturas em três dimensões, como por exemplo o programa *DEEPVIEW.*[Guex & Peitsch, 1997]

Os resultados da validação dos procedimentos de *dockings* para os **Modelos I** e **II** estão apresentados na **Figura 40** (**Modelo I**) e na **Figura 48** (**Modelo II**), respectivamente. Os resultados da aplicação dos procedimentos de *dockings* para os **Modelos I** e **II** estão apresentados na **Figura 41** (**Modelo I**) e **Figura 49** (**Modelo II**) , respectivamente.

2.6. Teste de inibição da cruzaína (Modelo I)

Conforme discutido na *Introdução* desta tese, item **1.4.**, os modelos de *virtual screening* devem ser validados experimentalmente para a confirmação de sua utilidade na busca de compostos bioativos. Neste caso, modelos de *virtual screening* de inibidores da cruzaína foram validados através do teste de inibição enzimática dos compostos sugeridos pelos modelos.

Os testes de inibição da cruzaína foram realizados nos laboratórios do Prof. Dr. Luiz Juliano, Departamento de Biofísica, Universidade Federal do Estado de São Paulo-SP; do Prof. Dr. Willem Baader do Departamento de Química Fundamental do Instituto de Química – Universidade de São Paulo – SP e da Profª. Drª. Shirley Schreier do Departamento de Bioquímica do Instituto de Química – Universidade de São Paulo – SP.

Este procedimento foi utilizado para os compostos selecionados pelo **Modelo I** (item **3.3.**) de inibidores da cruzaína.

Seguindo-se procedimento de *Tight-Binding-Titration,*[Gillmor *et al.*, 1997] modificado de acordo com protocolo de rotina implementado nos laboratórios do Prof. Dr. L. Juliano, UNFESP, SP,[McGovern *et al.*, 2003] em uma cubeta de quartzo de 1,4 mL e caminho óptico de 10 mm, foram adicionados 1000 μL de uma solução tampão fosfato dissódico (50 mM) e *EDTA* (2,5 mM), pH=6,5; 0,3 μL de uma solução de cruzaína (1,8 μM) e 1,0 μL de uma solução do agente redutor de grupos tiol *DTT* (1,0 M). Esta mistura é mantida em repouso, por 5 minutos, para a pré-incubação e a ativação da enzima. A reação é iniciada com a adição de 5,0 μL de uma solução do substrato fluorescente Z-FR-MCA (125 μM). O valor da velocidade de consumo do substrato pela cruzaína é registrado em um fluorímetro, equipado com agitação magnética e banho termostatizado (37,0 ± 0,1 °C), emitindo radiação em 380 nm e registrando-se a fluorescência em 460 nm. Em seguida,

alíquotas de uma solução do inibidor são adicionadas à cubeta e a velocidade da reação registrada, até que se observe ou a inibição total da enzima (*i.e.* valor da velocidade de consumo do substrato igual a zero) ou a perda da solubilidade do inibidor na cubeta. Soluções estoque dos inibidores, todas na concentração de 0,1 M, foram preparadas em DMSO. As concentrações de DMSO na cubeta nunca foram maiores que 5% (v/v) e os resultados foram controlados para o efeito do DMSO. Ao final, o valor da constante de inibição aparente (K_i^{ap}) são obtidos a partir do ajuste dos parâmetros da equação de *tight binding* (**Equação 2**) aos valores de velocidade de consumo do substrato em função da concentração do inibidor. Finalmente, os valores da constante de inibição (K_i) são obtidos a partir dos valores da constante de inibição aparente (K_i^{ap}) utilizando-se a **Equação 3**:

$$\mathbf{V = SA\ (\ \mathit{E_0} - 0{,}5\ \{(\mathit{E_0} + \mathit{I} + K_i^{ap}) - [(\mathit{E_0} + \mathit{I} + K_i^{ap})^2 - 4\mathit{E_0 I}]^{1/2}\})} \qquad \textbf{Equação 2}$$

onde **V** é o valor da velocidade de consumo do substrato; **SA** é uma constante de proporcionalidade entre a velocidade de consumo do substrato e a unidade de concentração da enzima; $\boldsymbol{E_0}$ é o valor da concentração da enzima; $\boldsymbol{I}$ é o valor da concentração do inibidor; e $\mathbf{K_i^{ap}}$ é o valor da constante de inibição aparente [Gillmor *et al.*, 1997].

$$\mathbf{K_i = K_i^{ap} / [1 + (S / K_m)]} \qquad \textbf{Equação 3}$$

onde $\mathbf{K_i}$ é o valor da constante de inibição; $\mathbf{K_i^{ap}}$ é o valor da constante de inibição aparente; **S** é o valor da concentração do substrato; e $\mathbf{K_m}$ é o valor da constante de Michaelis-Menten (medida da afinidade da enzima pelo substrato) [Gillmor *et al.*, 1997].

Os gráficos dos testes de inibição da cruzaína, mostrando a velocidade de consumo do substrato pela enzima, na presença de diferentes concentrações de cada um dos compostos selecionados pelo **Modelo I** estão apresentados na Figura 42.

2.7. Verificação do mecanismo promíscuo de inibição enzimática (Modelo I)

Conforme discutido na *Introdução* desta tese, item **1.4.**, os modelos experimentais de atividade biológica *in vitro* são sujeitos a artefatos experimentais que podem levar a resultados falso-positivos. Uma das causas conhecidas de resultados falso-positivos é a possibilidade da ocorrência de inibição enzimática através do mecanismo promíscuo.[McGovern *et al.*, 2002]

A possibilidade da ocorrência de inibição da cruzaína, através do mecanismo promíscuo, foi verificada avaliando-se a sensibilidade dos inibidores na presença de um detergente não-iônico durante o teste de inibição.[Seidler *et al.*, 2003; McGovern *et al.*, 2003] Nos casos em que foi observada a perda da inibição, ou seja, a retomada do valor original da atividade da cruzaína após a adição do detergente, foi sugerida a ocorrência de inibição pelo mecanismo promíscuo. Esta hipótese foi, a seguir, confirmada através de outro teste utilizando-se, neste caso, a enzima em uma concentração 10 vezes maior.[McGovern & Shoichet, 2003]

Assim, procedendo-se de acordo com o procedimento de teste de inibição enzimática da cruzaína descrito anteriormente, nos casos em que foi observada a inibição da cruzaína, foram adicionados 10,0 μL de uma solução 1% (v/v) de *Triton-X100* (um detergente não-iônico). A seguir, o valor da velocidade de consumo do substrato pela cruzaína foi registrado e, então, este valor foi comparado graficamente com o valor da velocidade de consumo do substrato antes da adição do detergente. Se a atividade da enzima retornasse aos valores próximos aos observados para a enzima sem o inibidor, então, se assume que a inibição enzimática está ocorrendo através do mecanismo promíscuo.

Na **Figura 42**, estão apresentados os gráficos de inibição enzimática, mostrando o consumo de substrato pela cruzaína, na presença de diferentes concentrações de cada um dos compostos testados e o efeito da adição de *Triton X-100*, nos testes em que houve inibição parcial da atividade da cruzaína.

Para confirmar esta hipótese de inibição pelo mecanismo promíscuo, um outro teste de inibição enzimática foi executado nas mesmas condições descritas anteriormente, utilizando-se, entretanto, neste caso uma concentração da enzima 10 vezes maior. Sendo assim, em uma cubeta de quartzo de 1,4 mL e caminho óptico de 10 mm, foram adicionados 1000 μL de uma solução tampão fosfato dissódico (50 mM) e *EDTA* (2,5 mM), pH=6,5; 3,3 μL de uma solução de cruzaína (1,8 μM) e 1,0 μL de uma solução do agente redutor de grupos tiol *DTT* (1,0 M). Esta mistura foi mantida em repouso, por 5 minutos, para a pré-incubação e a ativação da enzima. A reação é iniciada com a adição de 5,0 μL de uma solução do substrato fluorescente Z-FR-MCA (125 μM). O valor da velocidade de consumo do substrato pela cruzaína é registrado em um fluorímetro, equipado com agitação

BIBLIOTECA
INSTITUTO DE QUÍMICA
Universidade de São Paulo

magnética e banho termostatizado (37,0 ± 0,1 °C), emitindo radiação em 380 nm e registrando-se a fluorescência em 460 nm. Em seguida, uma alíquota de uma solução do inibidor na mesma concentração em que foi observada a inibição da enzima no teste anterior (com a enzima em uma concentração 10 vezes menor), é adicionada à cubeta e a velocidade de consumo do substrato é registrada. Se a inibição da enzima for significativamente menor que no teste com a enzima em uma concentração 10 vezes menor, então o mecanismo promíscuo de inibição enzimática é confirmado.

Na **Figura 43**, estão apresentados os gráficos de inibição enzimática, mostrando o consumo de substrato pela cruzaína na presença dos inibidores suspeitos de agirem segundo o mecanismo promíscuo, em duas concentrações diferentes da enzima (0,18 nM e 1,8 nM),.

2.8. Teste de inibição da cruzaína na presença de detergente (Modelo II)

Este procedimento foi utilizado com os compostos selecionado pelo **Modelo II** (item **3.4.**) de inibidores da cruzaína e é análogo ao procedimento realizado para o **Modelo I**, descrito no item **2.6.** A única mudança (sublinhada no texto) foi a adição do detergente *Triton X-100* antes do início da reação.

Seguindo-se procedimento de *Tight-Binding-Titration,*[Gillmor *et al.*, 1997] modificado de acordo com protocolo de rotina implementado nos laboratórios do Prof. Dr. L. Juliano, UNFESP, SP,[McGovern *et al.*, 2003] em uma cubeta de quartzo de 1,4 mL e caminho ótico de 10mm, foram adicionados 1000 μL de uma solução tampão fosfato dissódico (50mM) e *EDTA* (2,5mM), pH=6,5; 0,3 μL de uma solução de cruzaína (1,8 μM), 1,0 μL de uma solução do agente redutor de grupos tiol *DTT* (1,0 M) e 10 μL de uma solução 1% (v/v) de *Triton-X100.* Esta mistura é mantida em repouso, por 5 minutos, para a pré-encubação e a ativação da enzima. A reação é iniciada com a adição de 5,0 μL de uma solução do substrato fluorescente Z-FR-MCA (125 μM). O valor da velocidade de consumo do substrato pela cruzaína é registrado em um fluorímetro, equipado com agitação magnética e banho termostatizado (37,0 ± 0,1 °C), emitindo radiação em 380 nm e registrando-se a fluorescência em 460 nm. Em seguida, alíquotas de uma solução do inibidor são adicionadas à cubeta e a velocidade da reação registrada, até que se observe ou a inibição total da enzima (*i.e.* valor da velocidade de consumo do substrato igual a zero) ou

a perda da solubilidade do inibidor na cubeta. Soluções estoque dos inibidores, todas na concentração de 0,1M, foram preparadas em DMSO. As concentrações de DMSO na cubeta nunca foram maiores que 5% (v/v) e os resultados foram controlados para o efeito do DMSO. Ao final, o valor da constante de inibição aparente (K_i^{ap}) são obtidos a partir do ajuste dos parâmetros da equação de *tight binding* (**Equação 2**) aos valores de velocidade de consumo do substrato em função da concentração do inibidor. Finalmente, os valores da constante de inibição (K_i) são obtidos a partir dos valores da constante de inibição aparente (K_i^{ap}) utilizando-se a **Equação 3**.

Os gráficos dos testes de inibição da cruzaína, mostrando a velocidade de consumo do substrato pela enzima, na presença de diferentes concentrações de cada um dos compostos selecionados pelo **Modelo II** estão apresentados nas **Figura 52-Figura 70**.

3. Resultados e discussão

3.1. Introdução

O estudo de moléculas bioativas se estende a diferentes áreas de aplicação, incluindo o planejamento racional de novos fármacos ou de defensivos agrícolas, o estudo de seus mecanismos de ação, a previsão da toxicidade e o controle ambiental.[Kubinyi, 2005] Diferentes estratégias metodológicas complementares podem ser aplicadas iterativamente, focando-se em informações extraídas do sistema químico e/ou do sistema biológico e/ou da interação entre ambos. As informações do sistema químico podem ser obtidas do estudo das propriedades físico-químicas-moleculares de compostos bioativos, como por exemplo: (*i*) a determinação dos valores do coeficiente de partição (logP);[Rahman *et al.*, 2003] (*ii*) o cálculo de valores de energia HOMO/LUMO[Liu *et al.*, 2003]; (*iii*) estudos conformacionais[Perola & Charifson, 2004]; (*iv*) estudos espectroscópicos;[Pires *et al.*, 2001] e (*v*) estudos de reatividade.[Rishton, 1997] Informações do sistema biológico podem ser obtidas, entre outros: (*i*) de dados de cristalografia de raios-X ou de ressonância magnética nuclear de proteínas;[Kubinyi, 1998c] (*ii*) de estudos do metabolismo e síntese de biomoléculas;[Kubinyi, 1995] e (*iii*) de estudos para a descoberta e validação de alvos moleculares.[Wang, 1997] As informações sobre a interação entre os sistemas químico e biológico podem ser obtidas tanto do efeito da ação do composto bioativo sobre o sistema biológico, bem como da ação do sistema biológico sobre o composto bioativo, sendo que cada uma pode ser obtida de maneira isolada ou simultânea. Assim, por exemplo: (*i*) os valores da constante de afinidade (K_i) do ligante-receptor fornecem informações do efeito do composto sobre o sistema biológico;[Ritonja *et al.*, 1996] (*ii*) os dados de ADME fornecem informações do efeito do sistema biológico sobre o composto;[Eddershaw *et al.*, 2000] e (*iii*) valores de IC_{50}, determinados em um sistema *in vivo*, fornecem informações tanto do efeito da ação do composto bioativo sobre o sistema biológico, como da ação do sistema biológico sobre o composto bioativo.[Doyle *et al.*, 2007]

O <u>*Laboratório de QSAR e Modelagem Molecular de Compostos Bioativos*</u> do Instituto de Química da Universidade de São Paulo, liderado pela Prof[a]. Associada Antonia Tavares do Amaral, vem se dedicando, nos últimos 20 anos, ao estudo de diversos aspectos

de compostos bioativos. Entre eles, destacam-se estudos de Relações Quantitativas entre Estrutura química e Atividade biológica, incluindo-se aqueles envolvendo modelagem molecular (*QSAR/QSAR-3D*) e estudos para a determinação de propriedades físico-químicas e estruturais (como lipofilicidade e variação da densidade eletrônica do grupo carbonila), de compostos com atividade biológica, em especial antiarrítmicos e anestésicos locais;[Siqueira, 2001; Malvezzi, 2000; Tavares & Amaral, 1997; Sousa, 1997; do Amaral *et al.*, 1997; Tavares, 1987] antitumorais;[Raminelli, 2001] antibacterianos;[Pires, 1998; Tavares, 1993; Baroni, 1987] [Pires *et al.*, 2001] anti-hiv [Ishiki *et al.*, 2001; Ishiki, 1999]; tuberculostáticos [Ishiki, 2005] e antichagásicos [Malvezzi *et al.*, 2008]][Scotti, 2005]. Estes estudos foram planejados visando-se contribuir para elucidar o mecanismo de ação destes compostos e, quando possível, para prever derivados mais potentes e/ou menos tóxicos.

Além dos dados de atividade biológica, determinados no próprio grupo, ou retirados da literatura, os trabalhos desenvolvidos no *Laboratório de QSAR e Modelagem Molecular de Compostos Bioativos*, consideraram, até esta tese de doutorado, apenas informações do sistema químico, ou seja, dos parâmetros físico-químicos-moleculares determinados para as séries de compostos estudadas, no planejamento de compostos bioativos. Esta estratégia, também conhecida como *Ligand Based Drug Design,* ou desenvolvimento de fármacos com base no ligante, é aplicada quando não se considera a estrutura do alvo biológico. Nesta estratégia, o estudo do mecanismo de ação de determinado composto ou o mapeamento de seu receptor biológico é feito de modo indireto, ou seja, a partir da variação estrutural dos compostos da série em estudo e da variação da atividade biológica da série correspondente. Sendo assim, o sucesso desta estratégia no planejamento de compostos bioativos dependerá: (*i*) da qualidade dos dados de atividade biológica da série; (*ii*) da escolha e da determinação dos parâmetros físico-químicos usados como descritores da série; e (*iii*) do modelo estatístico proposto para ser usado para correlacionar os dados de atividade com os descritores físico-químicos.

Mais recentemente, com a disponibilidade de livre e amplo acesso a dados do sistema biológico, disponíveis, entre outros, no banco de dados *PDB,* e com a disponibilidade no *Laboratório de QSAR e Modelagem Molecular de Compostos Bioativos* de recursos de *hardware* e de *software* para manipulação e tratamento de grande número de

dados, foi possível incluir no desenvolvimento de compostos bioativos considerações sobre o sistema biológico, em especial, com relação à estrutura do alvo molecular e suas interações com os ligantes. Esta estratégia, denominada *Structure Based Drug Design,* ou desenvolvimento de fármacos com base no receptor, é complementar à estratégia anterior e permite o desenvolvimento de ligantes específicos para um determinado alvo molecular, explorando as características estruturais específicas daquele alvo.

No presente trabalho, desenvolvido no âmbito de um projeto de doutoramento no Instituto de Química da Universidade de São Paulo, sob a orientação da Profª. Associada Antonia Tavares do Amaral, foram implementadas e aplicadas estratégias de *Structure Based Drug Design,* até então inéditas ao *Laboratório de QSAR e Modelagem Molecular de Compostos Bioativos*. As estratégias implementadas foram: (*i*) a determinação de campos de interação molecular de proteínas (ver *Introdução* item **1.4.2.**), (*ii*) o *docking* de moléculas orgânicas pequenas no sítio de ligação de proteínas (ver *Introdução* item **1.4.3.**), (*iii*) a busca por similaridades em bancos de dados de proteínas (ver *Introdução* item **1.4.3.**) a elaboração de modelos farmacofóricos em três dimensões (ver *Introdução* item **1.4.2.**) e (*iv*) a dinâmica molecular de proteínas ou de complexos proteína-ligante (ver *Introdução* item **1.4.3.**). Estas estratégias foram aplicadas, junto com outras metodologias já implementadas no grupo, para a descoberta de novos inibidores da cruzaína, uma cisteíno-protease do *Trypanosoma cruzi*, o agente etiológico da doença de Chagas (ver *Introdução* item **1.7.3.**). Este trabalho está inserido no projeto Temático FAPESP Proc. 2001/01192-3 "Antichagásicos Potenciais, Derivados de Compostos Nitro-Heterocíclicos", coordenado pela Profa. Dra. Elizabeth Igne Ferreira da Faculdade de Ciências Farmacêuticas da Universidade de São Paulo e faz parte do sub-projeto "Abordagens Metodológicas Modernas Empregadas no Planejamento de Compostos Bioativos, Antichagásicos Potenciais: *QSAR/QSAR-3D*", coordenado pela Profª. Drª. Antonia Tavares do Amaral.

Neste relato introdutório do item **3.** - *Resultados e Discussão* - será apresentada, nos próximos dois parágrafos, uma breve narrativa histórica das etapas anteriores aos resultados apresentados nesta tese, sem a preocupação de apresentar e discutir, em detalhes, os resultados obtidos nestas etapas preliminares. Pretende-se com este relato introdutório, apenas, fornecer subsídios para que o leitor possa compreender as etapas que antecederam a

escolha dos sistemas que resultaram nos modelos de inibidores da cruzaína apresentados nos itens **3.3** e **3.4**.

No projeto originalmente proposto, para esta tese de doutorado pretendia-se aplicar metodologias usuais em *QSAR/QSAR-3D* para o estudo das características físico-químicas e estruturais de séries de derivados de fenilidrazonas (**Figura 26**), com potencial atividade antichagásica,[Li *et al.*, 1996] com o objetivo de contribuir para o entendimento da interação ligante-receptor e para propor modelos QSAR/QSAR-3D capazes de prever compostos mais ativos e menos tóxicos. Com este objetivo, foi preparada uma série de sete fenilidrazonas estruturalmente semelhantes que, a seguir, foram testadas como inibidores da cruzaína.[Malvezzi *et al.*, 2005] Os sete compostos testados apresentaram inibição da cruzaína (valores de K_i variaram entre 0,1 e 0,10 μM), entretanto, foi verificado que pelo menos um deles agia de acordo com o mecanismo promíscuo de inibição enzimática.[McGovern *et al.*, 2003] Estes resultados falso-positivos foram corroborados com relatos de literatura,[McGovern *et al.*, 2002] indicando que *bis*-fenilidrazonas - estruturalmente semelhantes à série preparada - também agem segundo o mecanismo promíscuo de inibição enzimática e levaram a mudança da abordagem, originalmente, proposta nesta tese que visa à busca de novos antichagásicos. Assim, considerando-se a disponibilidade no *PDB*, da estrutura cristalina da cruzaína complexada com diversos inibidores e a disponibilidade de bibliotecas virtuais de compostos orgânicos, potencialmente *drug-like* (ver *Introdução* itens **1.1.** e **1.4.1.**), foi proposta a busca de novas estruturas líder, agora, utilizando estratégias de *virtual screening* (ver *Introdução* item **1.4.**), baseadas em informações tanto da estrutura da cruzaína cristalizada (*structure based drug design*), como da estrutura dos inibidores complexados com a enzima (*ligand based drug design*).

Os modelos de *virtual screening* propostos, ainda nas etapas preliminares desta tese de doutorado, foram aplicados sobre bibliotecas de compostos, relativamente pequenas (< 300.000 compostos) incluindo-se: o banco de dados *Maybridge*[Maybridge Chemicals Ltd.,] (49.916 compostos),[Malvezzi *et al.*, 2004] o banco de dados *NCI*[NCI-NIH, 2005] (247.617 compostos),[Malvezzi *et al.*, 2006] e o banco de dados *Tocris*[Tocris, 2005] (824 compostos). Estes modelos, ainda preliminares, foram avaliados virtualmente empregando-se, apenas, os próprios inibidores da cruzaína usados para gerar o modelo. Os compostos sugeridos por estes modelos não foram submetidos à validação experimental, ou seja, ao

teste de inibição enzimática da cruzaína. Tais modelos preliminares foram considerados como exercícios de implementação e aplicação das ferramentas de *virtual screening* e para o domínio das potencialidades e limitações das estratégias recém-implantadas no grupo.

Os resultados que serão apresentados nesta tese, a seguir, se referem às etapas de reconhecimento da estrutura da cruzaína e de seu sítio de ligação e às etapas de preparação e validação experimental de dois modelos de *virtual screening* de inibidores da cruzaína (**Modelo I** e **Modelo II**), que foram aplicados sobre a biblioteca de compostos *ZINC*[Irwin & Shoichet, 2005] (3.294.741 compostos). Estes resultados serão apresentados, nos itens **3.2 à 3.4.**, a saber:

3.2. Reconhecimento da estrutura da cruzaína e características do sítio ativo. Neste item, são apresentados os resultados do reconhecimento e preparação da estrutura da cruzaína, para serem usados na proposição dos modelos de *virtual screening*. Este reconhecimento foi feito por: (*i*) inspeção visual; (*ii*) através de campos de interação molecular, calculados pelo programa *GRID*,[Goodford, 1985] e (*iii*) por propriedades de interação molecular na superfície da cavidade (pseudocentros), identificadas pelo programa *CAVBASE;*[Schmitt *et al.*, 2002] e (*iv*) por dinâmica molecular.[Lindahl *et al.*, 2001]

3.3. Obtenção e validação de modelo de *virtual screening* de inibidores da cruzaína - Modelo I. Neste ítem, são apresentados os resultados da obtenção e validação do **Modelo I**, proposto a partir das estruturas dos 13 complexos da cruzaína, depositadas no *PDB*. Os filtros, respectivamente, físico-químico, farmacofórico, de *docking* e de seleção por inspeção visual foram aplicados, nesta seqüência, sobre a biblioteca de compostos *ZINC*[Irwin & Shoichet, 2005], sendo que cada filtro foi individualmente validado utilizando-se todas as estruturas dos 13 complexos extraídos do *PDB*. Finalmente, são apresentados os resultados de validação experimental dos compostos selecionados pelo **Modelo I** e os resultados da verificação do mecanismo promíscuo [Malvezzi *et al.*, 2008; McGovern *et al.*, 2002]

3.4. Obtenção e validação de modelo de *virtual screening* de inibidores da cruzaína - Modelo II. Neste item, são apresentados os resultados da obtenção e da validação do **Modelo II**, proposto a partir das estruturas dos 33 complexos de cisteíno-proteases depositadas no *PDB* (13 complexos da cruzaína e 20 complexos de outras cisteíno-proteases, identificadas na busca por cavidades similares à cruzaína). Os filtros, respectivamente, farmacofórico, físico-químico, de *docking* e de seleção por inspeção visual, foram aplicados, nesta seqüência, sobre a biblioteca de compostos *ZINC*[Irwin & Shoichet, 2005], sendo que cada filtro foi, individualmente, validado utilizando-se todas as estruturas dos 33 complexos extraídos do *PDB*. Finalmente, são apresentados os resultados de validação experimental dos compostos selecionados pelo **Modelo II** [McGovern *et al.*, 2003].

3.2. Reconhecimento da estrutura da cruzaína e características do sítio ativo

Como discutido na *Introdução* desta tese (item **1.4.3.**), aspectos como: (*i*) o reconhecimento da função da enzima e seu mecanismo catalítico; (*ii*) a identificação da cavidade de ligação; (*iii*) a conformação; (*iv*) a flexibilidade; (*v*) a presença de moléculas de água conservadas; e (*vi*) a resolução da proteína são fundamentais para o sucesso de estratégias de *Structure Based Drug Design*.

O alvo biológico escolhido neste trabalho é uma cisteíno-protease do *T.cruzi*, a cruzipaína. A forma recombinante da cruzipaína, a cruzaína, foi expressa na bactéria *E. coli*[Lima *et al.*, 2001] e sua estrutura cristalina foi resolvida por cristalografia de raio-X.[Serveau *et al.*, 1996; Mcgrath *et al.*, 1995; Ring *et al.*, 1993; Meirelles *et al.*, 1992; Goodford, 1985] Existem hoje, pelo menos 13 (treze) estruturas da cruzaína depositadas no *PDB*, todas complexadas com um inibidor covalente diferente. A **Tabela 6** apresenta os códigos *PDB* dos treze complexos, com as correspondentes resoluções e estruturas dos inibidores.

Tabela 6. Códigos *PDB,* resolução e representação da fórmula estrutural dos inibidores presentes nos treze complexos da cruzaína, respectivamente. As estruturas dos inibidores foram modificadas para representar os compostos antes da formação do complexo covalente. Os círculos vermelhos indicam o átomo de carbono que se liga covalentemente à enzima.

Código *PDB* (Resolução)	Inibidor	Código *PDB* (Resolução)	Inibidor
1AIM[a] (2,0Å)		1F2B[c] (1,8)	
1EWL[b] (2,0)		1F2C[c] (2,0)	
1EWM[b] (2,0)		1ME3[d] (1,2)	
1EWO[b] (2,1)		1ME4[d] (1,2)	
1EWP[b] (1,7)		1U9Q[e] (2,3)	
1F29[c] (2,1)		2AIM[a] (2,2)	
1F2A[c] (1,6)			

a. [Gillmor *et al.*, 1997]; b Não publicado; c [Brinen *et al.*, 2000]; d [Huang *et al.*, 2003]; e [Choe *et al.*, 2005]

Conforme descrito no item **2.5.1.** desta tese, as estruturas dos treze complexos da cruzaína foram comparadas, entre si, após o alinhamento, átomo-a-átomo,[Guex & Peitsch,

1997] de todos os átomos da enzima, em relação à estrutura do complexo 1AIM. Foram excluídos do alinhamento os inibidores e moléculas de água presentes nos arquivos *PDB*. A **Figura 28** mostra o resultado do alinhamento das 13 estruturas da cruzaína depositadas no *PDB*. Todos as estruturas apresentam 100% de homologia, em relação à seqüência de resíduos de aminoácidos do complexo 1AIM. Os valores de *RMSD,* calculados para todos os átomos, variaram entre 0,21 Å e 0,41 Å (**Tabela 7**). Estas diferenças não são significativas, uma vez que as coordenadas atômicas têm um erro de aproximadamente 1/6 da resolução do cristal,[Bohm & Klebe, 1996] que no caso dos complexos de cruzaína utilizados tem resoluções variando entre 1,2 Å e 2,3 Å (**Tabela 6**), resultando em valores de incerteza em relação às coordenadas atômicas variando de 0,2 Å a 0,4 Å.

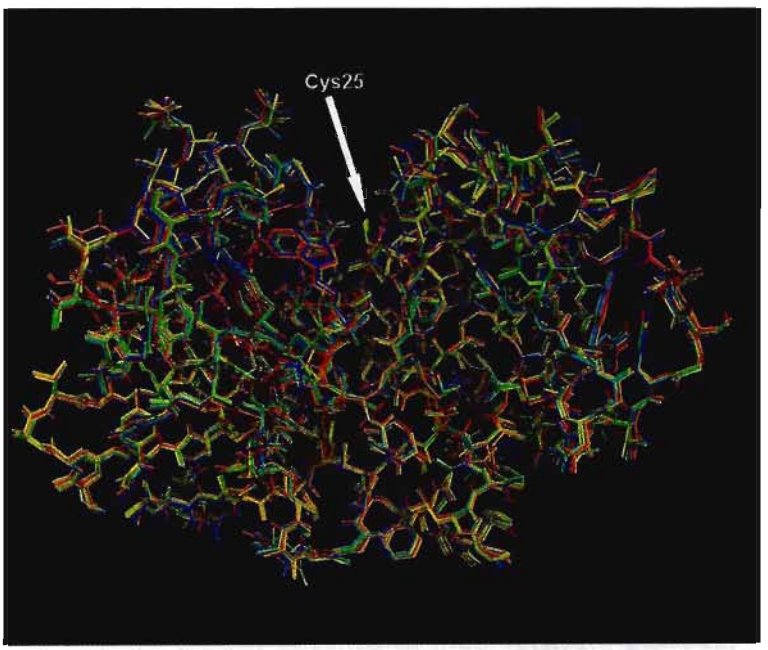


Figura 28. Estruturas dos 13 complexos da cruzaína alinhados átomo-a-átomo, em relação a estrutura do complexo 1AIM. Foram excluídos do alinhamento os inibidores e moléculas de água presentes nos arquivos *PDB*. Está indicado o resíduo catalítico *Cys25* no fundo da cavidade do sítio ativo da cruzaína.

A análise destes resultados indica a grande equivalência na estrutura terciária entre os complexos analisados, permitindo que uma única estrutura seja usada, como representante dos demais complexos da cruzaína, sem perda significativa das informações estruturais contidas nos outros complexos. A estrutura eleita como representante da cruzaína foi o complexo 1ME4, pois esta apresenta a melhor resolução cristalográfica

(1,2 Å). A partir deste ponto nesta tese, "estrutura da cruzaína" irá sempre se referir à estrutura contida no complexo 1ME4, a menos que indicado diferentemente.

Tabela 7. Valores de *RMSD* das 13 estruturas da cruzaína alinhadas átomo-a-átomo, em relação ao complexo 1AIM[a]

Código *PDB*	*RMSD*[b] (Å)
1AIM	-
2AIM	0,30
1EWL	0,21
1EWM	0,24
1EWO	0,31
1EWP	0,22
1F29	0,29
1F2A	0,28
1F2B	0,28
1F2C	0,38
1ME3	0,41
1ME4	0,39
1U9Q	0,30

a. Procedimentos realizados no programa *DeepView*.[Guex & Peitsch, 1997]
b. Valores de RMSD calculados conforme descrito no item **2.5.1.**

O alinhamento dos inibidores no sítio ativo da cruzaína, como resultado do alinhamento, átomo-a-átomo, das enzimas em seus complexos correspondentes, revelou a presença de choques entre os átomos dos inibidores e das enzimas com as quais estes não foram originalmente complexados. A presença deste choques é uma indicação da possibilidade da existência de ajuste induzido[Sotriffer *et al.*, 2004] da enzima devido à presença do inibidor. Este efeito será analisado mais a diante nesta tese, por dinâmica molecular.(Item **2.2.5.**)

3.2.1. Reconhecimento por inspeção visual

O reconhecimento da estrutura da cruzaína foi feito, inicialmente, através da inspeção visual do sítio ativo do complexo 1ME4, observando a natureza e a orientação dos resíduos de aminoácidos que o compõem. Conforme descrito na literatura,[Otto & Schirmeister, 1997] o centro catalítico da cruzaína é formado pela tríade *Cys25*, *His159* e *Asn175*, situada entre as cavidades ***S1*** e ***S1'***. A cavidade ***S1*** é aberta ao solvente e não tem resíduos que lhe atribuam especificidade. A cavidade ***S1'*** é ainda mais ampla e, igualmente, exposta ao solvente. A base da cavidade ***S1'*** é formada pelo anel indol do *Trp177*. A cavidade ***S2*** é bem delimitada pelas cadeias laterais dos resíduos hidrofóbicos *Met68*,

BIBLIOTECA
INSTITUTO DE QUÍMICA
Universidade de São Paulo

Ala133, *Leu157* e *Glu160*, respectivamente. No fundo da cavidade ***S2***, a *Glu205* pode expor, ora seus dois grupos metilenos intermediários, ora seu grupo carboxilato terminal. A flexibilidade da cadeia lateral da *Glu205* atribui uma dupla especificidade para a cavidade ***S2***, que pode assim acomodar resíduos tanto hidrofóbicos como positivos. A cavidade ***S3*** é parcialmente exposta ao solvente e é formada pelos resíduos *Ser61*, *Gly66*, *Leu67* e *Asp160*. A **Figura 29**-A apresenta uma representação gráfica da superfície de Van der Waals da cruzaína, onde estão indicadas as cavidades ***S3***, ***S2***, ***S1***, ***S1'***, enquanto que a **Figura 29**-B apresenta, em detalhe, uma representação gráfica dos resíduos que compõem as cavidades ***S3***, ***S2***, ***S1***, ***S1'*** da cruzaína.

A estrutura e o modo de ligação dos treze inibidores da cruzaína foi analisado a partir da estrutura dos ligantes alinhados no sítio ativo. Todos os inibidores da cruzaína contêm uma porção peptídica, ou peptideomimética, responsável pelo reconhecimento do sítio de ligação da enzima, e uma porção não-peptídica, responsável pela ligação covalente com a enzima. A **Figura 30**-A apresenta uma representação gráfica dos inibidores posicionados sobre a superfície de Van der Waals da cruzaína, alinhados em relação ao sítio ativo da 1ME4. A **Figura 30**-B mostra, em detalhe, uma representação gráfica dos inibidores alinhados e de alguns resíduos do sítio ativo da cruzaína.

Todos os treze inibidores da cruzaína analisados apresentam uma ligação aceptora de hidrogênio com o -NH da *Gly66* e uma ligação de hidrogênio doadora com o oxigênio carbonílico da Asp158. Adicionalmente, todos - exceto os complexos 1EWL e 1EWM - apresentam uma ligação aceptora de hidrogênio com o oxigênio carbonílico da Gly66. Todos os inibidores ocupam, ao menos parcialmente, as cavidades ***S2*** e ***S3***, com grupamentos hidrofóbicos ou positivamente carregados. A cavidade ***S1*** é ocupada por grupamentos hidrofóbicos em todos os complexos, exceto para 1EWL e 1EWM, que não ocupam a cavidade. A cavidade ***S1'***, é ocupada em apenas em 4 complexos (1F29; 1F2A; 1F2B e 1F2C), sendo todos com grupamento fenila. Todos os ligantes encontram-se covalentemente ligados ao enxofre da *Cys25*. As porções não-peptídicas que reagem com a *Cys25* são fluorometilcetonas (1AIM; 2AIM e 1EWP), epóxidos (1EWL e 1EWM); vinilsulfonas (1EWO; 1F29; 1F2A; 1F2B e 1F2C); α-hidroximetilcetona (1ME3 e 1ME4) e α-cetoéster (1U9Q). As estruturas dos ligantes podem ser observadas na **Tabela 6**.

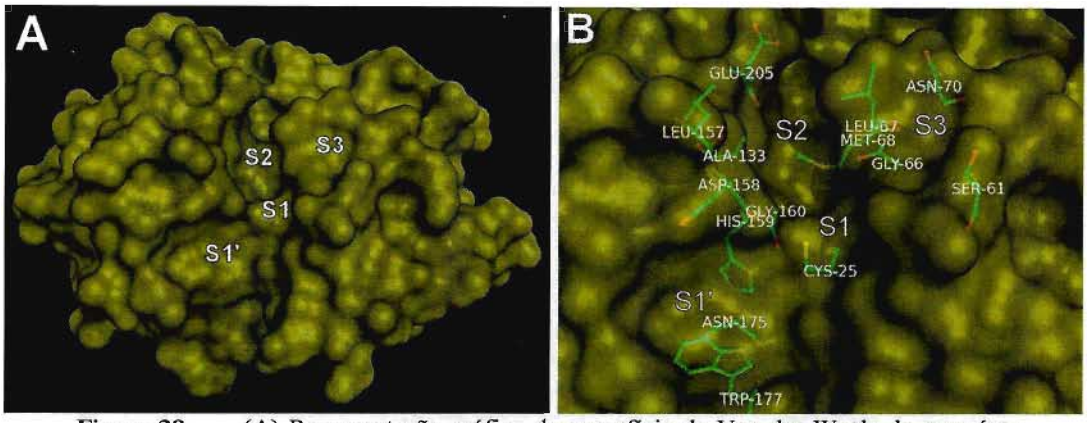


Figura 29. (A) Representação gráfica da superfície de Van der Waals da cruzaína.; (B) Representação gráfica dos resíduos que compõem o sítio ativo da cruzaína. As regiões assinaladas *S3*, *S2*, *S1*, *S1'* correspondem às respectivas cavidades do sítio ativo.

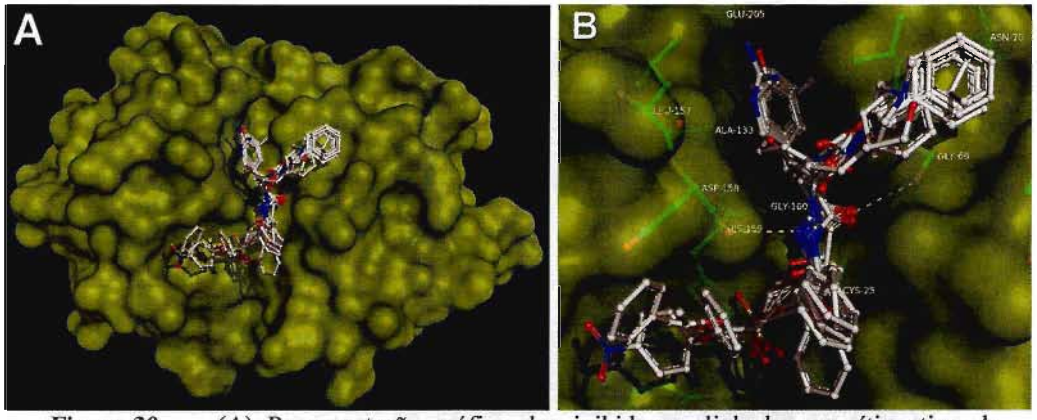


Figura 30. (A) Representação gráfica dos inibidores alinhados no sítio ativo da cruzaína; (B) Detalhe dos inibidores alinhados sobre a superfície da cruzaína, onde estão indicadas as ligações de hidrogênio entre os inibidores e os resíduos de *Gly66* e de *Asp158* no sítio ativo da enzima.

3.2.2. Reconhecimento por campos de interação molecular

Conforme apresentado nos itens **1.4.2.** e **1.4.3.**, o reconhecimento das propriedades do sitio de ligação de um alvo biológico pode ser feito a partir dos campos de interação molecular (*molecular interaction fields – MIF*),[Goodford, 1985] determinados ao redor da estrutura do alvo biológico.

Conforme descrito no item **2.5.2.** (*Materiais e Métodos*) desta tese, o reconhecimento das propriedades físico-químicas do sítio ativo da cruzaína foi realizado com base nos campos de interação molecular gerados pelo programa *GRID*.[Kastenholz *et al.*, 2000] Inicialmente, a sonda molecular "H_2O" foi utilizada para prever as posições mais favoráveis de uma molécula de água, na estrutura da cruzaína. Estas posições foram

comparadas com as posições de moléculas de água cristalográficas presentes dentro da cavidade do sítio ativo, em cada uma das 13 estruturas da cruzaína depositadas no *PDB*. A análise dos resultados mostrou que não houve sobreposição entre as posições das moléculas de água previstas pelo programa *GRID* e as moléculas de molécula de água cristalográficas. Igualmente, considerando-se apenas as moléculas de água cristalográficas dentro da cavidade do sítio ativo, nenhuma posição foi conservada em mais que 50% das estruturas, quando analisada por inspeção visual. Estes resultados sugerem que não existem moléculas de água conservadas na estrutura da cruzaína e que para os próximos estudos apresentados nesta tese envolvendo as estruturas da cruzaína depositadas no *PDB*, as moléculas de água cristalográficas podem ser desconsideradas.

Em seguida, outras três sondas moleculares disponíveis no programa *GRID*, "*Carbono aromático*", "*Oxigênio carbonílico*" e "*Nitrogênio de amida*", foram usadas para mapear dentro do sítio ativo da cruzaína, as regiões favoráveis de interação hidrofóbica, doadora de ligação de hidrogênio e aceptora de ligação de hidrogênio. As regiões de mínimo de energia dos campos de interação molecular obtidos para cada sonda representam "*hot spots*"[Malvezzi *et al.*, 2008], ou em outras palavras, regiões favoráveis para posicionar um átomo (ou grupamento) com propriedades equivalentes às de cada sonda. Estas regiões de mínimo de energia foram comparadas com as posições dos átomos dos inibidores presentes, respectivamente, nos 13 complexos da cruzaína depositados no *PDB*. A análise dos resultados mostra que os átomos dos inibidores se sobrepõem às regiões de mínimo de energia dos campos de interação molecular, obtidos para as três sondas moleculares. A correspondência observada entre a posição dos inibidores e os "*hot spots*" identificados pelo programa *GRID* indica que os inibidores são capazes de reconhecer os principais pontos de interação molecular da cruzaína e, portanto, são bons modelos para o desenvolvimento de um modelo farmacofórico de ligantes da enzima (*ligand based drug design*). A **Figura 31** apresenta uma representação gráfica dos contornos isopotenciais da cruzaína calculados pelo programa *GRID* [Goodford, 1985] usando quatro diferentes sondas moleculares.

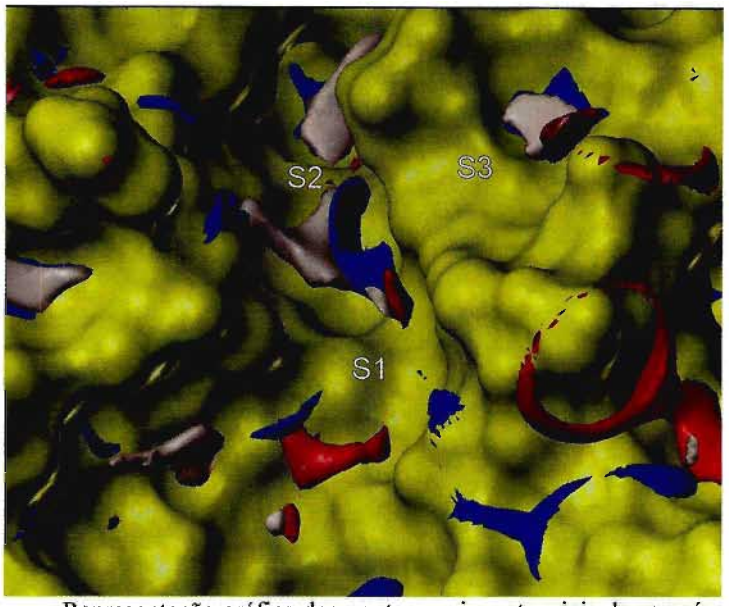


Figura 31. Representação gráfica dos contornos isopotenciais da cruzaína calculados pelo programa *GRID* [Goodford, 1985], usando-se diferentes sondas moleculares, sendo:. Branco - carbono alifático. nível de energia de –3,0 kcal/mol); Vermelho: oxigênio carbonílico - nível de energia de -3,0 kcal/mol; Azul: nitrogênio de amida - nível de energia de –6,0 kcal/mol). A superfície em amarelo corresponde a superfície de Van der Waals da cruzaína (1ME4). As regiões assinaladas *S1*, *S2* e *S3* correspondem às respectivas cavidades do sítio ativo.

3.2.3. Reconhecimento por pseudocentros

Conforme apresentado no **1.4.3.**, as propriedades de reconhecimento molecular da cavidade do sítio de ligação podem ser analisadas por pseudocentros.[Schmitt *et al.*, 2001] Estes são pontos na superfície da cavidade, que representam as propriedades físico-químicas dos resíduos que compõem esta cavidade e que são expostas na superfície.

Conforme descrito no item **2.5.3.** (*Materiais e Métodos*) desta tese, a análise dos pseudocentros da cavidade do sítio ativo da cruzaína foi realizada utilizando-se o programa *CAVBASE*,[Schmitt *et al.*, 2001] que é um módulo do programa *RELIBASE+*.[Gohlke *et al.*, 2000a] Através do programa *CAVBASE* as propriedades de reconhecimento molecular são transformadas em cinco tipos de pseudocentros, respectivamente: (*i*) doador de ligação de hidrogênio; (*ii*) aceptor de ligação de hidrogênio; (*iii*) doador-aceptor de ligação de hidrogênio; (*iv*) hidrofóbico alifático e (*v*) hidrofóbico aromático. Para a cavidade da cruzaína o programa identificou 63 pseudocentros, representando as propriedades dos resíduos expostas na superfície desta cavidade. A **Figura 32** apresenta todos os pseudocentros identificados na cavidade da cruzaína.

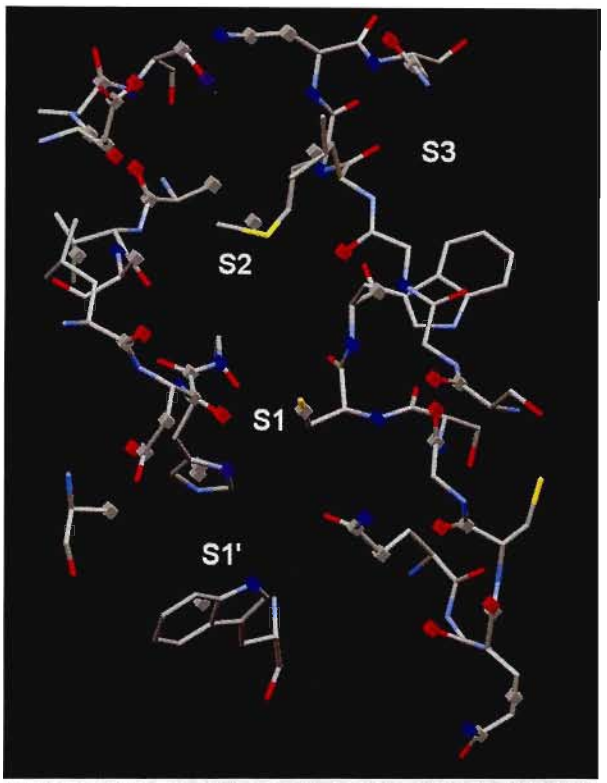


Figura 32. Representação dos 63 pseudocentros (PC) identificados pelo programa *CAVBASE*, alinhados sobre os resíduos que compõem a cavidade de ligação da cruzaína (estrutura extraída do complexo *PDB 1ME4*). Os cubos em azul, em vermelho e, em cinza representam os PC de resíduos doadores de ligação de hidrogênio, aceptores de ligação de hidrogênio e hidrofóbicos alifáticos, respectivamente. As regiões assinaladas *S1*, *S2*, *S3* e *S1'* correspondem as respectivas cavidades do sítio ativo.

Diferentemente da análise dos campos de interação molecular realizada anteriormente, os pseudocentros identificados pelo programa *CAVBASE* representam pontos de interação na superfície da proteína, enquanto que o programa *GRID* identifica regiões complementares aos pontos de interação da proteína. Sendo assim, a comparação dos resultados das duas análises não pode ser feita diretamente. Entretanto, uma análise comparativa, feita por inspeção visual no programa *DEEPVIEW*, entre os "*hot spots*" identificados pelo programa *GRID* e os pseudocentros identificados pelo programa *CAVBASE*, permite a identificação de quais pseudocentros correspondem às propriedades de reconhecimento molecular mais importantes na cavidade da cruzaína. Assim, foram escolhidos 10 pseudocentros para caracterizar a cavidade da cruzaína considerando tanto os "*hot spots*", identificados pelo programa *GRID*, quanto a posição dos treze inibidores da cruzaína cristalizados nos complexos *PDB*. A **Figura 33** apresenta os 10 pseudocentros selecionados, alinhados sobre os resíduos que compõem a cavidade de ligação da enzima.

Estes pseudocentros (PC) escolhidos correspondem às seguintes propriedades: doadora de ligação de hidrogênio do "*oxianion hole*" do centro catalítico (PC 1 e 2 na **Figura 33**); aceptora de ligação de hidrogênio da *Asp158* e doadora e aceptora de ligação de hidrogênio da *Gly66* (PC 3, 4 e 5, respectivamente, na **Figura 33**) e hidrofóbicas dos resíduos que formam as cavidades *S2* (PC 6, 7, 8 e 9 na **Figura 33**) e *S3* (PC 9 e 10 na **Figura 33**), respectivamente.

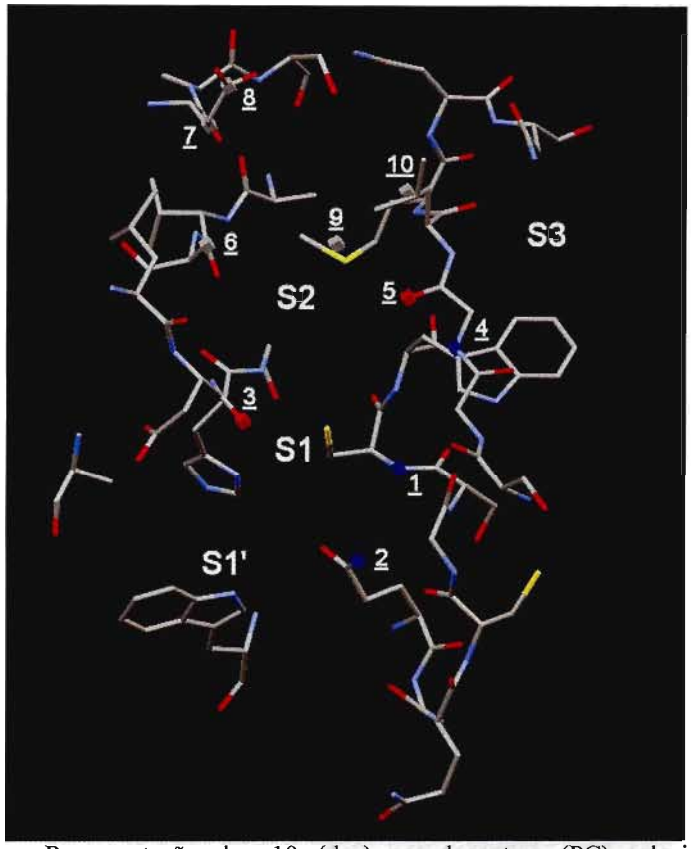


Figura 33. Representação dos 10 (dez) pseudocentros (PC) selecionados pelo programa *CAVBASE*, alinhados sobre os resíduos que compõem a cavidade de ligação da cruzaína (estrutura extraída do complexo *PDB 1ME4*). Os cubos em azul, em vermelho e em cinza representam os PC de resíduos doadores de ligação de hidrogênio, aceptores de ligação de hidrogênio e hidrofóbicos alifáticos, respectivamente.

3.2.4. Busca por cavidades similares

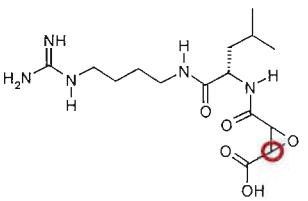
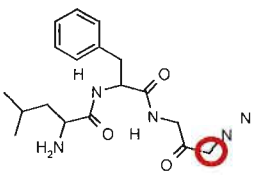
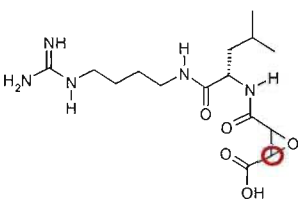
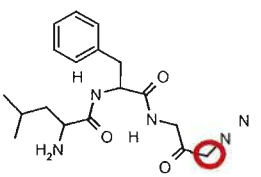
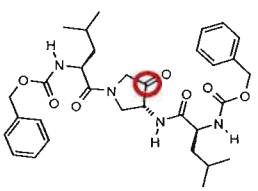
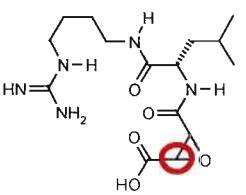
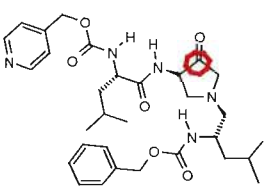
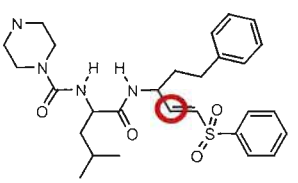
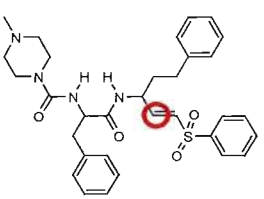
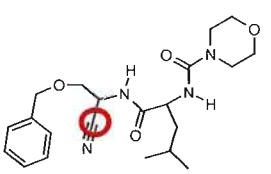
Conforme descrito no item **2.5.3.** (*Materiais e Métodos*), o modelo de cavidade da cruzaína com 10 pseudocentros foi utilizado para uma busca no banco de dados *PDB* por proteínas apresentando cavidades similares (ou seja, com a mesma natureza e distribuição espacial de pseudocentros) e que tenham um ligante posicionado dentro desta cavidade. O

objetivo desta busca é encontrar ligantes em cavidades similares à cruzaína para que estes possam ser usados na proposição e validação de modelos virtuais de atividade da enzima.

A busca no banco de dados *PDB* por cavidades similares à cruzaína foi realizada pelo programa *CAVBASE*, que listou as 100 melhores soluções, classificadas em função do número de pseudocentros correspondentes e da homologia de seqüência. Foram excluídos os resultados em que a proteína apresentava 100% de homologia com a *1ME4* e os resultados em que o ligante posicionado no sítio ativo era um metal, ou molécula inorgânica. A análise dos resultados selecionou 20 estruturas, com pelo menos 6 pseudocentros correspondentes aos 10 pseudocentros do modelo da cruzaína. Todos os complexos selecionados eram estruturas de cisteíno-proteases covalentemente ligadas a um inibidor peptídico. As homologias de seqüência entre as 20 proteínas selecionadas e a cruzaína variaram de 25,5% à 43,4%. A **Tabela 8** apresenta os códigos *PDB* dos 20 complexos selecionados, com as correspondentes resoluções e estrutura dos inibidores.

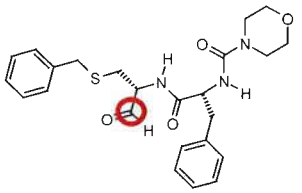
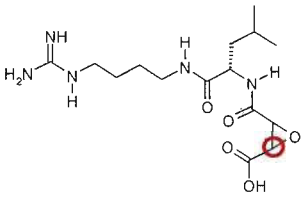
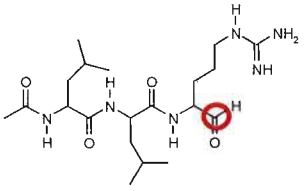
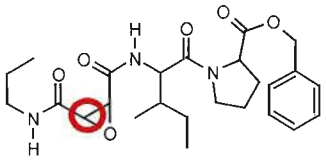
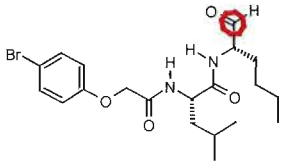
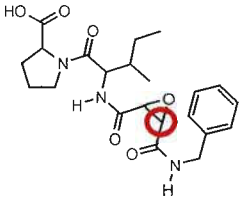
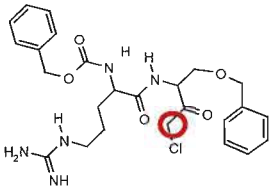
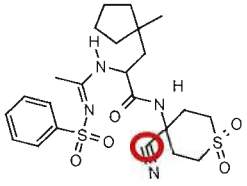
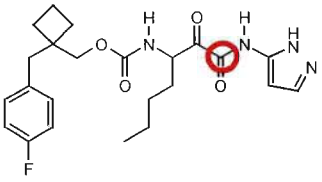
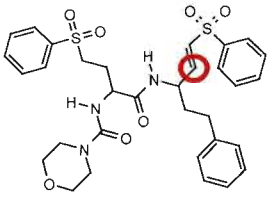
Conforme descrito no item **2.5.1.** (*Materiais e Métodos*), as 20 estruturas dos complexos selecionados na busca com o *CAVBASE* foram alinhadas, em relação ao sítio ativo da cruzaína, tomando-se como referência os carbonos-alfa dos resíduos *Cys25, Gly66, Asp158* e *His159,* respectivamente. A **Figura 34** mostra o resultado do alinhamento das 20 estruturas dos complexos de cisteíno-protease selecionados do banco de dados *PDB*.

Tabela 8. Códigos *PDB*, resolução e representação da fórmula estrutural dos inibidores presentes, respectivamente, nos 20 complexos selecionados pelo programa *CAVBASE*. As estruturas dos inibidores foram modificadas para representar os compostos antes da formação do complexo covalente. Os círculos em vermelho indicam o átomo de carbono que se liga covalentemente à enzima.

Código *PDB* (Resolução)	Inibidor	Código *PDB* (Resolução)	Inibidor
1AEC (1,86Å)		1KHP (2,00Å)	
1ATK (2,20Å)		1KHQ (1,60Å)	
1AU3 (2,50Å)		1MEG (2,00Å)	
1AU4 (2,30Å)		1MEM (1,80Å)	
1FH0 (1,60Å)		1MS6 (1,90Å)	

Continua

Tabela 8.(continuação) Códigos *PDB,* resolução, número de pseudocentros equivalentes à cruzaína e representação da fórmula estrutural dos inibidores presentes, respectivamente, nos 20 complexos selecionados pelo programa *CAVBASE.* As estruturas dos inibidores foram modificadas para representar os compostos antes da formação do complexo covalente. Os círculos em vermelho indicam o átomo de carbono que se liga covalentemente à enzima.

Código *PDB* (Resolução)	Inibidor	Código *PDB* (Resolução)	Inibidor
1NQC (1,80Å)		2BDZ (2,10Å)	
1POP (2,10Å)		2DC6 (2,30Å)	
1SNK (2,40Å)		2DCD (2,50Å)	
1THE (1,90Å)		2FRQ (1,60Å)	
1TU6 (1,75Å)		2G6D (2,50Å)	

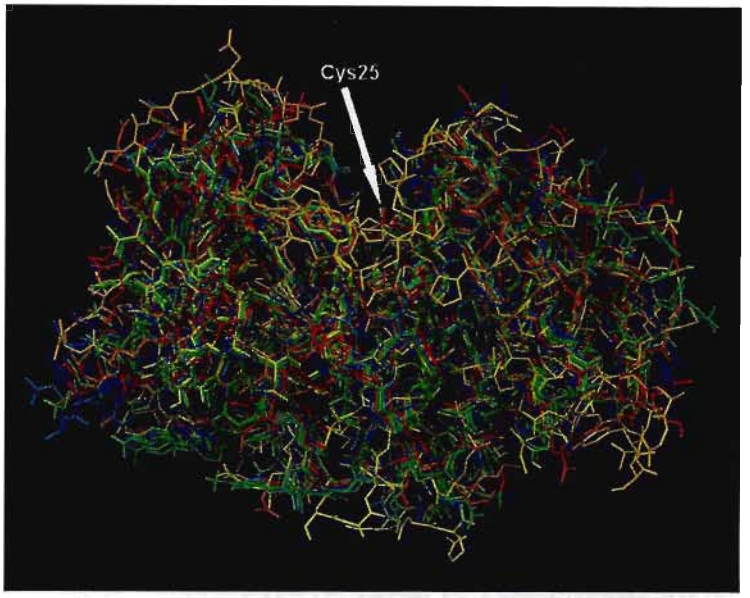


Figura 34. Estruturas das 20 cisteíno-protease com cavidades semelhantes à cruzaína, selecionados do banco de dados *PDB* utilizando-se o programa *CAVBASE*, alinhadas átomo-a-átomo, em relação a estrutura do complexo 1AIM. Foram excluídos do alinhamento os inibidores e moléculas de água presentes nos arquivos *PDB*. Está indicado o resíduo catalítico *Cys25* no fundo da cavidade do sítio ativo catalítico.

A **Figura 35** mostra a sobreposição dos inibidores extraídos dos 20 complexos de cisteíno-protease com cavidades semelhantes à cruzaína, selecionados do banco de dados *PDB* pelo programa *CAVBASE*, alinhados sobre a cavidade do sítio ativo da cruzaína. Todos, exceto os inibidores dos complexos 1AU3 e 1AU4, apresentam uma ligação aceptora de hidrogênio com o -NH da *Gly66* e uma ligação de hidrogênio doadora com o oxigênio carbonílico da *Asp158*. Adicionalmente, todos, exceto os inibidores dos complexos 1TU6, 2DC6 e 2DCD, apresentam uma ligação aceptora de hidrogênio com o oxigênio carbonílico da *Gly66*. Todos os inibidores ocupam, ao menos parcialmente, as cavidades ***S2*** e ***S3***, com grupamentos hidrofóbicos ou positivamente carregados (a única exceção é o inibidor do complexo 2DCD que não ocupa a cavidade ***S3***). A cavidade ***S1*** é ocupada por menos da metade dos inibidores (1FHO, 1MEM, 1NQC, 1POP, 1SNK, 1THE, 1TU6, 2FRQ e 2G6D), enquanto que os demais deixam a cavidade livre. A cavidade ***S1'*** é ocupada em apenas em 8 complexos (1AU3, 1AU4, 1FH0, 1MEM, 1TU6, 2DC6, 2DCD e 2G6D). As porções não-peptídicas que reagem com a *Cys25* são: epóxidos (1AEC; 1ATK; 1MEG; 2BDZ; 2DCD e 2DC6); vinilsulfonas (1FHO; 1MEM e 2G6D); α-azometilcetona

(1KHP e 1KHQ); clorometilcetona (1THE); ciano (1MS6 e 2FRQ); aldeídos (1NQC; 1POP e 1SHK); cetonas cíclicas (1AU3 e 1AU4) e α-cetamino (1TU6).

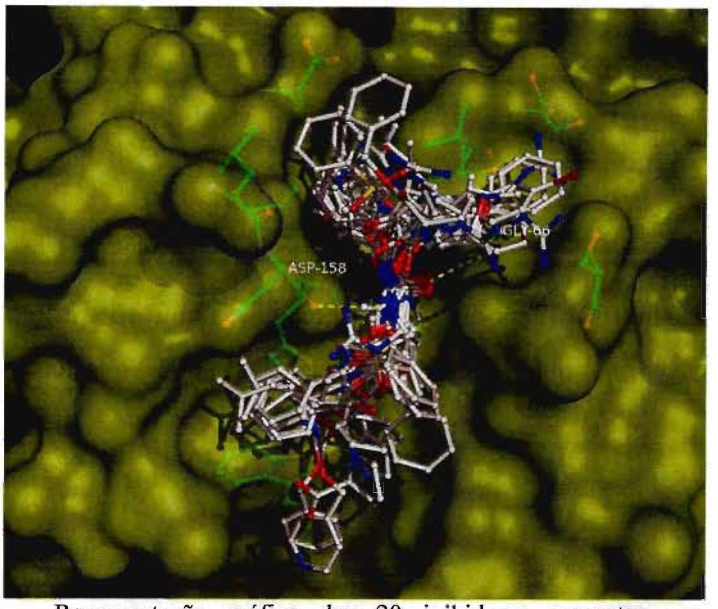


Figura 35. Representação gráfica dos 20 inibidores presentes nos complexos selecionados pela busca de cavidades similares à cruzaína, utilizando o programa *CAVBASE*,[Schmitt *et al.*, 2001] alinhados no sítio ativo da cruzaína. Estão indicadas as ligações de hidrogênio entre os inibidores e os resíduos de *Gly66* e de *Asp158* no sítio ativo da enzima.

3.2.5. Reconhecimento por dinâmica molecular

Conforme discutido na *Introdução* desta tese, item **1.4.3.**, uma das estratégias para se considerar tanto a flexibilidade da proteína, quanto a possibilidade da ocorrência de ajuste induzido na conformação da proteína devido à presença de um ligante é a aplicação de protocolos de dinâmica molecular. Sendo assim, conforme descrito no item **2.5.4.** (*Materiais e Métodos*), foram realizados estudos de dinâmica molecular da enzima livre e da enzima complexada com inibidor.

Antes de serem apresentados e analisados os resultados da dinâmica molecular deve-se destacar que no modelo em que a enzima se encontra complexada com o inibidor, a ligação covalente entre o inibidor e a enzima não pode ser reproduzida, nas condições de dinâmica molecular empregadas. Neste caso, foi considerado um complexo não-covalente, em que a *Cys25* encontra-se na forma de tiol, e o ligante em sua estrutura antes de sofrer o ataque nucleofílico. Apesar desta não ser a mesma estrutura que foi cristalizada, assume-se

que o complexo não-covalente como representado neste caso deve existir, por um certo período de tempo, antes do ataque da *Cys25* e a conseqüente formação do complexo covalente e irreversível.

Partindo-se da estrutura da cruzaína, extraída do complexo 1ME4, foram registrados 3ns das trajetórias da enzima, respectivamente, livre e complexada com seu próprio inibidor. Os resultados foram analisados quanto à variação da posição dos átomos em relação à estrutura cristalina, durante toda a trajetória. A **Figura 36** apresenta os gráficos dos valores de *RMSD* em função do tempo, obtidos a partir das trajetórias da cruzaína livre (vermelho) e complexada com o ligante (azul), para todos os átomos da proteína (**Figura 36**-A); para os átomos do sítio ativo (**Figura 36**-B); para os átomos de carbonos-alfa (**Figura 36**-C) e para os átomos de cadeias laterais (**Figura 36**-D). Estes resultados estão resumidos na **Tabela 9** que apresenta valores de *RMSD* médios e os respectivos desvios padrão de cada uma das trajetórias analisadas.

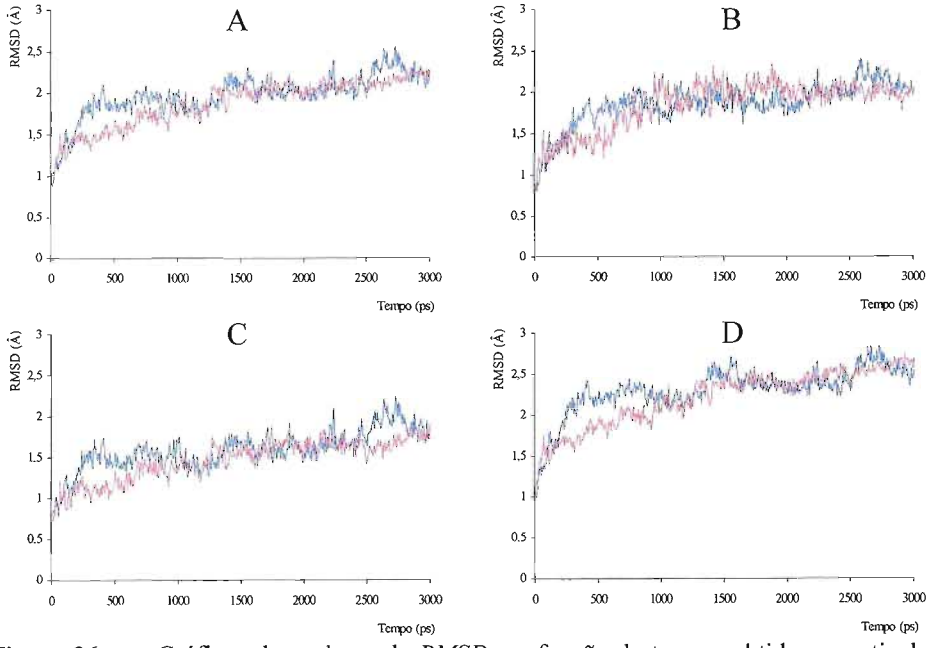


Figura 36. Gráficos dos valores de *RMSD* em função do tempo, obtidos a partir das trajetórias[a] da cruzaína[b] sem o ligante (vermelho) e complexada com o ligante 1ME4 (azul), para: todos os átomos da proteína (A); os átomos do sítio ativo (B); os átomos de carbono-alfa (C) e os átomos de cadeias laterais (D), respectivamente.

[a.] Obtidas por dinâmica molecular no período de 3 ns, à temperatura de 300 K, utilizando-se o programa *Gromacs3.2.1* [Lindahl *et al.*, 2001; Berendsen *et al.*, 1995], e o campo de força G43a1.

[b.] As coordenadas iniciais foram obtidas a partir da estrutura *PDB* do complexo 1ME4.

Tabela 9. Valores médios de *RMSD* (± desvios padrão) observados nas trajetórias[a] da cruzaína[b] sem o ligante e complexada com o ligante 1ME4, para todos os átomos, respectivamente, da proteína; do sítio ativo; de carbono-alfa e de cadeias laterais.

Grupo de átomos	# átomos	*RMSD* (Å)	
		cruzaína livre	cruzaína complexada
Toda a Proteína	2017	1,88 ± 0,29	1,97 ± 0,27
Sitio Ativo	349	1,83 ± 0,30	1,86 ± 0,26
Carbono-Alfa	215	1,47 ± 0,25	1,59 ± 0,25
Cadeia Lateral	947	2,21 ± 0,35	2,30 ± 0,30

[a] Obtidas por dinâmica molecular no período de 3ns, a uma temperatura de 300K, utilizando-se o programa *Gromacs3.2.1* [Lindahl *et al.*, 2001; Berendsen *et al.*, 1995], e o campo de força G43a1.

[b] As coordenadas iniciais obtidas a partir da estrutura *PDB* do complexo 1ME4.

Observando os valores na **Tabela 9,** não se nota diferença significativa entre as trajetórias da enzima livre e da enzima complexada. Os valores baixos de RMSD (menores que 2 Å) observados quando se consideram todos os átomos da proteína, ou apenas os átomos de carbonos-alfa, sugerem que a cruzaína apresenta uma significativa estabilidade de sua estrutura terciária e um enovelamento globular compacto.[Sotriffer *et al.*, 2004] Considerando apenas os átomos do sítio ativo, as trajetórias da enzima livre, ou complexada, também são semelhantes, dentro do desvio padrão, indicando que o sítio ativo não sofre grandes mudanças conformacionais, tanto quando este se encontra livre do ligante, ou quando este se encontra complexado, com um ligante. Finalmente, como esperado, os valores de *RMSD,* considerando apenas os átomos de cadeias laterais, são os maiores entre os grupos analisados, revelando que as cadeias laterais apresentam alguma flexibilidade, principalmente em resíduos na superfície da proteína e expostos ao solvente.

Para avaliar mais especificamente a possibilidade de ajuste induzido do sítio ativo devido à presença do ligante, foram registrados os valores das distâncias entre pontos selecionados do sítio ativo. A comparação entre estas distâncias na enzima livre, e na enzima complexada, permite identificar se existem resíduos que sofrem mudanças conformacionais (ajuste induzido) devido à presença do ligante. A **Figura 37**, apresenta os gráficos dos valores das distâncias em função do tempo entre os seguintes átomos: *CA-Asp60 – CA-Gly66* (**Figura 37-A**); *CA-Cys25 - CA-His159* (**Figura 37-B**); *CA-Met68 – CA-Val132* (**Figura 37-C**); *CA-Ser139 – CA-Asp158* (**Figura 37-D**); CD *Glu205 – CA-Cys25* (**Figura 37-E**); *CG-Asp158 – CA-Gly65* (**Figura 37-F**); *CG-Leu67 – CA-Gly65* (**Figura 37-G**); e entre os resíduos *Asp60 – Gly66* (**Figura 37-H**); *Cys25 – His159* (**Figura 37-I**); *Met68 - Val132* (**Figura 37-J**) e *Ser139 - Asp158* (**Figura 37-K**). Para o cálculo dos valores das distâncias entre dois resíduos foi considerado o centro geométrico de cada

resíduo. Na **Tabela 10** estão apresentados os valores das distâncias, respectivamente, média, máxima e mínima, observadas nas trajetórias da cruzaína livre e complexada, entre os mesmos conjuntos de átomos descritos anteriormente.

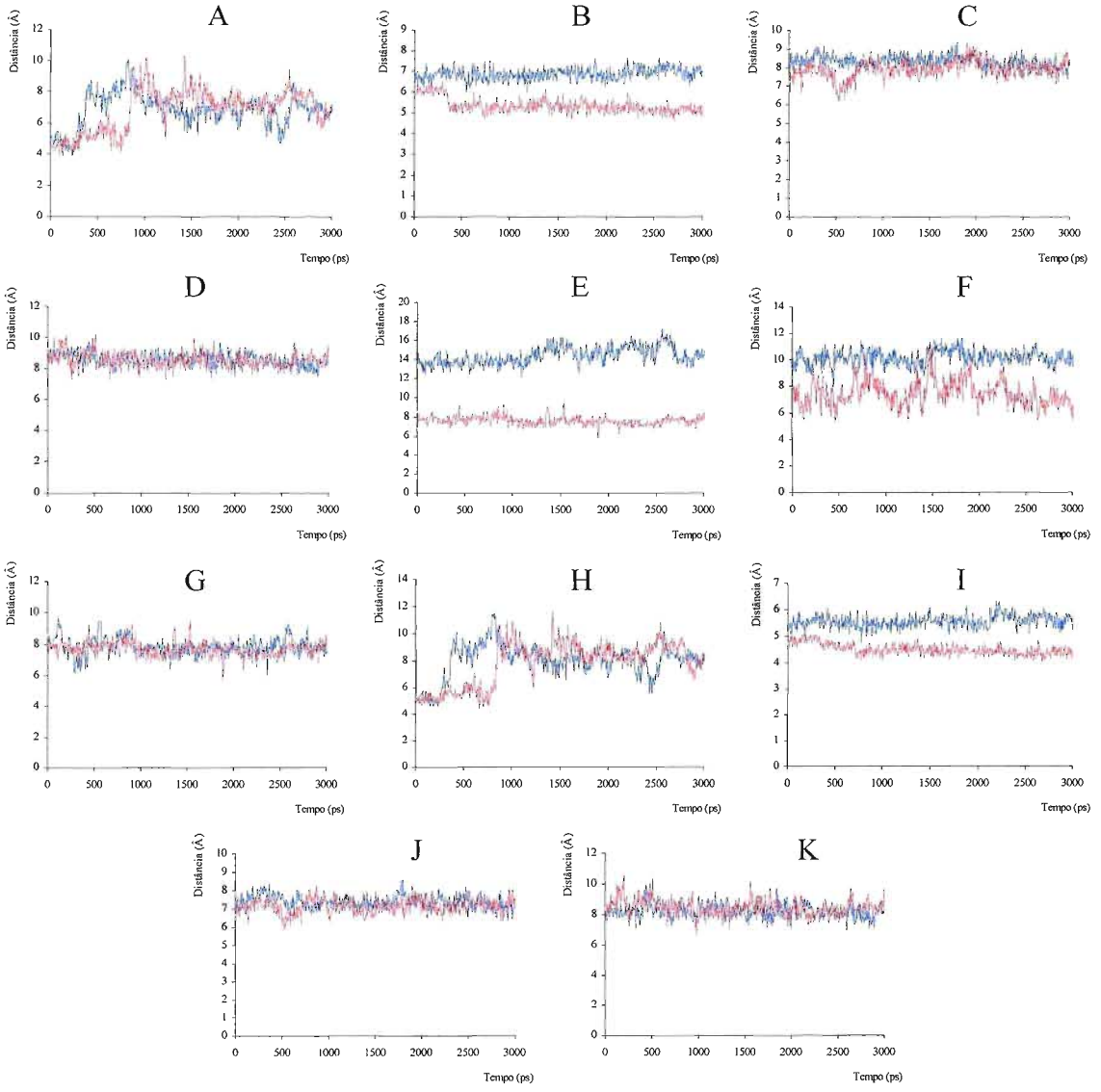


Figura 37. Gráficos dos valores das distâncias em função do tempo, obtidos a partir das trajetórias[a] da cruzaína[b] livre (vermelho) e complexada (azul), respectivamente, entre os seguintes átomos: *CA-Asp60 – CA-Gly66* (A); *CA-Cys25 - CA-His159* (B); *CA-Met68 – CA-Val132* (C); *CA-Ser139 – CA-Asp158* (D); CD *Glu205 – CA-Cys25* (E); *CG-Asp158 – CA*-Gly65 (F); *CG*-Leu67 – *CA*-Gly65 (G); e entre os resíduos: Asp60[c] – Gly66[c] (H); *Cys25*[c] – *His159*[c] (I); *Met68*[c] - *Val132*[c] (J) e, *Ser139*[c] - *Asp158*[c] (K), respectivamente.

[a] Obtidas por dinâmica molecular no período de 3 ns, a uma temperatura de 300 K, utilizando-se o programa *Gromacs3.2.1* [Lindahl *et al.*, 2001; Berendsen *et al.*, 1995], e o campo de força G43a1.

[b] As coordenadas iniciais foram obtidas a partir da estrutura *PDB* do complexo 1ME4.

[c] Para o cálculo dos valores das distâncias entre resíduos foi considerado o centro geométrico de cada resíduo.

Tabela 10. Valores das distâncias, respectivamente, média, máxima e mínima, observadas nas trajetórias[a] da cruzaína[b] livre e a correspondente complexada, entre átomos do sítio ativo.

Distâncias entre os átomos ou resíduos (Å)	cruzaína livre			cruzaína complexada		
	Média ± SD	Max.	Min.	Média ± SD	Max.	min
CA *Asp60 – CA-Gly66*	6,86 ± 1,34	10,28	3,88	6,84 ± 1,06	10,14	3,97
CA *Cys25 – CA-His159*	5,38 ± 0,35	6,51	4,66	6,90 ± 0,29	7,73	5,97
CA *Met68 – CA-Val132*	7,90 ± 0,48	9,37	6,16	8,30 ± 0,35	9,39	7,20
CA *Ser139 – CA-Asp158*	8,59 ± 0,46	10,29	7,33	8,60 ± 0,43	10,03	7,38
CD *Glu205 – CA-Cys25*	7,75 ± 0,44	9,58	5,76	14,41 ± 0,91	17,31	12,1
CG-Asp158 – CA-Gly65	7,48 ± 0,95	11,34	5,44	10,13 ± 0,55	11,67	8,02
CG-Leu67 – CA-Gly65	7,75 ± 0,44	9,58	5,76	7,90 ± 0,56	9,82	6,07
Asp60[c] – Gly66[c]	7,84 ± 1,62	11,65	4,43	8,01 ± 1,28	11,60	4,59
Cys25[c] – His159[c]	4,51 ± 0,21	5,07	4,04	5,57 ± 0,23	6,36	4,74
Met68[c] - Val132[c]	7,19 ± 0,38	8,33	5,95	7,41 ± 0,36	8,67	6,38
Ser139[c] - Asp158[c]	8,46 ± 0,54	10,64	6,69	8,23 ± 0,45	10,00	6,98

[a] Obtidas por dinâmica molecular no período de 3ns, a uma temperatura de 300K, utilizando-se o programa *Gromacs3.2.1* [Lindahl *et al.*, 2001; Berendsen *et al.*, 1995], e o campo de força G43a1.
[b] As coordenadas iniciais foram obtidas a partir da estrutura *PDB* do complexo 1ME4.
[c] Para o cálculo dos valores das distâncias entre resíduos foi considerado o centro geométrico de cada resíduo.

Observa-se, nos valores das distâncias entre átomos do sítio ativo da cruzaína, apresentados na **Figura 37** e na **Tabela 10**, que a presença do ligante diminuiu o grau de liberdade de alguns átomos do sítio ativo. A perda de flexibilidade pode ser observada através dos valores de desvio padrão das distâncias entre os pontos do sítio ativo selecionados, que são em geral menores durante a trajetória do complexo quando comparados com os observados na trajetória da enzima livre. Por exemplo, na **Figura 37**-C, observa-se que a distância entre o *CA-Met68* e *CA-Val132* apresenta valores de desvio padrão de 0,48 Å e de 0,35 Å, durante a trajetória da cruzaína livre e complexada, respectivamente. Este efeito é ainda mais significativo quando os átomos considerados se encontram nas cadeias laterais de resíduos do sítio ativo, como por exemplo na **Figura 37**-F, em que a distância entre o *CG-Asp158* e o *CA-Gly65* variou com um desvio padrão de 0,95 Å, durante a trajetória da cruzaína livre, e de 0,55 Å, durante a trajetória da enzima complexada. Outro aspecto a ser observado, é a significativa mudança de posição de cadeias laterais de resíduos do sítio ativo próximos ao ligante. Estas mudanças podem ser observadas na **Figura 37**-E, onde o valor da distância entre os átomos *CD-Glu205* e *CA-Cys25*, foi em média de 7,75 Å durante a trajetória da cruzaína livre e de 14,41 Å durante a trajetória da enzima complexada. Outro exemplo desta mudança de posição pode ser observado na **Figura 37**-F, onde a distância entre os átomos *CG-Asp158* e *CA-Gly65*, foi em média de 7,48 Å durante a trajetória da cruzaína livre e de 10,13 Å durante a trajetória da enzima complexada.

Os resultados da dinâmica molecular obtidos, analisando-se os valores das distâncias entre átomos do sítio ativo em conjunto com os resultados de *RMSD* descritos anteriormente, confirmam tanto a baixa flexibilidade da estrutura da cruzaína (considerando-se a proteína como um todo) como a ocorrência de ajuste induzido, em alguns pontos do sítio ativo após a complexação desta com um ligante, principalmente quando considerados os átomos de cadeias laterais de resíduos do sítio ativo. Estes resultados sugerem que ligantes, obtidos a partir de procedimentos de *docking* (que consideram a proteína rígida), devem ter suas posições dentro do sítio ativo confirmadas, quando possível, por testes de dinâmica molecular, para que a possibilidade do ajuste induzido dos resíduos do sítio ativo possa ser também levada em consideração.

3.3. Obtenção e validação de modelo de *virtual screening* de inibidores da cruzaína - Modelo I

Conforme discutido na *Introdução* desta tese, item **1.4.**, a estratégia de *virtual screening*, utilizada na descoberta de compostos bioativos, consiste na aplicação seqüencial de diferentes filtros virtuais em bibliotecas de compostos. Assim, partindo-se do banco de dados *ZINC*[Irwin & Shoichet, 2005] versão 5, foram aplicados filtros virtuais de diferentes naturezas para a seleção de inibidores da cruzaína. Cada um dos diferentes filtros virtuais, a saber, físico-químico, farmacofórico, de *docking* e de inspeção visual, foi individualmente validado utilizando-se as estruturas dos 13 inibidores da cruzaína, extraídos dos complexos depositados no banco de dados *PDB*. Os compostos selecionados pelo modelo foram testados como inibidores da cruzaína. A **Figura 38** mostra uma representação esquemática da seqüência de aplicação de cada filtro, acompanhada dos valores da porcentagem de redução do número de compostos da biblioteca, em cada etapa.

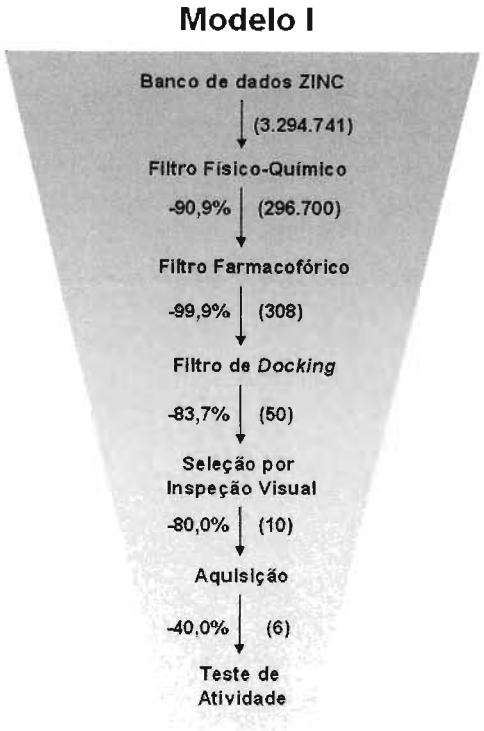


Figura 38. Representação esquemática da seqüência filtros aplicados sobre a biblioteca de compostos *ZINC*,[Irwin & Shoichet, 2005] representando o **Modelo I** de *virtual screening*. Em parêntesis são indicados os números de compostos selecionados em cada etapa e ao lado a correspondente porcentagem de redução, por cada filtro.

3.3.1. Filtro Físico-Químico

Uma discussão detalhada sobre filtros virtuais de compostos, baseados em propriedades físico-químicas, está apresentada no item **1.4.1.** da *Introdução* desta tese.

A primeira etapa do procedimento de *virtual screening* foi a aplicação de um filtro de propriedades físico-químicas sobre todos os compostos do banco de dados *ZINC*. Todos os procedimentos nesta etapa foram executados utilizando-se o programa DBfilter,[Wang, 2003] e estão descritos no item **2.5.6.** (*Materiais e Métodos*). As propriedades físico-químicas, utilizadas como filtros, foram baseadas na regra-dos-5 de Lipinski[Lipinski *et al.*, 1997] e os seus valores de máximo e de mínimo foram baseados nos treze compostos extraídos dos complexos da cruzaína. Entretanto, considerando-se que estes inibidores são derivados peptídicos, ou peptidomiméticos, e que um dos objetivos deste trabalho era encontrar inibidores não-peptídicos da cruzaína, os valores de máximo e de mínimo, das propriedades físico-químicas utilizadas como filtros foram alterados para que fossem selecionados compostos menores, menos lipofílicos e menos flexíveis do que os inibidores

presentes nos complexos da cruzaína, usados para validar o modelo de *virtual screening*. A **Tabela 11** apresenta as propriedades físico-químicas utilizadas e os seus correspondentes valores de máximo e de mínimo, utilizados para selecionar inibidores da cruzaína.

Tabela 11. Valores das propriedades físico-químicas utilizadas para selecionar inibidores da cruzaína

Propriedade Físico-Química	Valor (Min e Máx)	
Átomos permitidos	H;C;N;O;F;Cl;Br;I;P;S	
Massa Molecular	250	600
logP	-2	5
Grupos doadores de ligação de hidrogênio	2	5
Grupos aceptores de ligação de hidrogênio	2	10
Soma dos átomos de oxigênio e nitrogênio	3	8
Halogênios	0	2
Número de anéis	0	4
Tamanho do anel	0	8
Ligações com rotação livre	0	6
Extensão da cadeia alifática $CH_3(CH_2)_n$-	0	2

Em função da alteração dos valores de máximo e de mínimo das propriedades utilizadas como filtros, em especial o valor máximo de seis ligações com rotação livre, nenhum dos 13 inibidores da cruzaína, extraídos dos complexos *PDB*, foram reconhecidos pelo filtro físico-químico. A aplicação destes filtros físico-químicos sobre o banco de dados *ZINC* encontrou 296.700 moléculas apresentando as propriedades *drug-like* pré-selecionadas. A aplicação deste filtro representou uma redução de 90,9% do número de moléculas do banco de dados *ZINC* original (3.294.741 moléculas).

3.3.2. Filtro farmacofórico

A segunda etapa do procedimento de *virtual screening* foi a aplicação de um filtro farmacofórico sobre os compostos selecionados na etapa anterior. Uma discussão detalhada sobre filtros farmacofóricos aplicados para a seleção de compostos bioativos está apresentada no item **1.4.2.** da *Introdução* desta tese.

Todas as etapas envolvendo a preparação, a validação e a aplicação do modelo farmacofórico foram realizadas no programa *CATALYST*,[Catalyst, 2005] e estão descritas no item **2.5.5.** (*Materiais e Métodos*). Para a construção do modelo farmacofórico foram utilizadas apenas as estruturas dos inibidores presentes nos complexos, respectivamente, 1AIM; 1EWO; 1F29; 1F2A; 1F2B; 1F2C; 1ME3; 1ME4 e 1U9Q. Os inibidores presentes

nos complexos 2AIM; 1EWL e 1EWP não foram utilizados pois estes não apresentam um grupamento fenila na cavidade ***S2***. Conforme descrito na literatura,[Cazzulo *et al.*, 2001] a cavidade ***S2*** da cruzaína tem afinidade por grupamentos hidrofóbicos aromáticos, como o grupo fenila. A presença deste grupamento, na estrutura do ligante, aumenta a afinidade e favorece o correto posicionamento deste no sítio ativo da cruzaína. Adicionalmente, o inibidor presente no complexo 1EWM foi excluído por apresentar uma amida invertida na cavidade ***S1***. A exclusão desta estrutura permite uma localização espacial mais precisa das propriedades doadora e aceptora de ligação de hidrogênio no modelo farmacofórico final. A **Figura 39**-A apresenta o alinhamento dos ligantes utilizados para a construção do modelo farmacofórico. Observa-se o perfeito alinhamento do grupamento fenila, de todos os inibidores na posição ***S2*** e a sobreposição dos átomos de oxigênio e de nitrogênio da amida central, em todos os ligantes.

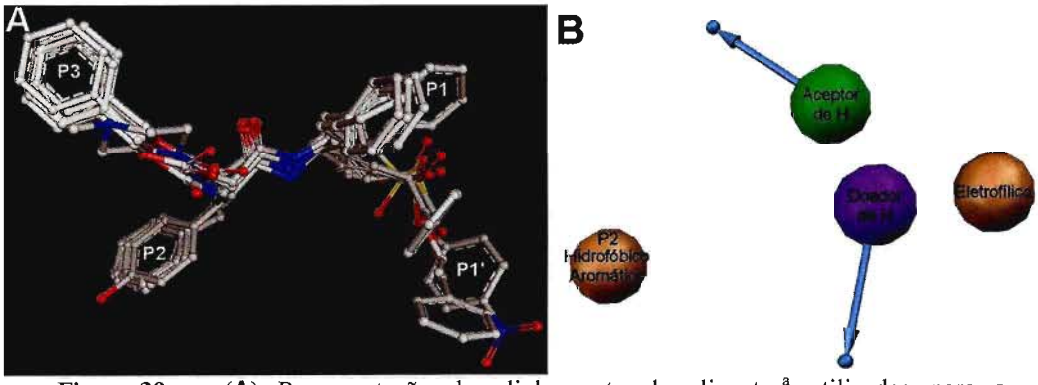


Figura 39. (**A**) Representação do alinhamento dos ligantes[a] utilizados para a construção do modelo farmacofórico. (**B**) Representação do modelo farmacofórico, gerado com o programa *CATALYST*,[Catalyst, 2005] a partir dos ligantes[a] alinhados.
[a] 1AIM; 1EWO; 1F29; 1F2A; 1F2B; 1F2C; 1ME3; 1ME4; e 1U9Q, respectivamente, alinhados no sítio ativo da cruzaína.

A **Figura 39**-B mostra uma representação do modelo farmacofórico gerado. O modelo contém o arranjo espacial, em três dimensões, de quatro regiões com diferentes propriedades, respectivamente: (*i*) uma região que deve ser ocupada por um grupo aceptor de ligação de hidrogênio; (*ii*) uma região que deve ser ocupada por um grupo doador de ligação de hidrogênio, (*iii*) uma região que deve ser ocupada por um grupamento hidrofóbico aromático e (*iv*) uma região que deve ser ocupada por um átomo eletrofílico. Adicionalmente, o modelo apresenta regiões de exclusão de volume (não mostradas na **Figura 39**-B) aleatoriamente espalhadas ao redor das regiões de interação supra citadas. A região, aceptora de ligação de hidrogênio, representa a interação com o nitrogênio da amida

da *Gly66* e a região, doadora de ligação de hidrogênio, representa a interação com o oxigênio carbonílico da *Asp158*. Estas regiões descrevem tanto a posição do grupamento envolvido na interação quanto a direção em que a ligação de hidrogênio deve ser formada. A região hidrofóbica aromática representa a ocupação da cavidade ***S2,*** do sítio ativo da cruzaína e a região eletrofílica corresponde à posição do átomo que pode sofrer o ataque nucleofílico do enxofre da *Cys25.*

Este modelo farmacofórico foi capaz de reconhecer os 9 inibidores da cruzaína utilizados na preparação do modelo. A aplicação deste modelo, sobre as 296.700 moléculas selecionadas na etapa anterior (filtro físico-químico), encontrou 308 compostos capazes de satisfazer as restrições espaciais impostas pelo modelo. Esta etapa representou uma redução de 99,9% no número de moléculas selecionadas pelo filtro físico-químico (296.700 moléculas).

3.3.3. Filtro de *docking*

Uma discussão detalhada sobre filtros de *docking* aplicados para a seleção de compostos bioativos, está apresentada no item **1.4.3.** da *Introdução* desta tese.

As moléculas selecionadas pelo filtro farmacofórico foram submetidas, a seguir, ao procedimento de *docking* no sítio ativo da cruzaína. Todos os procedimentos de *docking* foram executados pelo programa *GOLD,*[Verdonk *et al.*, 2003] e estão descritos no item **2.5.7.** (*Materiais e Métodos*).

Para a validação desta etapa, foram realizados dez *dockings* de cada um dos treze inibidores extraídos dos complexos *PDB,* sobre suas correspondentes estruturas da cruzaína. A similaridade, entre as poses apresentadas pelo programa e a pose encontrada no cristal foi avaliada pelo correspondente valor de *RMSD* calculado entre as duas poses. Os resultados foram avaliados quanto à capacidade do programa de prever, corretamente, a pose do ligante no sítio ativo da proteína e quanto à posição relativa da pose mais próxima ao cristal, em relação aos 10 *dockings* realizados para cada inibidor. A **Tabela 12** apresenta o valor de *RMSD* da solução mais próxima à pose no cristal e à posição relativa desta pose, em relação às outras poses, encontradas pelo programa *GOLD.*

Tabela 12. Resultados dos *dockings*[a] dos treze complexos do *PDB*[b], apresentando os correspondentes valores de *RMSD* da solução mais próxima à pose do cristal e à posição relativa [c] desta pose, em relação às outras possíveis poses encontradas

Código *PDB*	*RMSD* (Å)	Posição relativa [c]
1AIM	1,02	1
1EWL	2,16	5
1EWM	1,76	1
1EWO	1,34	1
1EWP	1,50	3
1F29	1,29	1
1F2A	1,89	1
1F2B	1,65 [d]	1
1F2C	1,34 [d]	1
1ME3	1,00	1
1ME4	1,35	1
1U9Q	1,31	1
2AIM	1,84	8

[a] Utilizando o programa *GOLD* [Verdonk *et al.*, 2003]
[b] Os valores de resolução dos cristais variaram de 1,2 a 2,2 (Å)
[c] Em relação aos 10 *dockings* executados por ligante
[d] No cálculo dos valores de *RMSD* para estes compostos foram excluídos os átomos da cavidade *S1'*

A **Figura 40** apresenta dois exemplos representativos dos resultados de validação do procedimento de *docking* utilizando-se os inibidores extraídos do *PDB*. São apresentadas a sobreposição da pose de maior valor de *Fitness* obtida pelo programa *GOLD* e da pose encontrada no cristal, para os inibidores dos complexos1ME3 (**Figura 40**-A) e 1F2B (**Figura 40**-B), respectivamente.

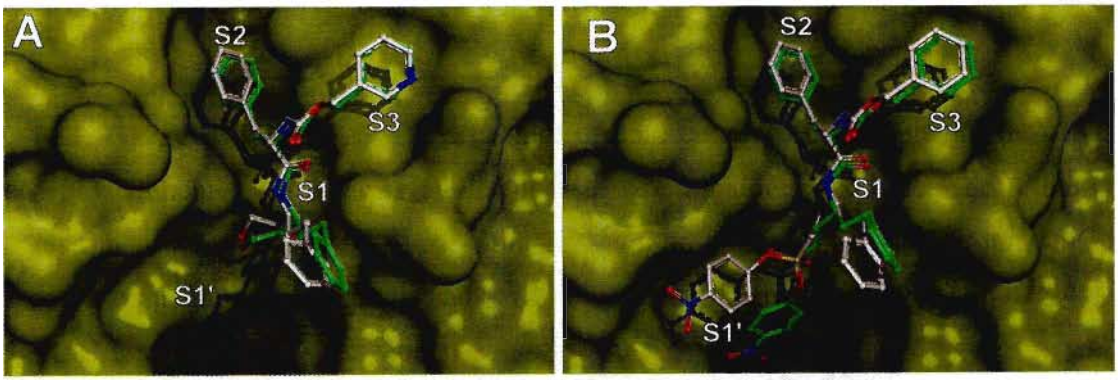


Figura 40. Sobreposição das poses de maior valor de *Fitness* obtidas pelo *GOLD 3.0* (branco) e as correspondentes poses encontradas no cristal (verde) para os ligantes dos complexos 1ME3 (**A**) e 1F2B (**B**), respectivamente. As regiões assinaladas *S3*, *S2*, *S1*, *S1'* correspondem às respectivas cavidades do sítio ativo.

Observa-se na **Figura 40**, o perfeito ajuste entre as poses obtidas pelo programa *GOLD* e as correspondentes poses encontradas extraídas dos complexos, para os inibidores

1ME3 (**Figura 40**-A) e 1F2B (**Figura 40**-B), respectivamente, em especial nas cavidades ***S2*** e ***S3*** do sítio ativo da cruzaína. Observa-se ainda na **Figura 40**-B, que não houve ajuste da porção do ligante 1F2B posicionada na cavidade ***S1'*** do sítio ativo. A justificativa para esta falta de ajuste é o fato de a cavidade ***S1'*** ser mais exposta ao solvente do que as cavidades ***S2*** e ***S3*** do sítio ativo da cruzaína, permitindo que o ligante se posicione livremente dentro da cavidade ***S1'***.

Conforme pode ser observado na **Tabela 12**, os valores de *RMSD*, das poses mais próximas à pose no cristal encontradas pelo programa, variaram de 1,00 a 2,16 Å. Adicionalmente, em dez dos treze *dockings* realizados, a pose mais próxima à do cristal foi também a de maior valor de *Fitness*, ou seja, a pose eleita pelo programa como a de maior ajuste. Estes resultados indicam que, na maior parte dos casos (10/13), o programa tanto é capaz de propor a pose correta, como também reconhecer esta pose como a de maior ajuste (*Fitness*), entre qualquer outra pose possível. Em apenas três casos, o programa não classificou corretamente a pose mais próxima ao cristal como a de maior valor de *Fitness*, respectivamente, nos complexos 1EWL, 1EWP e 2AIM. Em comum, estes três complexos (junto com o complexo 1F29), não apresentam um grupo aromático posicionado dentro da cavidade ***S2*** do sítio ativo da cruzaína. Sabidamente,[Cazzulo *et al.*, 2001] a cavidade ***S2*** tem afinidade pela cadeia lateral de resíduos de fenilalanina, ou seja, grupamentos hidrofóbicos aromáticos. A falta de grupamentos desta natureza na posição ***S2*** pode ter contribuído para que o programa não posicionasse corretamente os inibidores na cavidade da cruzaína. Estes resultados obtidos validam o procedimento de *docking*, utilizando o programa *GOLD*, para a previsão de modos de ligação de ligantes no sítio ativo da cruzaína.

Após a validação do procedimento de *docking*, os 308 compostos, selecionados pelo filtro farmacofórico na etapa anterior, foram submetidos a 10 *dockings* na estrutura da cruzaína, extraída do complexo 1ME4, utilizando-se as mesmas condições da etapa de validação. Os resultados foram avaliados em relação às poses de maior valor de *Fitness* atribuído pelo programa *GOLD,* para cada um dos compostos docados. Os 50 compostos com maiores valores de *Fitness* foram selecionados para serem analisados, por inspeção visual, dentro do sítio ativo da cruzaína, em suas poses de mínimo de energia atribuídas

pelo programa *GOLD*. A aplicação deste filtro de *docking* representou uma redução de 83,7% no número de moléculas selecionadas pelo filtro farmacofórico (308 moléculas).

A **Figura 41** apresenta as poses de maior valor de *Fitness* identificadas pelo programa *GOLD*, para os compostos ZINC04506284 (**Figura 41**-A) e **ZINC03326243** (**Figura 41**-B), posicionados dentro da cavidade do sítio ativo da cruzaína. São indicados os valores das distâncias das ligações de hidrogênio entre os átomos dos ligantes e os resíduos Gly66 e Asp158 da cavidade do sítio ativo da cruzaína.

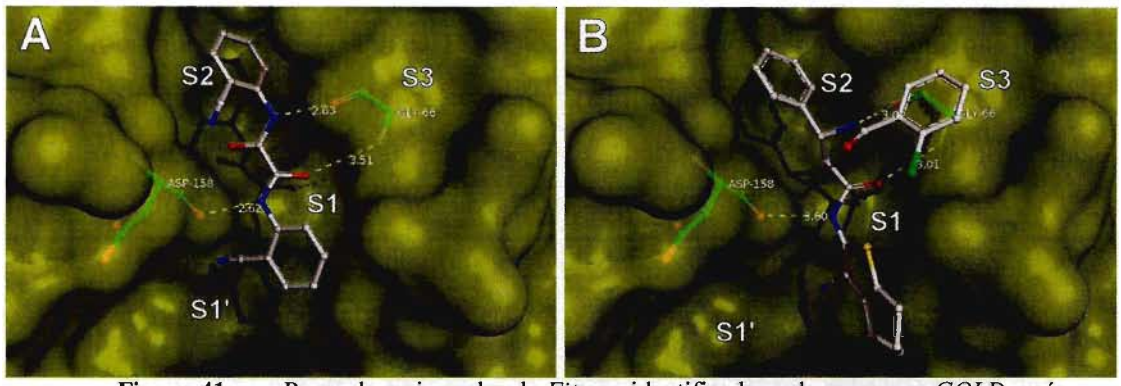


Figura 41. Poses de maior valor de *Fitness* identificadas pelo programa *GOLD*, após o procedimento de *docking* dos compostos ZINC04506284 (**A**) e **ZINC03326243** (**B**), posicionados dentro da cavidade do sítio ativo da cruzaína. São indicados os valores das distâncias das ligações de hidrogênio entre os átomos dos ligantes e os resíduos Gly66 e Asp158 da cavidade do sítio ativo da cruzaína. As regiões assinaladas *S3*, *S2*, *S1*, *S1'* correspondem às respectivas cavidades do sítio ativo.

3.3.4. Filtro visual e seleção dos compostos para o teste de inibição enzimática

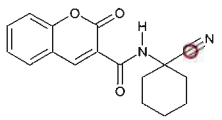
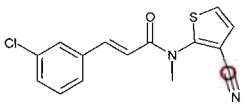
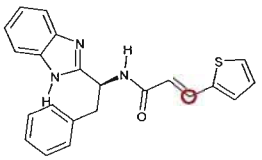
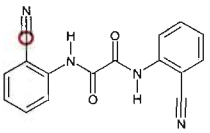
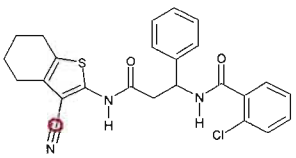
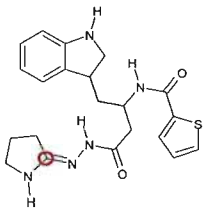
Conforme discutido na *Introdução* item **1.4.3.**, a inspeção visual dos resultados de *virtual screening* é uma etapa crítica no procedimento de *docking* uma vez que os modelos de seleção são rudimentares e não superam o conhecimento e experiência química do pesquisador envolvido com os aspectos multifatoriais da interação entre o ligante e seu alvo biológico.[Sotriffer *et al.*, 2004]

Os 50 compostos selecionados na etapa anterior de *docking* foram visualmente inspecionados, dentro do sítio ativo da cruzaína, avaliando-se a qualidade da interação entre o composto e os resíduos que compõem o sítio de ligação, levando-se em conta os seguintes critérios: (*i*) a reatividade do grupo eletrofílico; (*ii*) a distância e a geometria da possível interação entre o enxofre da *Cys25* e o átomo eletrofílico do ligante; (*iii*) a complementaridade estrutural entre o ligante e as cavidades da enzima; e (*iv*) a distância e a

geometria das ligações de hidrogênio entre os átomos oxigênio da *Asp158* e Nitrogênio da *Gly66* e o composto. O resultado desta análise visual selecionou 10 potenciais inibidores da cruzaína. Esta seleção por inspeção visual resultou em uma redução de 80,0% no número de moléculas selecionadas pelo filtro de *docking* (50 moléculas).

Dos dez compostos selecionados na última etapa do **Modelo I** (inspeção visual), apenas seis encontravam-se comercialmente disponíveis (**ZINC02470662**, **ZINC02682879**, **ZINC02663001**, **ZINC01936854**, **ZINC03326243**, **ZINC03192044**, respectivamente). Estes seis compostos foram adquiridos das correspondentes empresas fornecedoras. Esta etapa de aquisição dos compostos, representou uma redução de 40% em relação ao compostos selecionados pelo **Modelo I**. A **Tabela 13** mostra a representação estrutural dos seis compostos selecionados pelo **Modelo I** de *virtual screening* e que foram adquiridos para a validação experimental do modelo.

Tabela 13. Representação esquemática dos 6 compostos selecionados pelo **Modelo I** de inibidores da cruzaína e que foram adquiridos e submetidos ao teste de inibição enzimática. São apresentados os correspondentes valores de massa molecular, logP[a] e empresa fornecedora, de cada um dos compostos adquiridos. Os círculos em vermelho indicam o átomo de carbono eletrofílico susceptível ao ataque nucleofílico do enxofre catalítico da *Cys25*.

# Ref	Composto	Código *ZINC*	Peso Molecular	logP[a]	Vendedor
1		**ZINC02470662**	296,3	2,9	*Specs*
2		**ZINC02682879**	288,7	3,4	*Life Chemicals*
3		**ZINC02663001**	373,5	3,7	*Enamine Ltd*
4		**ZINC01936854**	290,3	1,3	*Asinex Limited*
5		**ZINC03326243**	463,99	4,6	*Enamine Ltd*
6		**ZINC03192044**	411,5	2,5	*Enamine Ltd*

[a]Valores calculados usando o programa XLOGP[Wang *et al.*, 1997].

3.3.5. Testes Enzimáticos e Validação Experimental do Modelo de *Virtual screening*

Conforme discutido na *Introdução* desta tese, item **1.4.**, os modelos de *virtual screening* devem ser validados experimentalmente para a confirmação de sua utilidade na

busca de compostos bioativos. Assim, os seis compostos selecionados pelo **Modelo I** de inibidores da cruzaína e que foram adquiridos das correspondentes empresas fornecedoras, foram submetidos ao teste de inibição enzimática da cruzaína, segundo o teste de inibição de *tight-binding* modificado,[Judice *et al.*, 2001] como descrito no item **2.6.** (*Materiais e Métodos*).

Dos seis compostos testados, três não mostraram inibição significativa da cruzaína até as concentrações testadas de 7,22 mM; de 38,71 mM e de 0,67 mM, para os compostos **ZINC02470662**, **ZINC02682879** e **ZINC03192044,** respectivamente. Em concentrações mais altas, estes compostos se tornaram insolúveis na solução tampão. A **Figura 42** apresenta os gráficos de inibição enzimática, mostrando o consumo de substrato pela cruzaína na presença de cada um dos compostos testados, em diferentes concentrações.

Os outros três compostos foram capazes de inibir, parcialmente, a cruzaína nas concentrações de 44,91 μM; de 17,26 μM e de 4,17 μM, para os compostos **ZINC02663001**, **ZINC01936854** e **ZINC03326243,** respectivamente. Conforme descrito no item **2.7.** (*Materiais e Métodos*), para verificar a possibilidade destes compostos estarem agindo segundo o mecanismo promíscuo de inibição enzimática,[McGovern *et al.*, 2002] foi determinado o efeito da adição de *Triton X-100* (0,1%, v/v) sobre a potência dos compostos (ver *Introdução* item **1.3.3.**) Para os três compostos que mostraram inibição parcial da cruzaína, a adição do detergente não-iônico *Triton X-100* diminuiu, significativamente, a potência fazendo com que a enzima retornasse a valores de atividade antes da inibição. Segundo o modelo de inibição promíscua[McGovern *et al.*, 2002], compostos que formam agregados moleculares a concentrações micro ou submicromolares são capazes de inibir inespecificamente enzimas de diferentes naturezas. A adição de um detergente não-iônico desfaz estes agregados moleculares e interrompe a inibição enzimática. A **Figura 42** apresenta os gráficos de inibição enzimática, mostrando o consumo de substrato pela cruzaína, na presença de diferentes concentrações de cada um dos compostos testados e o efeito da adição de *Triton X-100* nos testes onde houve inibição parcial da atividade da cruzaína.

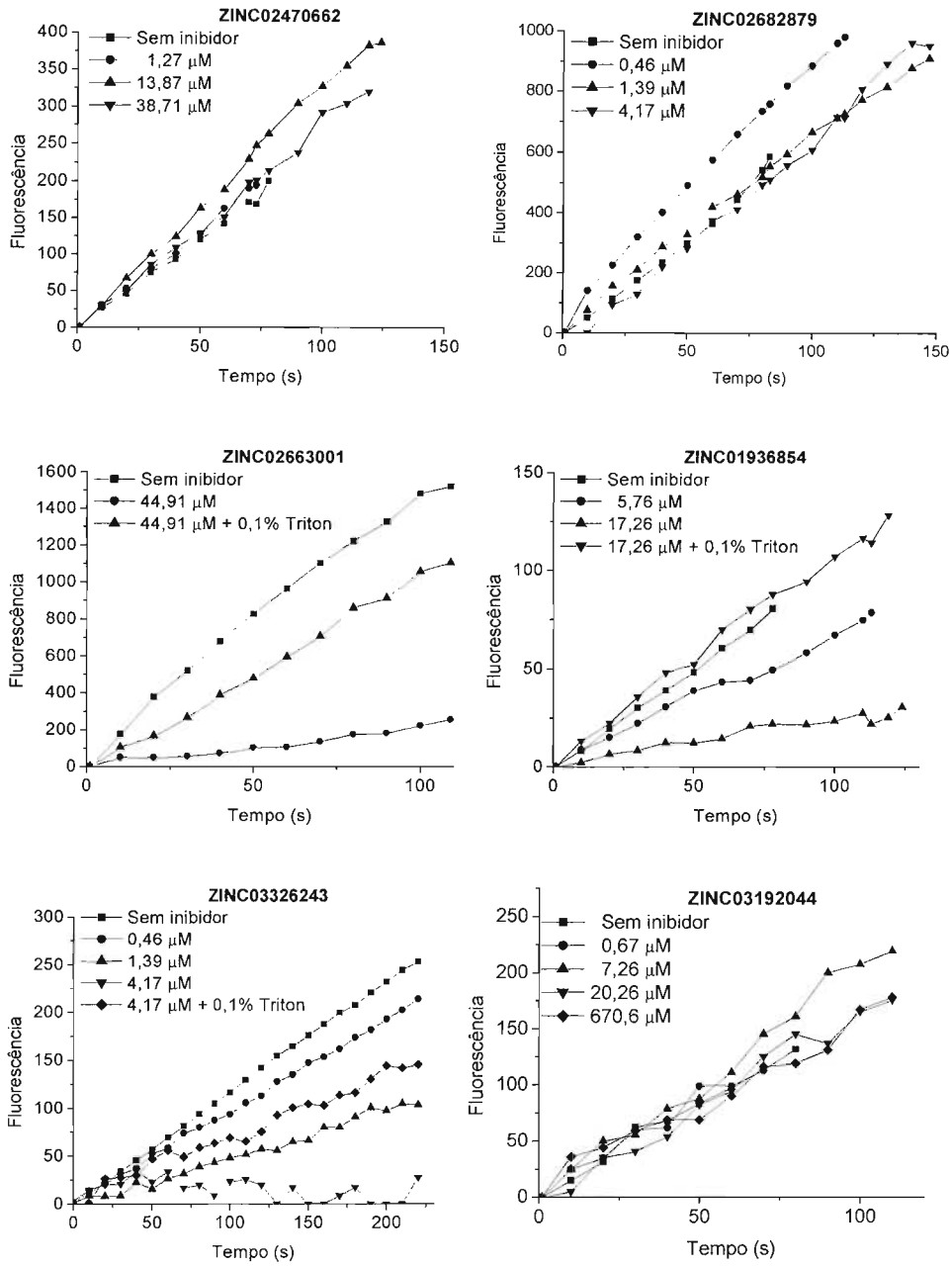


Figura 42. Gráficos dos valores da velocidade[a] de consumo do substrato[b] pela cruzaína[c] na presença de cada um dos compostos selecionados (usando pelo menos duas concentrações) e o efeito da adição de TritonX-100 (0,1%) nos testes onde foi observada inibição da atividade enzimática (compostos **ZINC02663001**, **ZINC01936854** e **ZINC03326243**, respectivamente).

[a] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e *EDTA* (2,5 mM), pH=6,5; sob agitação magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380nm; Em: 460nm).

[b] 5 μL de Bz-FR-MCA (170 μM).

[c] 0,3 μL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 μL de DTT

Conforme descrito no item **2.7.** (*Materiais e Métodos*), para confirmar a ocorrência de inibição enzimática, via mecanismo promíscuo para os compostos **ZINC02663001**, **ZINC01936854** e **ZINC03326243** foi realizado o teste do efeito da concentração da enzima sobre a potência inibitória dos compostos. Sabe-se que inibidores micromolares específicos, que agem segundo o mecanismo competitivo, não devem ser, significativamente, afetados por um aumento de nanomolar na concentração da enzima, pois a concentração dos inibidores é muitas vezes maior que a concentração da enzima. [McGovern *et al.*, 2002] Entretanto, inibidores promíscuos que formam agregados moleculares podem perder, significativamente, sua capacidade de inibição enzimática, em função do aumento na concentração da enzima.[Malvezzi *et al.*, 2008] Isto ocorre, pois os agregados moleculares e a enzima encontram-se em concentrações equivalentes e um pequeno aumento na concentração da enzima pode, facilmente, superar a capacidade de inibição dos agregados moleculares (ver *Introdução* item **1.3.3.**). Para os três compostos testados, foi observada perda significativa da potência de inibição da cruzaína, quando a concentração da enzima foi aumentada 10 vezes, corroborando a hipótese destes compostos estarem agindo segundo o mecanismo promíscuo de inibição enzimática. A **Figura 43** apresenta os testes enzimáticos, mostrando o consumo de substrato pela cruzaína na presença dos compostos, respectivamente, **ZINC02663001**, **ZINC01936854** e **ZINC03326243** em duas concentrações diferentes da enzima (0,18 nM e 1,8 nM). Em todos os casos, os compostos testados foram menos potentes com a enzima na concentração mais alta (1,8 nM) do que com a enzima na concentração mais baixa (0,18 nM)

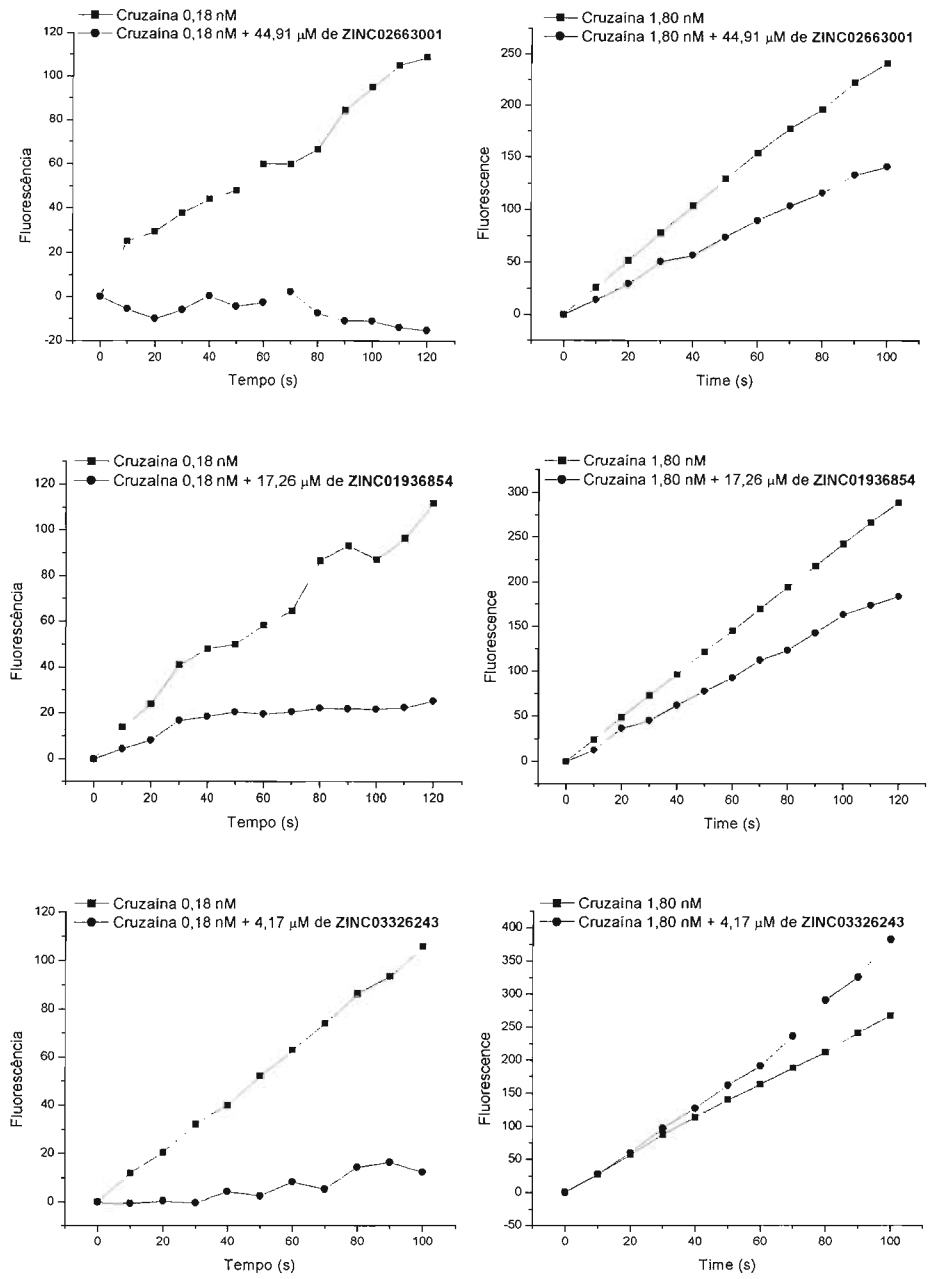


Figura 43. Gráficos dos valores[a] de consumo do substrato[b] pela cruzaína na presença dos compostos **ZINC02663001**, **ZINC01936854** e **ZINC03326243,** respectivamente. Os valores das concentrações da enzima foram aumentadas 10 vezes, respectivamente, de 0,18 nM (gráficos à esquerda) para 1,8 nM (gráficos à direita) para confirmar a ocorrência de inibição enzimática via mecanismo promíscuo.

[a] Valores determinados fluorimetricamente, em 1000μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5; sob agitação magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 μL de Bz-FR-MCA (170 μM).

Estes resultados indicam que, apesar do modelo de *virtual screening* ter sido desenvolvido para selecionar inibidores específicos da cruzaína, incluindo compostos potencialmente reativos frente ao enxofre catalítico de cisteíno-proteases, os compostos escolhidos e testados não se mostraram capazes de inibir a cruzaína, ou apresentaram inibição de acordo com o mecanismo promíscuo. Estes resultados mostram que, conforme discutido na *Introdução* desta tese, item **1.4.**, resultados falso-positivos ocorrem tanto em modelos de *virtual screening* quanto em testes enzimáticos experimentais.

Estes resultados negativos são esperados, uma vez que modelos de *virtual screening* são, naturalmente, apenas modelos de atividade e não se espera que sejam capazes de prever, corretamente, todas as possíveis interações ligantes e o alvo molecular. Dentre os dez compostos sugeridos pelo **Modelo I**, apenas seis compostos foram escolhidos para a validação experimental do modelo de *virtual screening*. Este número de compostos é pequeno tendo em vista o índice acerto de modelos de *virtual screening* em trabalhos publicados na literatura[Kubinyi, 2006] (entre 1/2 e 1/100 dos compostos sugeridos pelo modelo). Ou seja, considerando-se a ineficiência dos modelos é preciso testar vários compostos (pelo menos algumas dezenas) para que um, ou mais compostos ativos, possam ser encontrados. Modelos de *virtual screening* devem ser considerados apenas como ferramentas para enriquecer bibliotecas de compostos e não testes virtuais, propriamente ditos.

Uma das vantagens do uso de modelos de *virtual screening* para o desenvolvimento de ligantes é a velocidade com que estes modelos podem ser modificados e reaplicados sobre a biblioteca de compostos, sem qualquer custo adicional. Em função disso, o desenvolvimento de ligantes pode ser um processo interativo, ou seja, os resultados negativos de um modelo podem auxiliar na elaboração do próximo modelo, levando em conta as possíveis razões para o insucesso do primeiro. Dentre as possíveis causas para a falta de atividade dos compostos sugeridos pelo modelo de *virtual screening* estão a baixa reatividade de nitrilas aromáticas (quatro entre os seis compostos testados eram nitrilas aromáticas, compostos **ZINC02470662**, **ZINC02682879**, **ZINC01936854** e **ZINC03326243**) e a não ocupação da cavidade *S3* da cruzaína (três dos seis compostos testados eram relativamente pequenos e não ocupavam a cavidade *S3*, compostos **ZINC02470662**, **ZINC02682879** e **ZINC01936854**). Compostos pequenos, que não

ocupam mais de uma cavidade do sítio ativo da cruzaína, não apresentam afinidade suficiente pela enzima, a ponto de deslocar o substrato usado no teste de inibição.

Finalmente, as evidências de inibição pelo mecanismo promíscuo em três dos seis compostos testados, mostra que a inibição da cruzaína, nas condições experimentais utilizadas é passível deste tipo de artefato experimental e que estes resultados, falso-positivos, podem ser freqüentes em compostos selecionados por *virtual screening*. As razões que levam um composto a formar um agregado molecular capaz de inibir inespecificamente enzimas não são completamente conhecidas e, portanto, ainda não é possível prever com confiança este tipo de comportamento indesejado.[Malvezzi *et al.*, 2008; McGovern *et al.*, 2002] Para evitar estes resultados falso-positivos, a principal estratégia recomendada é o uso de detergentes (*Triton X-100*) no ensaio enzimático, evitando-se, assim, a formação dos agregados moleculares, que promove a inibição inespecífica da enzima.[Seidler *et al.*, 2003; McGovern *et al.*, 2003]

3.4. Obtenção e Validação de modelo de *virtual screening* de inibidores da cruzaína - Modelo II

A partir dos conhecimentos e experiências adquiridas na preparação do **Modelo I**, foi proposto um novo modelo de *virtual screening* de inibidores da cruzaína, introduzindo-se modificações em cada uma das etapas de elaboração do modelo. Foram propostas as seguintes modificações: (*i*) na série de compostos utilizados como referência para a preparação do modelo, (*ii*) nos critérios de cada um dos diferentes filtros aplicados sobre a biblioteca de compostos, (*iii*) na seqüência de aplicação dos diferentes filtros sobre a biblioteca de compostos, (*iv*) no número de compostos selecionados para a validação experimental do modelo e, (*v*) no procedimento de teste enzimático utilizado para a validação experimental do modelo.

A primeira mudança em relação ao **Modelo I** aplicada na elaboração deste **Modelo II** foi o aumento do número de inibidores de cisteíno-proteases utilizados como compostos de referências para a preparação do modelo. Além dos 13 inibidores extraídos dos complexos da cruzaína depositados no *PDB* utilizados no **Modelo I**, foram incluídos outros 20 inibidores extraídos de complexos de cisteíno-proteases identificados na busca por cavidades similares

BIBLIOTECA
INSTITUTO DE QUÍMICA
Universidade de São Paulo

à cruzaína, utilizando-se o programa *CAVBASE* (ver *Resultados e Discussão* itens **3.2.3.** e **3.2.4.**).

Assim, partir das estruturas de 33 complexos de cisteíno-proteases depositados no *PDB* (13 complexos da cruzaína + 20 complexos de outras cisteíno-proteases) foi proposto o **Modelo II** de *virtual screening* de inibidores da cruzaína. Foram aplicados, nesta seqüência, os filtros farmacofórico, físico-químico, de *docking* e de seleção por inspeção visual sobre o banco de dados *ZINC*[Irwin & Shoichet, 2005] versão 5. A seqüência de aplicação dos filtros foi alterada pois ambos os filtros farmacofórico e físico-químico são rápidos (i.e. podem ser aplicados sobre o banco de dados *ZINC* inteiro gastando-se apenas alguns minutos), entretanto, o filtro farmacofórico é mais específico que o filtro físico-químico, permitindo uma redução mais criteriosa no número de compostos do banco de dados *ZINC*. Como discutido na *Introdução*, item **1.4.**, a seqüência de aplicação dos filtros não deve alterar o resultado final da seleção, entretanto, esta afirmação só é válida quando os critérios de cada filtro já foram definidos. Quando a escolha dos critérios de cada filtro é dependente do número de compostos selecionados/rejeitados pelo filtro, a aplicação de filtros pouco seletivos, como o filtro físico-químico, nas primeiras etapas da seleção, pode resultar na perda de diversidade estrutural.

Cada um dos diferentes filtros virtuais foi validado individualmente utilizando-se as estruturas dos 33 inibidores de cisteíno-protease extraídos dos correspondentes complexos. Os compostos selecionados pelo **Modelo II** foram testados como inibidores da cruzaína. A **Figura 44** mostra a representação esquemática da seqüência de aplicação de cada filtro no **Modelo II**, juntamente com a porcentagem de redução da biblioteca de compostos em cada etapa.

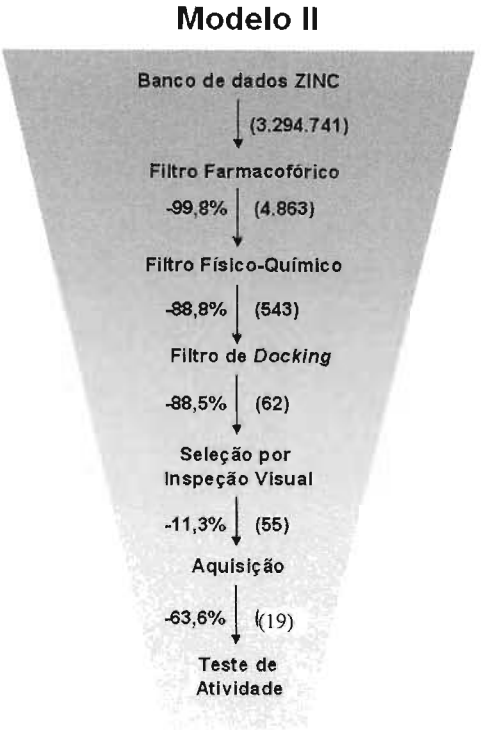


Figura 44. Representação esquemática da seqüência filtros aplicados sobre a biblioteca de compostos *ZINC*,[Irwin & Shoichet, 2005] representando o **Modelo II** de *virtual screening*. Em parêntesis são indicados os números de compostos selecionados em cada etapa e ao lado a correspondente porcentagem de redução por cada filtro.

3.4.1. Filtro farmacofórico

Uma discussão detalhada sobre filtros farmacofóricos aplicados para a seleção de compostos bioativos, está apresentada no item **1.4.2.** da *Introdução* desta tese.

A primeira etapa do procedimento de *virtual screening* foi a preparação e validação de um modelo farmacofórico de inibidores da cruzaína, para sua posterior aplicação como um filtro virtual sobre todo o banco de dados *ZINC)*.[Irwin & Shoichet, 2005] Todas as etapas envolvendo a preparação, a validação e a aplicação do modelo farmacofórico foram realizadas no programa *CATALYST*,[Catalyst, 2005] e estão descritas no item **2.5.5.** (*Materiais e Métodos*). Para a construção do modelo farmacofórico foram utilizadas as estruturas dos 13 inibidores extraídos dos complexos da cruzaína, depositados no *PDB* e 17 dos 20 inibidores selecionados, a partir da busca com o *CAVBASE*. Justifica-se a exclusão desses três compostos (1MEG, 2BDZ e 2DCD), pois estes foram rejeitados pelo programa *CATALYST*. Todos os 30 inibidores aceitos pelo programa *CATALYST* foram utilizados em suas conformações extraídas do complexo cristalino e alinhados pelos resíduos *Cys25*;

Gly66; *Asp158* e *His159* do sítio ativo. Suas estruturas foram modificadas para representar os compostos antes da formação do complexo covalente.

Durante o procedimento de criação do modelo farmacofórico, um modelo tentativo inicial é gerado e este é aplicado tanto sobre os 30 compostos de referência e validação, quanto sobre todos os compostos do banco de dados *ZINC*. Este procedimento é repetido inúmeras vezes e interativamente, visando-se maximizar o número de compostos de referência aceitos pelo modelo e, ao mesmo tempo, buscando-se minimizar o número de compostos do banco de dados *ZINC* aceitos pelo modelo.

Partindo-se de um modelo inicial simples, com apenas um componente, foi criado um modelo farmacofórico de inibidores da cruzaína composto de 5 propriedades físico-químicas distribuídas em um espaço de três dimensões. Duas regiões, que representam um espaço que deve ser ocupado por grupamentos hidrofóbicos, ou com carga positiva; uma região que deve ser ocupada por um grupo doador de ligação de hidrogênio, uma região que deve ser ocupada por um grupo aceptor de ligação de hidrogênio e uma região que deve ser ocupada por um átomo eletrofílico, suscetível ao ataque nucleofílico do enxofre da *Cys25*. A **Figura 45** mostra uma representação do modelo farmacofórico gerado.

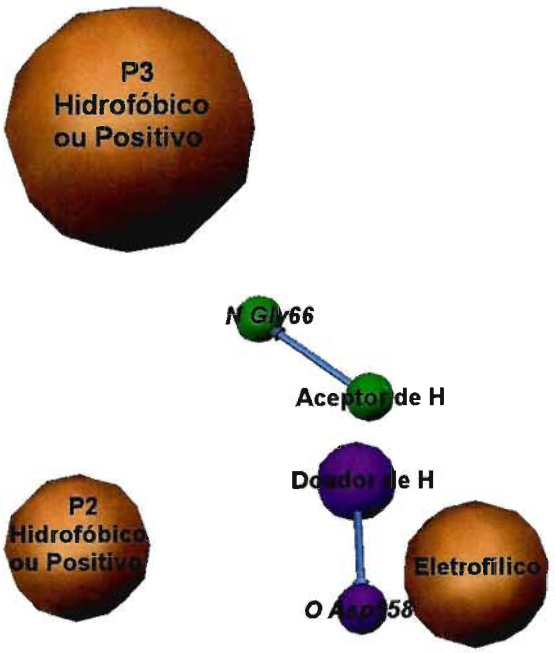


Figura 45. Representação do modelo farmacofórico de 5 pontos gerado a partir das estruturas dos ligantes extraídos dos 30[(a)] complexos, selecionados para a criação e validação do modelo de *virtual screening*
(a). 13 compostos extraídos dos complexos da cruzaína e 17 complexos selecionados, após a busca com o programa *CAVBASE*.

Cada uma das regiões hidrofóbicas/positivas corresponde a uma cavidade do sítio ativo da cruzaína, ou seja, às cavidades *S2* e *S3,* respectivamente. A região correspondente à cavidade *S2* tem um raio de 1,5 Å, enquanto que a região correspondente à cavidade *S3* tem um raio de 2,5 Å. Esta diferença se deve ao fato de a cavidade *S2* ser menor e menos exposta ao solvente do que a cavidade *S3*, fazendo com que os ligantes tenham menor liberdade para se posicionarem na cavidade *S2* do que na *S3*. O fato de se permitir tanto grupamentos positivamente carregados quanto grupamentos hidrofóbicos nas cavidades *S2* e *S3* é uma característica exclusiva[Otto & Schirmeister, 1997] da cruzaína, em relação a outras cisteíno-proteases. A cavidade *S2* é formada lateralmente pelos resíduos hidrofóbicos, respectivamente, *Leu67*, *Ala133* e *Met68,* e no seu fundo, encontra-se o resíduo *Glu205*. A cadeia lateral da *Glu205* tem flexibilidade e pode tanto expor para o interior da cavidade o grupo carboxilato negativamente carregado, e assim permitir a acomodação de um ligante positivamente carregado, bem como pode mover a porção terminal da sua cadeia lateral para o solvente, expondo para o interior da cavidade os dois grupamentos metilenos da sua cadeia intermediária, para poder assim acomodar ligantes hidrofóbicos. A cavidade *S3* é formada pelos resíduos *Leu67*, Asp60, Ser61 e Ser64, que entretanto, por se encontrar exposta ao solvente, pode igualmente acomodar tanto um grupamento carregado quanto um grupamento hidrofóbico. Todos os 30 (trinta) ligantes utilizados como referências para a criação (e posterior validação) do modelo farmacofórico apresentam grupamentos hidrofóbicos, ou positivamente carregados, posicionados tanto na cavidade *S2* quanto na *S3*.

As regiões doadora e aceptora de ligação de hidrogênio do modelo farmacofórico correspondem às posições complementares aos átomos oxigênio da *Asp158* e hidrogênio da amida da *Gly66*, respectivamente. O valor do raio da região aceptora de ligação de hidrogênio é de 0,5 Å enquanto que o da região doadora de ligação de hidrogênio é de 0,8 Å. A região doadora de ligação de hidrogênio é maior que a aceptora para poder acomodar amidas invertidas, como é o caso dos ligantes da cruzaína para *1EWL* e *1EWM.* Todos, exceto os inibidores dos complexos 1AU3 e 1AU4, apresentam uma ligação aceptora de hidrogênio com o -NH da *Gly66*, e uma ligação de hidrogênio doadora com o oxigênio carbonílico da *Asp158*. A região eletrofílica tem um raio de 1,5 Å e corresponde à posição do átomo que pode sofrer o ataque nucleofílico do enxofre da *Cys25*. Todos os 30

ligantes, utilizados como referência para a criação do modelo apresentam um átomo eletrofílico que se liga covalentemente ao enxofre da *Cys25*. Estes átomos eletrofílicos fazem parte dos seguintes grupos funcionais: aldeído, cetona, epóxido, vinil sulfona, nitrila, α-azometilcetona, α-clorometilcetona, α-flúormetil cetona, α-hidroximetilcetona, α-amido cetona e α-ester cetona. Além destes, são descritos na literatura[Otto & Schirmeister, 1997] inibidores de cisteíno-proteases com seguintes grupos funcionais: carbazona, dissulfeto e cetonas αβ-insaturadas. A partir destes grupos funcionais eletrofílicos presentes em inibidores da cruzaína e de outras cisteíno-proteases, foram definidos os tipos de átomos que devem ocupar a região eletrofílica do modelo farmacofórico. A **Figura 46** apresenta os grupos funcionais eletrofílicos suscetíveis ao ataque do enxofre da *Cys25*. As setas apontam para o átomo eletrofílico, que sofre o ataque nucleofílico em cada um dos grupamentos apresentados.

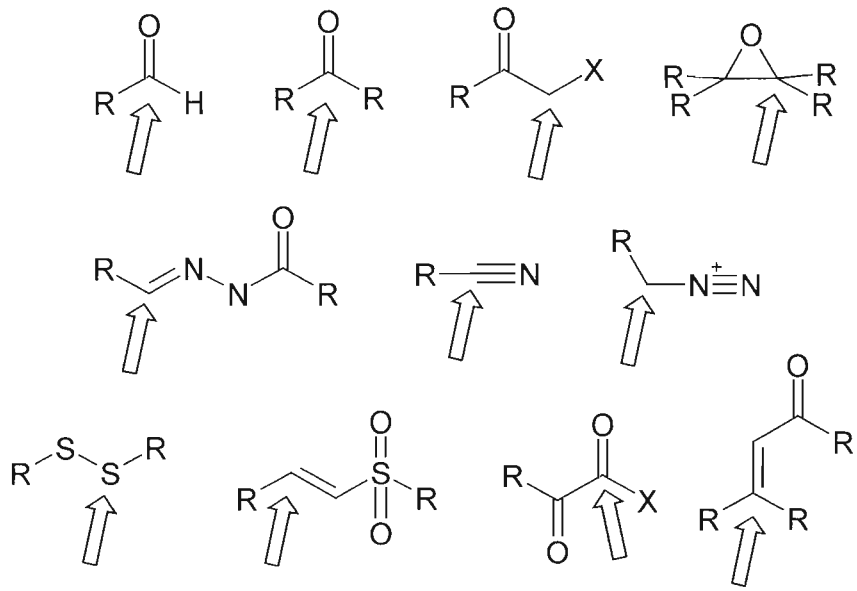


Figura 46. Grupos funcionais eletrofílicos suscetíveis ao ataque do enxofre da *Cys25*. As setas apontam para o átomo eletrofílico, que sofre o ataque nucleofílico em cada um dos grupamentos apresentados: Aldeído, Cetona, α-Halo Metil Cetona, Epóxido, Semicarbazona, Nitrila, Azo Composto, Dissulfeto, Vinilsulfona, α-Amido Cetona, e cetona α,β-insaturada.

O modelo famacofórico proposto foi capaz de reconhecer vinte e sete dos trinta compostos utilizados na criação e validação do modelo. Este desempenho representa uma identificação de 90% dos compostos de referência. Os três compostos não reconhecidos pelo modelo são 1EWL, 1U9Q e 1AU3.

A aplicação deste modelo sobre o banco de dados *ZINC* encontrou 15.847 compostos capazes de satisfazer todas as características do modelo. Em relação aos 3.294.741 compostos do banco de dados *ZINC*, a aplicação deste filtro farmacofórico representou uma redução de 99,51% do número de compostos.

Para refinar ainda mais o modelo farmacofórico e incluir informações sobre a estrutura da cruzaína, foram adicionadas regiões de exclusão de volume para representar o espaço correspondente aos átomos da cavidade da enzima ao redor dos 5 pontos de interação molecular que compõem o modelo farmacofórico desenvolvido. Para definir os pontos de exclusão de volume, foram usadas as coordenadas dos átomos dos resíduos que formam o sítio de ligação da cruzaína. As regiões de exclusão de volume foram criadas com raios de 0,5 Å. As regiões de exclusão de volume, correspondentes aos átomos da cadeia lateral da *Cys25,* foram excluídas do modelo, pois, como se pressupõe[Otto & Schirmeister, 1997] uma interação covalente, entre o átomo eletrofílico do ligante e o enxofre da *Cys25*, a distância que seria esperada entre estes átomos é interpretada pelo programa *CATALYST* como sendo um choque entre átomos e, assim, o ligante passa a ser rejeitado pelo modelo. A **Figura 47** representa o modelo farmacofórico com as respectivas regiões de exclusão de volume.

Este modelo farmacofórico (com as regiões de exclusão de volume) foi capaz de reconhecer vinte e dois dos vinte e sete compostos de referência reconhecidos pelo modelo anterior (sem as regiões de exclusão de volume). Este desempenho representa uma identificação de 75% em relação aos 30 (trinta) compostos inicialmente selecionados para a validação do modelo. As estruturas rejeitadas pelo modelo são .2DC6, 2G6D, 1AU4, 1KHP, 1KHQ e 2DCD.

A aplicação deste modelo final (com as regiões de exclusão de volume) sobre todo o banco de dados *ZINC* encontrou 4.863 compostos com as características do modelo farmacofórico. Esta etapa representou uma redução de 99,85% do banco de dados *ZINC* original (3.294.741).

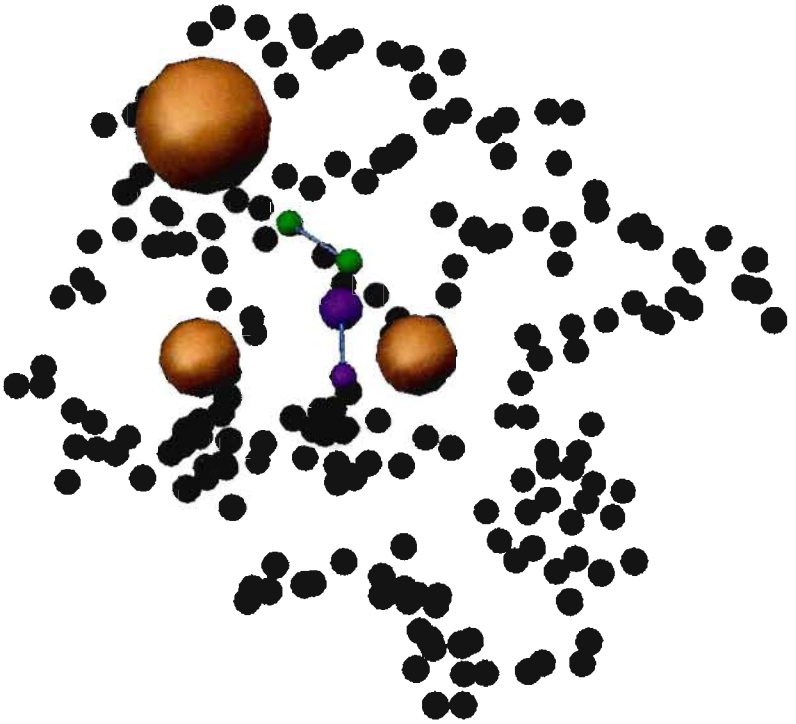


Figura 47. Representação do modelo farmacofórico de 5 pontos[(a)] com as correspondentes regiões de exclusão de volume, definidas a partir das coordenadas dos átomos dos resíduos que formam o sítio de ligação da cruzaína (estrutura extraída do complexo *PDB* 1ME4).
(a) Gerado a partir das estruturas dos ligantes, extraídos dos 30 complexos selecionados para a criação e validação do modelo de *virtual screening*

Este modelo farmacofórico preparado para o **Modelo II** apresentou quatro mudanças significativas, em relação ao modelo farmacofórico preparado para o **Modelo I**, a saber: (*i*) uma região hidrofóbica/positiva correspondendo a cavidade *S3* do sítio ativo da cruzaína; (*ii*) valores menores para o raio das regiões doadora e aceptora de ligação de menores; (*iii*) possibilidade de acomodação de grupamentos hidrofóbicos *ou positivamente carregados* na cavidade *S2* e; (*iv*) regiões de exclusão de volume representando todos os átimos da cavidade de ligação da proteína. Todas estas modificações tornam este modelo farmacofórico (preparado para o **Modelo II**) mais específico e mais restritivo para seleção de inibidores da cruzaína.

3.4.2. Filtro Físico-Químico

Uma discussão detalhada sobre filtros virtuais de compostos, baseados em propriedades físico-químicas está apresentada no item **1.4.1** da *Introdução* desta tese.

A segunda etapa do procedimento de *virtual screening* foi a aplicação de um filtro de propriedades físico-químicas sobre os compostos selecionados na etapa anterior. Todas os procedimentos nesta etapa foram executados utilizando-se o programa DBfilter,[Wang, 2003] e estão descritos no item **2.5.6.** (*Materiais e Métodos*). As propriedades físico-químicas utilizadas como filtros foram baseadas na regra-dos-5 de Lipinski,[Lipinski *et al.*, 1997] e os correspondentes valores de máximo e de mínimo foram derivados dos treze compostos, extraídos dos complexos da cruzaína e dos vinte compostos selecionados na busca de cavidades similares à cruzaína utilizando o programa *CAVBASE*. A **Tabela 14** apresenta os valores do massa molecular, do logP, do número de grupos doadores de H, do número de grupos aceptores de ligação de H, da soma dos átomos de Nitrogênios e Oxigênio, do número de halogênios, do número de anéis, do tamanho dos anéis, do número de ligações com rotação livre e do número de grupos metilenos em cadeias alifáticas para os 33 inibidores da cruzaína selecionados

A partir das características estruturais dos 33 inibidores selecionados, foram determinados valores, respectivamente, de máximo e de mínimo das propriedades físico-químicas utilizadas para selecionar inibidores da cruzaína. A **Tabela 15** apresenta as propriedades físico-químicas utilizadas e os seus correspondentes valores de máximo e mínimo, utilizados para selecionar inibidores da cruzaína.

Tabela 14. Valores do peso molecular, do *logP*, do número de grupos doadores de H, do número de grupos aceptores de ligação de H, da soma dos átomos de Nitrogênios e Oxigênio, do número de halogênios, do número de anéis, do tamanho dos anéis, do número de ligações com rotação livre e, do número de grupos metalinas em cadeias alifáticas, para os 33 inibidores de cisteíno-protease selecionados

Código *PDB*	Massa molecular	logP	Número							
			Doadores de H	Aceptores de H	N + O	Halogênios	Anéis	Tamanho do anel	Ligações com rotação livre	$-(CH_2)_nCH_3$
1AIM	402	2,6	3	5	7	1	2	6	10	1
1EWL	364	1,7	4	5	7	0	1	6	11	1
1EWM	364	1,7	4	5	7	0	1	6	11	1
1EWO	524	4,2	2	6	8	0	3	6	14	0
1EWP	314	0,3	2	4	7	0	1	6	8	1
1F29	562	4,2	2	5	8	0	4	6	11	0
1F2A	597	6,7	2	5	7	0	4	6	15	0
1F2B	644	6,6	2	8	11	0	4	6	16	0
1F2C	619	4,7	2	7	10	0	4	6	15	1
1ME3	476	2,4	3	6	8	0	3	6	13	0
1ME4	461	3,3	3	5	7	0	3	6	12	0
1U9Q	452	3,6	2	6	8	0	2	6	13	2
2AIM	395	1,4	5	5	9	1	1	6	13	1
1AEC	356	-0,2	6	6	10	0	1	3	11	1
1ATK	356	-1,4	6	5	10	0	1	3	11	1
1AU3	595	5,0	3	7	11	0	3	6	15	1
1AU4	582	5,6	3	8	11	0	3	6	16	1
1FH0	575	4,3	2	5	8	0	4	6	11	1
1KHP	359	0,8	3	6	8	0	1	6	11	1
1KHQ	360	0,8	3	6	8	0	1	6	11	1
1MEG	356	-0,6	5	6	10	0	1	3	11	1
1MEM	527	3,9	3	5	8	0	3	6	11	1
1MS6	402	1,3	2	5	8	0	2	6	9	1
1NQC	456	2,6	2	4	7	0	3	6	10	0
1POP	427	1,9	6	5	10	0	0	0	14	1
1SNK	441	5,3	2	4	6	1	1	6	12	2
1THE	518	2,4	5	6	10	1	2	6	16	0
1TU6	444	3,6	3	5	8	1	3	6	12	1
2BDZ	356	-0,6	5	6	10	0	1	3	11	1
2DC6	474	2,2	2	6	9	0	3	6	12	2
2DCD	431	1,1	3	6	9	0	3	6	9	2
2FRQ	509	3,3	2	8	9	0	3	6	8	1
2G6D	640	4,0	2	7	10	0	4	6	13	0

Tabela 15. Valores das propriedades físico-químicas utilizadas para selecionar inibidores da cruzaína

Propriedade Físico-Química	Valor (Min e Máx)	
Átomos permitidos	**H;C;N;O;F;Cl;Br;S**	
Peso Molecular	**300**	**700**
logP	**-2**	**7**
Grupos doadores de ligação de Hidrogênio	**2**	**6**
Grupos aceptores de Ligação de Hidrogênio	**4**	**8**
Soma dos átomos de Oxigênio e Nitrogênio	**6**	**11**
Halogênios	**0**	**3**
Número de anéis	**0**	**4**
Tamanho do anel	**0**	**6**
Ligações com rotação livre	**0**	**16**
Extensão da cadeia alifática $CH_3(CH_2)_n$-	**0**	**2**

(a) **Derivados dos 33 inibidores de cisteíno-protease selecionados para a criação do modelo de *virtual screening***

Em relação aos valores dos 33 compostos de referência, dois deles apresentados na **Tabela 15** foram alterados: o número máximo de átomos halogênios foi aumentado de 1 para 3, para que também fossem aceitos compostos com o grupo funcional trifluormetil e o número de ligações com rotação livre foi considerado a partir do zero, para incluir também compostos rígidos ou com menor flexibilidade (o valor derivado dos compostos de referência seria um mínimo de 8 ligações com rotação livre). Em relação aos valores usados no filtro físico-químico do **Modelo I**, todos os valores foram alterados (exceto o "número de anéis" e a "extensão da cadeia alifática $CH_3(CH_2)_n$-"). Estas mudanças resultaram na possibilidade de seleção de compostos maiores, mais lipofílicos, mais flexíveis e com maior diversidade estrutural, do que os selecionados no **Modelo I**.

Como os valores das propriedades físico-químicas apresentados na **Tabela 15** foram derivados diretamente dos compostos selecionados para a validação do modelo, nenhum destes compostos seriam rejeitados pela aplicação destes filtros. A aplicação destes filtros físico-químicos sobre as 4.863 moléculas, selecionadas pelo filtro farmacofórico da etapa anterior encontrou 543 moléculas com as mesmas propriedades dos compostos de referência. A aplicação deste filtro representou uma redução de 88,83% do número de moléculas selecionadas pelo filtro farmacofórico (4.863 moléculas). O critério de seleção mais restritivo, ou seja, o critério que mais rejeitou os compostos selecionados pelo filtro farmacofórico foi o número de grupos doadores de ligação de hidrogênio. A maior parte (aproximadamente 80%) dos compostos selecionados pelo filtro farmacofórico apresentava

apenas um grupamento doador de ligação de hidrogênio, enquanto que o critério de seleção utilizado requeria a presença de, no mínimo, dois grupamentos doadores de ligação de hidrogênio.

3.4.3. Filtro de *Docking*

Uma discussão detalhada sobre filtros de *docking* aplicados para a seleção de compostos bioativos, está apresentada no item **1.4.3.** da *Introdução* desta tese.

A terceira etapa do procedimento de *virtual screening* foi a aplicação de um filtro de *docking* sobre os 543 compostos selecionados na etapa anterior. Todos os procedimentos de *docking* foram realizados utilizando-se o programa *GOLD,*[Verdonk *et al.*, 2003] e estão descritas no item **2.5.7.** (*Materiais e Métodos*).

Para a validação do procedimento de *docking,* os inibidores - extraídos dos 33 complexos selecionados para a criação e validação do modelo de *virtual screening*, ou seja, os 13 compostos extraídos dos complexos da cruzaína, e os 20 complexos de outras cisteíno-proteases, selecionados após a busca de cavidades similares à cruzaína com o programa *CAVBASE* - foram docados sobre a estrutura da cruzaína, como depositada no complexo 1ME4, utilizando-se os parâmetros *default* do programa *GOLD.* Foram utilizadas as estruturas dos ligantes modificadas para representar suas estruturas antes da formação do complexo covalente. Isto representa mudança significativa em relação ao procedimento de validação do *docking* do **Modelo I**, que utilizou as estruturas dos inibidores diretamente como extraídas dos correspondentes complexos com a cruzaína. Adicionalmente, na validação do **Modelo I**, os ligantes foram docados sobre as correspondentes estruturas da cruzaína, com as quais eles foram complexados. A justificativa para estas mudanças é que o programa de *docking* deve ser capaz de reconhecer o inibidor antes deste reagir com a enzima e uma única estrutura da cruzaína deve ser usada na etapa de validação e aplicação do filtro de *docking*.

Para a análise dos resultados do *docking*, foram consideradas apenas as poses com o maior valor de *Fitness,* para cada um dos inibidores docados. O resultado do *docking* foi considerado corretamente posicionado, somente quando a pose de maior valor de *Fitness* apresentasse as seguintes características: ocupação da cavidade *S2* da enzima; ligações de hidrogênio com os átomos oxigênio da *Asp158* e de nitrogênio da *Gly66* e proximidade

(distância $\leq$ 2Å) entre o átomo eletrofílico do ligante e o enxofre da *Cys25*. Assumiu-se que, caso qualquer um destes critérios não fosse observado, o resultado do *docking* seria considerado incorreto. Adicionalmente, foi calculado o valor de *RMSD* entre a pose do cristal e a pose com o maior valor de *Fitness*. A **Tabela 16** apresenta os valores, respectivamente, de *Fitness* e de *RMSD* calculados para as poses de maior valor de *Fitness,* para cada um dos inibidores docados.

Conforme pode ser observado na **Tabela 16**, dos 33 ligantes docados, apenas 14 foram corretamente posicionados pelo programa *GOLD*, segundo os critérios de avaliação dos resultados do *docking* descritos acima. Este resultado indica que, aplicando-se os seus parâmetros *default*, o programa *GOLD* erra o posicionamento de 60% dos compostos usados para a validação. Este é um resultado que pode ser considerado não satisfatório, levando-se em conta que os compostos usados nesta etapa de validação são todos ligantes específicos, com alta afinidade pelo sítio de ligação da cruzaína, ou de cisteíno-proteases similares.

Dentre as possíveis razões para a incapacidade do programa *GOLD* de prever corretamente a pose de ligação dos compostos selecionados para a validação está o fato do programa não ser capaz tanto de formar, ou de romper, ligações covalentes entre átomos bem como de alterar a estrutura do ligante docado.

Todos os inibidores utilizados nesta etapa de validação do procedimento de *docking* são covalentemente ligados ao enxofre da *Cys25* e suas estruturas não são, portanto, aquelas dos compostos que se encontram ligados no sítio ativo da enzima. Isto ocorre, pois, após o ataque nucleofílico da enzima os compostos sofrem mudanças estruturais, como por exemplo a saída de grupos de partida e mudanças na hibridização (por exemplo, sp^2 para sp^3). Por estas razões, o programa *GOLD* é incapaz de acomodar, corretamente, o grupo eletrofílico dos inibidores, responsáveis pela formação da ligação covalente com a enzima. Deste modo, o inibidor assume posição diferente daquela descrita originalmente no complexo cristalino. Em outras palavras, para o programa *GOLD*, a presença do enxofre da *Cys25* impede a aproximação do ligante, pois seu raio de *Van der Waals* passa a representar uma barreira contra a aproximação de qualquer inibidor covalente e, conseqüentemente, no resultado do *docking* o posicionamento dos ligantes é incorreto.

Tabela 16. Valores de Fitness[(a)] e de RMSD[(b)], calculados para os correspondentes 33 compostos da série de validação docados [(c)] sobre o sitio de ligação da estrutura da cruzaína, extraída do complexo 1ME4. Os valores em vermelho indicam os ligantes que foram corretamente posicionados, segundo os critérios de avaliação dos resultados do *docking* descritos no texto.

Complexo PDB	*Fitness*	*RMSD*
1AEC	47,11	9,4
1AIM	49,86	2,0
1ATK	46,03	9,1
1AU3	75,66	1,5
1AU4	55,02	1,7
1EWL	40,00	5,1
1EWM	45,54	8,1
1EWO	63,82	6,0
1EWP	45,47	5,3
1F29	66,40	1,1
1F2A	69,16	10,7
1F2B	65,38	3,8
1F2C	68,02	1,8
1FH0	64,28	7,5
1KHP	55,94	7,5
1KHQ	52,58	4,5
1ME3	64,10	3,8
1ME4	69,80	2,6
1MEG	54,98	8,7
1MEM	55,05	9,0
1MS6	56,69	2,9
1NQC	63,97	3,6
1POP	52,42	7,3
1SNK	48,35	1,9
1THE	47,95	10,3
1TU6	48,43	3,1
1U9Q	58,86	1,9
2AIM	42,61	7,5
2BDZ	48,07	3,1
2DC6	59,18	13,4
2DCD	55,80	2,0
2FRQ	59,44	7,5
2G6D	63,46	2,3

(a) O valor de *Fitness* é a medida qualidade da interação entre o ligante e o sítio ativo da proteína. Valores maiores de *Fitness* indicam maior afinidade entre o ligante e a proteína [Verdonk *et al.*, 2003]
(b) Diferença entre a pose no cristal e a pose com o maior valor de *Fitness*.
(c) Utilizando-se os parâmetros *default* do programa *GOLD* .[Verdonk *et al.*, 2003] Para cada ligante foram executados 10 *dockings*.

Para contornar esta limitação do programa *GOLD*, foi proposta, neste trabalho, a utilização da estrutura de uma cruzaína mutante, *C25G*, com uma Glicina substituindo a da *Cys25*. A justificativa para esta modificação estrutural da enzima é que a glicina, por não ter uma cadeia lateral, supõe-se que não deva apresentar o mesmo grau de impedimento

estérico contra a aproximação de ligantes que a cisteína. Adicionalmente, além da utilização da enzima mutante, foram propostas e introduzidas outras mudanças na função de ranqueamento (*score function*) do programa, para favorecer o posicionamento correto dos compostos no sítio de ligação da enzima. Estas mudanças, descritas a seguir, promovem o aumento nos valores de *Fitness* para os compostos que tanto satisfazem certas restrições espaciais bem como que estabelecem ligações específicas com a enzima, ajudando a selecionar as poses, que são corretamente posicionadas pelo programa de *docking*.

As mudanças introduzidas na função de ranqueamento do programa *GOLD* visam reproduzir as mesmas condições espaciais impostas pelo modelo farmacofórico utilizado anteriormente, quais sejam:

(i) *Protein HBond Constraints* (ligação de hidrogênio). Os átomos de oxigênio da *Asp158*, e nitrogênio da *Gly66* foram indicados como sendo os átomos da proteína que devem obrigatoriamente formar ligações de hidrogênio com o ligante. Foram mantidos os valores *default* do programa para a contribuição de compostos que satisfazem esta condição.

(ii) *Region Constraints* (Restrições espaciais). Foram criadas três regiões espaciais, para indicar três posições que os ligantes devem obrigatoriamente ocupar ao se posicionarem no sítio de ligação da enzima. Estas três regiões têm, exatamente, os mesmos tamanhos e posições que as das três regiões correspondentes criadas no modelo farmacofórico. Deste modo, uma região indica o posicionamento de átomos hidrofóbicos do ligante na cavidade *S2* da enzima, uma outra região indica o posicionamento de átomos hidrofóbicos do ligante na cavidade *S3* da enzima e a última região indica o posicionamento de qualquer átomo do ligante próximo ao átomo de enxofre da *Cys25*.

A validação do procedimento de *docking* foi repetida, considerando-se, desta vez, o *docking* dos 33 inibidores da série de validação sobre a estrutura da cruzaína mutante *C25G* e incluindo-se as mudanças na função de ranqueamento descritas acima. Os critérios para a avaliação dos resultados do *docking* foram os mesmos descritos para a validação anterior. De modo análogo, foram calculados os valores de *RMSD* entre a pose do cristal e a pose

com o maior valor de *Fitness*. A **Tabela 17** apresenta os valores de *Fitness* e de *RMSD* calculados para as poses de maior valor de *Fitness*.

Tabela 17. Valores de Fitness[(a)] e de RMSD[(b)], calculados para os 33 compostos da série de validação docados [(c)] sobre o sitio de ligação da estrutura da cruzaína 1ME4, modificada com a mutação *C25G*. Os valores em vermelho indicam os ligantes que foram corretamente posicionados, segundo os critérios de avaliação dos resultados do *docking* descritos no texto.

Complexo PDB	*Fitness*	*RMSD*
1AEC	48,84	2,6
1AIM	61,68	1,3
1ATK	51,22	1,4
1AU3	77,42	1,7
1AU4	66,79	3,9
1EWL	55,13	5,5
1EWM	58,66	0,8
1EWO	73,83	1,6
1EWP	43,51	2,3
1F29	71,28	7,3
1F2A	87,31	2,5
1F2B	62,19	4,3
1F2C	64,71	6,7
1FH0	67,48	8,2
1KHP	58,98	2,3
1KHQ	59,65	2,5
1ME3	74,33	1,5
1ME4	79,13	2,5
1MEG	55,36	3,6
1MEM	61,73	7,9
1MS6	57,48	1,6
1NQC	73,87	7,9
1POP	47,01	2,4
1SNK	55,42	1,9
1THE	45,72	10,6
1TU6	55,03	6,0
1U9Q	61,39	2,1
2AIM	46,70	7,8
2BDZ	49,04	2,9
2DC6	62,03	10,8
2DCD	61,63	1,6
2FRQ	59,55	4,1
2G6D	60,86	8,0

(a) O valor de *Fitness* é a medida qualidade da interação entre o ligante e o sítio ativo da proteína. Valores maiores de *Fitness* indicam maior afinidade entre o ligante e a proteína

(b) Diferença entre a pose no cristal e a pose com o maior valor de *Fitness*.

(c) Utilizando-se as mudanças na função de ranqueamento como descritas no texto. Para cada ligante foram executados 10 *dockings*

.

Conforme pode ser observado na **Tabela 17**, dos trinta e três ligantes docados sob as novas condições da função de ranqueamento, vinte e três foram corretamente posicionados pelo programa *GOLD*, ou seja, ocupando as cavidades *S2* e *S3* da enzima, estabelecendo, deste modo, ligações de hidrogênio tanto com os átomos de oxigênio da *Asp158* e de nitrogênio da *Gly66*, bem como com o átomo eletrofílico do ligante próximo (distância ≤ 2 Å) ao átomo de enxofre da *Cys25*. Estes resultados indicam que, com os novos parâmetros estabelecidos para o procedimento de *docking*, o programa *GOLD* posicionou corretamente 70% dos compostos usados para a validação. Estes resultados são significativamente melhores do que o procedimento de validação realizado anteriormente, ou seja, utilizando a enzima não-mutante e os parâmetros *default* do programa. A **Figura 48** apresenta quatros exemplos representativos dos resultados de validação do procedimento de *docking* utilizando-se os inibidores extraídos do *PDB*. São apresentadas a sobreposição da pose de maior valor de *Fitness* obtida pelo programa *GOLD* e da pose encontrada no cristal, para os inibidores dos complexos 1ATK (**Figura 48**-A); 1AU3 (**Figura 48**-B), 1MS6 (**Figura 48**-C) e 2DCD (**Figura 48**-D), respectivamente.

A seguir, o procedimento de *docking* como descrito e validado acima, foi aplicado sobre os 543 compostos previamente selecionados pelo filtro físico-químico. Para cada um dos compostos foram realizados 10 (dez) *dockings* sobre a cruzaína mutante *C25G*, aplicando-se as mudanças na função de ranqueamento apresentadas durante a etapa de validação. As poses com maior valor de *Fitness* de cada um dos 543 compostos docados foram analisadas automaticamente pelo programa *GOLD*. O programa selecionou todos os compostos que apresentavam valores de *Fitness* maiores que 50 e que, ao mesmo tempo, foram posicionados ocupando as cavidades *S2* e *S3* da enzima; estabelecendo ligações de hidrogênio com os átomos de oxigênio da *Asp158* e de Nitrogênio da *Gly66* e com qualquer átomo próximo ao átomo de enxofre da *Cys25*. De acordo com estes critérios, 119 compostos foram selecionados. Cada um destes compostos foi individualmente examinado, por inspeção visual, dentro da cavidade da cruzaína *1ME4*, na pose sugerida pelo programa *GOLD*. Nesta primeira análise visual, foram eliminados todos os compostos cujo átomo eletrofílico foi posicionado distante do enxofre da *Cys25*, em uma posição visivelmente incompatível com a possível formação de uma ligação covalente entre o composto e a enzima. De acordo com esta avaliação preliminar, foram selecionados 62 compostos. A

aplicação deste filtro de *docking* representou uma redução de 88,5% no número de moléculas selecionadas pelo filtro físico-químico (543 moléculas).

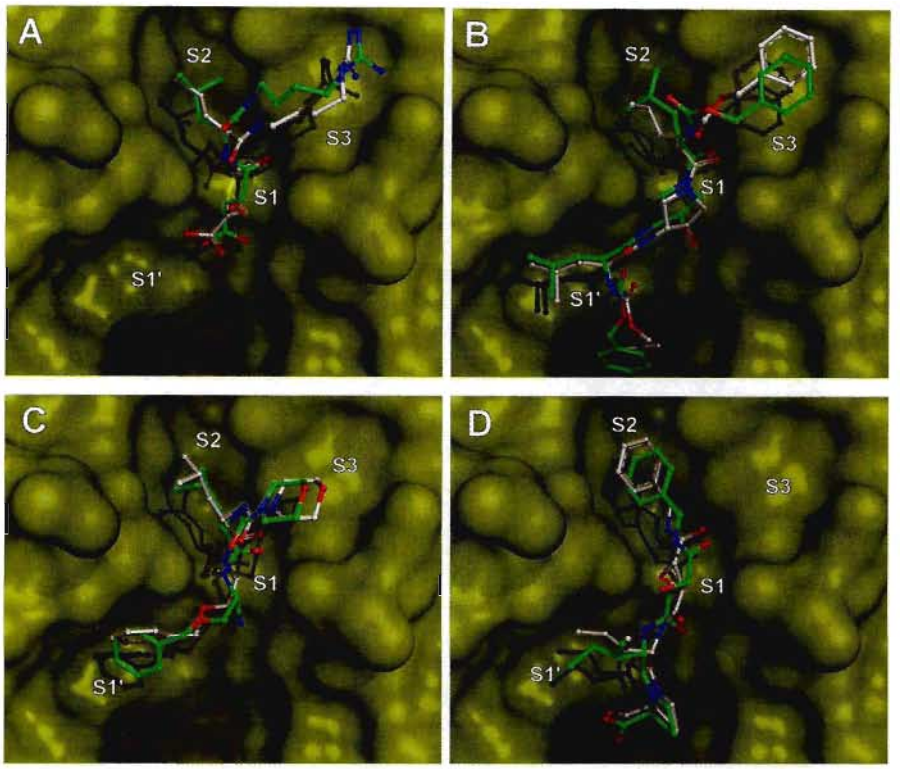


Figura 48. Sobreposição da pose de maior valor de *Fitness* obtida pelo *GOLD 3.0* (branco) e a correspondente pose encontrada no cristal (verde) para os ligantes extraídos dos seguintes complexos: 1ATK (**A**); 1AU3 (**B**), 1MS6 (**C**) e 2DCD (**D**), respectivamente. As regiões assinaladas *S3*, *S2*, *S1*, *S1'* correspondem às respectivas cavidades do sítio ativo.

A **Figura 49** mostra quatros exemplos representativos dos resultados de aplicação do procedimento de *docking* sobre os compostos selecionados pelo filtro físico-químico. São apresentadas as poses de maior valor de *Fitness* identificadas pelo programa *GOLD*, para os compostos **ZINC02338249** (**Figura 49**-A) **ZINC02342142** (**Figura 49**-B), **ZINC02399860** (**Figura 49**-C) e **ZINC02400672** (**Figura 49**-D), posicionados dentro da cavidade do sítio ativo da cruzaína. São indicados os valores das distâncias das ligações de hidrogênio entre os átomos dos ligantes e os resíduos *Gly66* e *Asp158* da cavidade do sítio ativo da cruzaína.

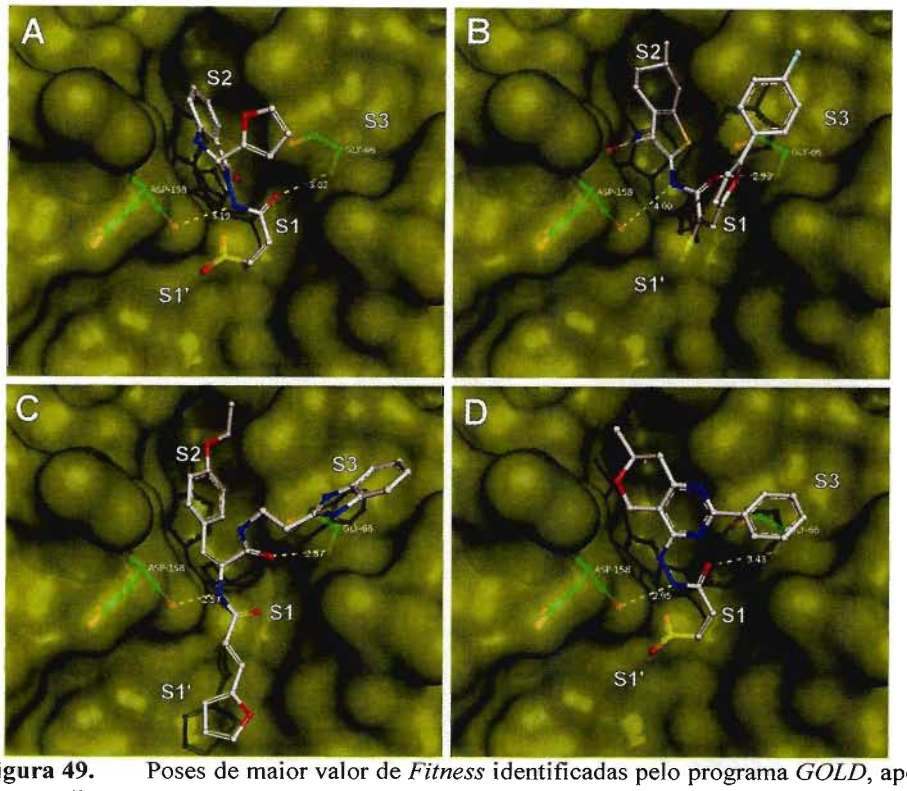


Figura 49. Poses de maior valor de *Fitness* identificadas pelo programa *GOLD*, após o procedimento de *docking* dos compostos **ZINC02338249** (**A**) **ZINC02342142** (**B**), **ZINC02399860** (**C**) e **ZINC02400672** (**D**) posicionados dentro da cavidade do sítio ativo da cruzaína. São indicados os valores das distâncias das ligações de hidrogênio entre os átomos dos ligantes e os resíduos Gly66 e Asp158 da cavidade do sítio ativo da cruzaína. As regiões assinaladas *S3*, *S2*, *S1*, *S1'* correspondem às respectivas cavidades do sítio ativo.

3.4.4. Filtro visual e seleção dos compostos para o teste de inibição enzimática

Conforme discutido na *Introdução* item **1.4.3.**, a inspeção visual dos resultados de *virtual screening* é uma etapa crítica no procedimento de *docking*.[Sotriffer *et al.*, 2004] O objetivo desta etapa é utilizar a intuição e os conhecimentos químicos para a seleção final dos compostos que, a seguir, serão comprados e testados experimentalmente como inibidores da cruzaína. Os 62 compostos selecionados na etapa anterior de *docking* foram visualmente inspecionados, dentro do sítio ativo da cruzaína, avaliando-se a qualidade da interação entre o composto e os resíduos que compõem o sítio de ligação, levando-se em conta os seguintes critérios: (*i*) a reatividade do grupo eletrofílico; (*ii*) a distância e a geometria da possível interação entre o enxofre da *Cys25* e o átomo eletrofílico do ligante; (*iii*) a complementaridade estrutural entre o ligante e as cavidades da enzima; (*iv*) a distância e a geometria das ligações de hidrogênio entre os átomos, de oxigênio da *Asp158*

e de nitrogênio da *Gly66* com os correspondentes átomos doadores e aceptores de ligação de hidrogênio do ligante. De acordo com estes critérios, 55 compostos foram selecionados como potenciais inibidores da cruzaína. Esta seleção por inspeção visual resultou em uma redução de 11,3% no número de moléculas selecionadas pelo filtro de *docking* (62 moléculas).

Dentre os compostos selecionados, foram identificados dois inibidores da cruzaína já descritos na literatura, ambos apresentando valores de inibição IC_{50} de 2,2 μM.[Harth *et al.*, 1993; Crawford *et al.*, 1988] Esta constatação é bastante importante, considerando-se que a confirmação de que estes dois compostos selecionados são comprovadamente inibidores da cruzaína representa uma validação externa do modelo de *virtual screening*. Ou seja, dentre os mais de 3 milhões de compostos do banco de dados *ZINC*, o modelo de *virtual screening* proposto nesta tese foi capaz de identificar estes inibidores da cruzaína. A **Figura 50** mostra uma representação estrutural dos dois inibidores de cisteíno-protease identificados pelo modelo de *virtual screening* de inibidores da cruzaína, respectivamente, ZINC04899534 e ZINC01547017. Uma vez que um dos objetivos desta tese de doutorado é a identificação de novos inibidores da cruzaína, estes dois compostos foram excluídos da lista de compostos para serem adquiridos e testados experimentalmente.

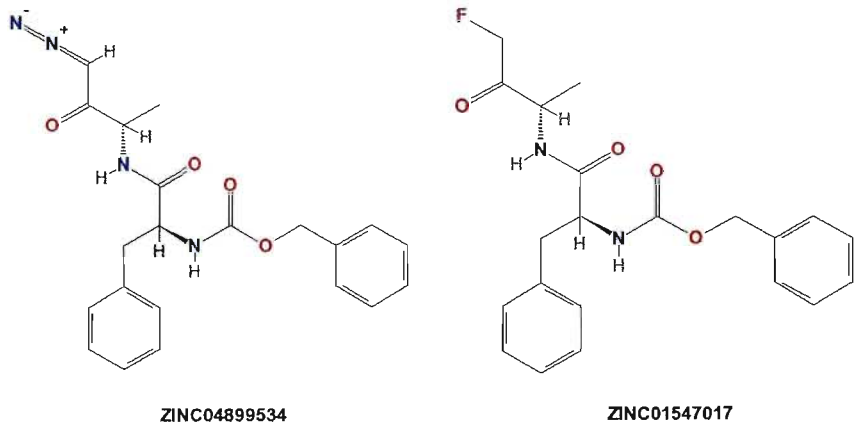


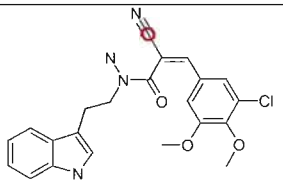
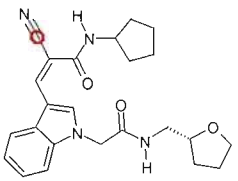
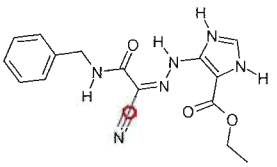
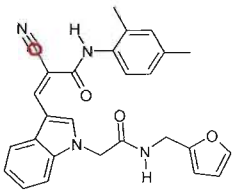
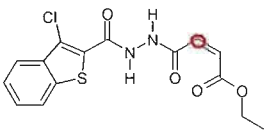
Figura 50. Representação estrutural esquemática dos dois inibidores de cisteíno-protease identificados pelo modelo de *virtual screening* de inibidores da cruzaína, respectivamente, ZINC04899534 e ZINC01547017 .

Dos outros 53 compostos identificados como possíveis inibidores da cruzaína apenas 19 estavam comercialmente disponíveis para pronta entrega. Estes 19 compostos foram adquiridos pelas correspondentes empresas fornecedoras. Esta etapa de aquisição dos

compostos representou uma redução de 65,5% em relação ao compostos selecionados pelo **Modelo II** (55 compostos).

A **Tabela 18** mostra uma representação estrutural dos 19 compostos selecionados pelo **Modelo II** de *virtual screening* e que foram adquiridos para serem submetidos ao teste de inibição da cruzaína. São eles os compostos: **ZINC00874200**, **ZINC00890529**, **ZINC01414485**, **ZINC01758885**, **ZINC01794384**, **ZINC01794422**, **ZINC01794844**, **ZINC01806509**, **ZINC02188675**, **ZINC02198693**, **ZINC02209285**, **ZINC02342142**, **ZINC02399860**, **ZINC02400672**, **ZINC02411240**, **ZINC02768688**, **ZINC04964356**, **ZINC03243155**, **ZINC03306238**.

Tabela 18. Representação estrutural esquemática dos 19 compostos selecionados pelo **Modelo II** de *virtual screening* e que foram adquiridos para serem submetidos ao teste de inibição da cruzaína. São apresentados os correspondentes códigos do banco de dados ZINC , os valores de peso molecular e a empresa fornecedora, de cada um dos compostos adquiridos. Os círculos em vermelho indicam o átomo de carbono eletrofílico susceptível ao ataque nucleofílico do enxofre catalítico da *Cys25*.

# Ref	Composto	Código *ZINC*	Massa Molecular	Vendedor
1		**ZINC00874200**	424,89	IBScreen (STOCKS2S-49269)
2		**ZINC00890529**	420,52	Asinex (BAS02985564)
3		**ZINC01414485**	342,36	Toslab (600063)
4		**ZINC01758885**	452,52	IBScreen (STOCK3S-57280)
5		**ZINC01794384**	352,80	Life Chemicals (F3145-4620)

# Ref	Composto	Código *ZINC*	Massa Molecular	Vendedor
6		**ZINC01794422**	393,20	Life Chemicals (F3145-4606)
7		**ZINC01794844**	320,35	Life Chemicals (F3145-3436)
8		**ZINC01806509**	392,50	Life Chemicals (F3047-0312
9		**ZINC02188675**	407,53	Specs (AG-205/37047002
10		**ZINC02198693**	432,87	IBScreen (STOCK2S-31029
11		**ZINC02209285**	403,49	IBScreen (STOCK2S-31935
12		**ZINC02342142**	449,51	IBScreen (STOCK2S-99288
13		**ZINC02399860**	484,56	IBScreen (STOCK4S-32328)
14		**ZINC02400672**	368,40	IBScreen (STOCK4S-33348)
15		**ZINC02411240**	385,44	IBScreen (STOCK4S-44363)

# Ref	Composto	Código *ZINC*	Massa Molecular	Vendedor
16		**ZINC02768688**	393,45	Vitas-M (STK008910)
17		**ZINC04964356**	410,48	ChemBridge (5230379)
18		**ZINC03243155**	408,50	Enamine (T0507-1260)
19		**ZINC03306238**	366,38	Enamine (T0515-7498)

3.4.5. Testes de inibição enzimática e validação experimental do modelo de *virtual screening*

Conforme discutido na *Introdução* desta tese, item **1.4.**, os modelos de *virtual screening* devem ser validados experimentalmente para a confirmação de sua utilidade na busca de compostos bioativos.

Assim, conforme descrito no item **2.8.** (*Materiais e Métodos*), cada um dos 19 compostos selecionados pelo **Modelo II** e que foram adquiridos das correspondentes empresas fornecedoras, foram testados como inibidores da cruzaína, de acordo com o teste de inibição de *tight-binding* modificado[Judice *et al.*, 2001], diretamente, na presença do detergente *Triton X-100*. Esta mudança, em relação ao **Modelo I**, no procedimento experimental do teste de inibição enzimática da cruzaína foi introduzida tendo em vista o grande número de compostos selecionados e testados pelo **Modelo I**, que apresentaram inibição enzimática através do mecanismo promíscuo. Sendo assim, considerando-se que um dos objetivos desta tese de doutorado é a identificação de inibidores específicos da cruzaína e não de inibidores que agem pelo mecanismo promíscuo, justifica-se a presença

de *Triton X-100* nos ensaios enzimáticos, para que a ocorrência deste tipo de artefato experimental seja minimizada.

A atividade da enzima, bem como as condições experimentais do teste enzimático foram validados utilizando-se como controle positivo de inibição da cruzaína, o composto ***E-64***, um inibidor específico de cisteíno-proteases.[Otto & Schirmeister, 1997]. A **Figura 51** apresenta o gráfico dos valores da velocidade de consumo do substrato pela cruzaína, respectivamente, sem inibidor e, na presença de diferentes concentrações do inibidor específico ***E-64***.

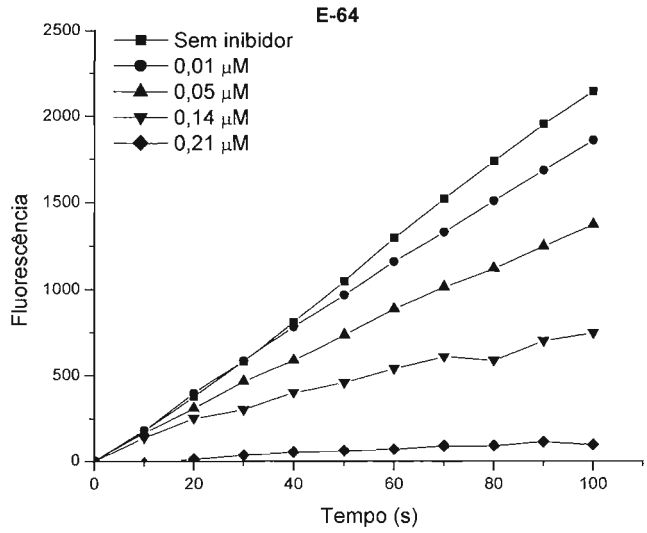


Figura 51. Gráficos dos valores da velocidade[a] de consumo do substrato[b] pela cruzaína[c] sem inibidor e na presença de três diferentes concentrações do composto ***E-64***.
[a] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).
[b] 5 μL de Bz-FR-MCA (170 μM).
[c] 0,3 μL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 μL de DTT

As **Figuras 52-70** apresentam os gráficos dos valores da velocidade de consumo do substrato pela cruzaína em diferentes concentrações, respectivamente, para cada um dos compostos selecionados do banco de dados *ZINC*, por *virtual screening*.

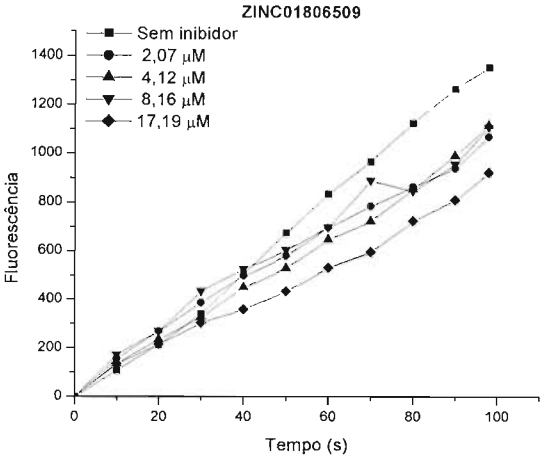


Figura 52. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC00890529** (*Asinex BAS02985564*).

[a] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 μL de Bz-FR-MCA (170 μM).

[c] 0,3 μL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 μL de DTT

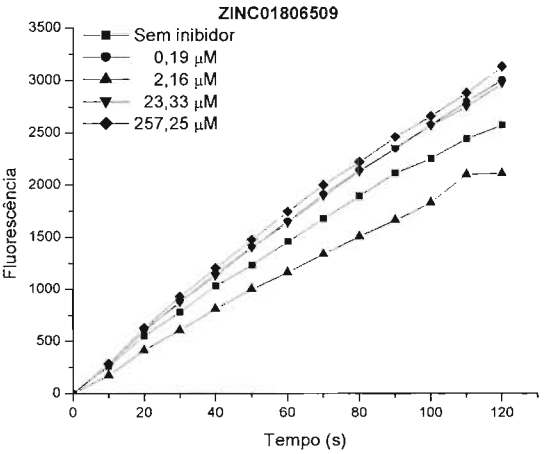


Figura 53. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC01806509** (*LifeChemicals F3047-0312*).

[a] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 μL de Bz-FR-MCA (170 μM).

[c] 0,3 μL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 μL de DTT

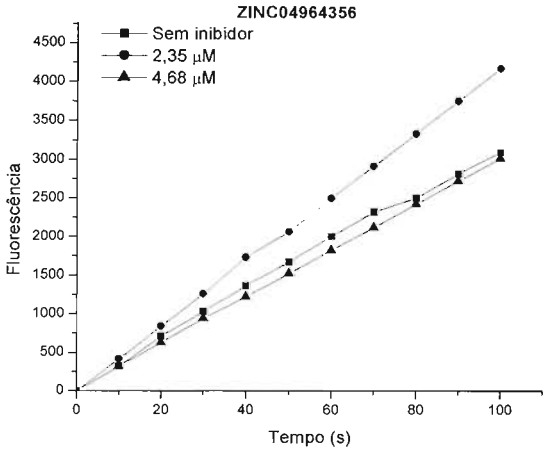


Figura 54. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC04964356** (*ChemBridge 5230379*).

[a] Valores determinados fluorimetricamente, em 1000 µL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 µL de Bz-FR-MCA (170 µM).

[c] 0,3 µL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 µL de DTT

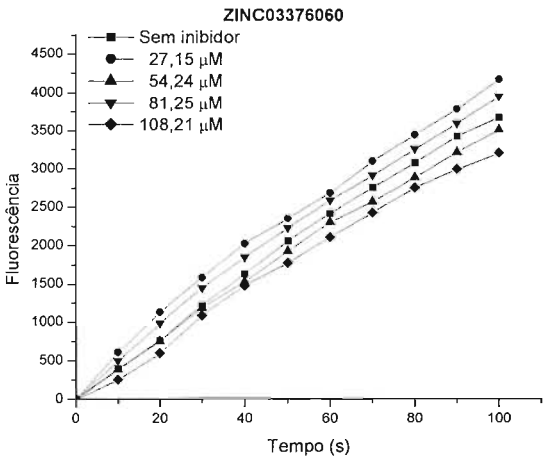


Figura 55. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto ZINC03376060 (*Enamine T0519-6812*).

[a] Valores determinados fluorimetricamente, em 1000 µL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 µL de Bz-FR-MCA (170 µM).

[c] 0,3 µL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 µL de DTT

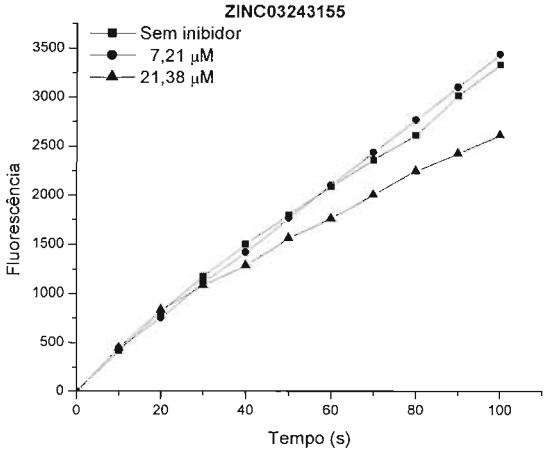


Figura 56. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC03243155** (*Enamine T0507-1260*).

[a] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 μL de Bz-FR-MCA (170 μM).

[c] 0,3 μL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 μL de DTT

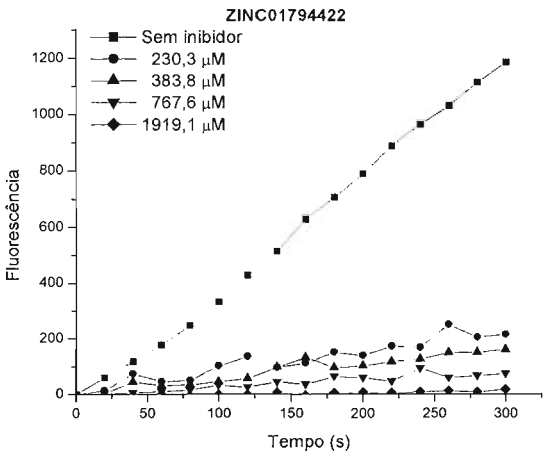


Figura 57. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC01794422** (*LifeChemicals F3145-4606*).

[a] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 μL de Bz-FR-MCA (170 μM).

[c] 0,3 μL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 μL de DTT

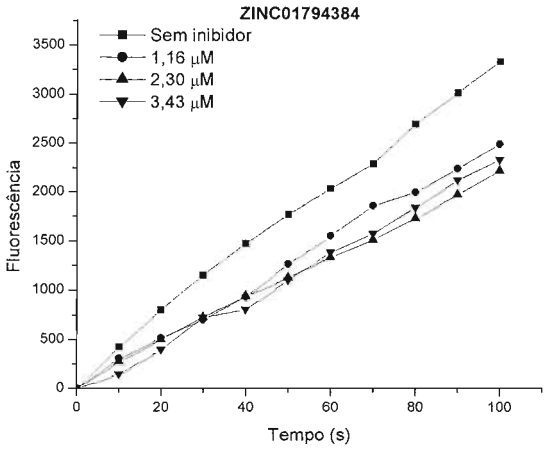


Figura 58. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC01794384** (*LifeChemicals F3145-4620*).

[a] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 μL de Bz-FR-MCA (170 μM).

[c] 0,3 μL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 μL de DTT

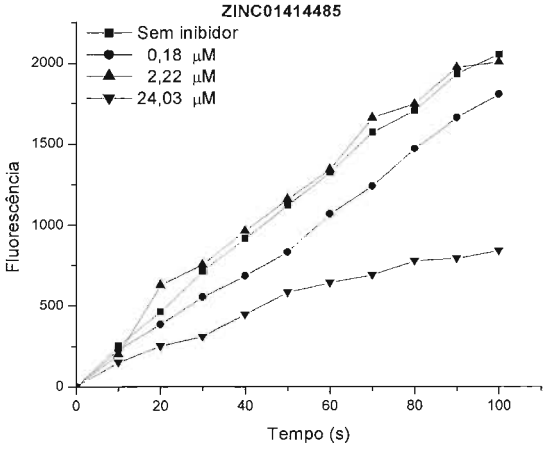


Figura 59. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC01414485** (*Toslab 60063*).

[a] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 μL de Bz-FR-MCA (170 μM).

[c] 0,3 μL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 μL de DTT

BIBLIOTECA
INSTITUTO DE QUÍMICA
Universidade de São Paulo

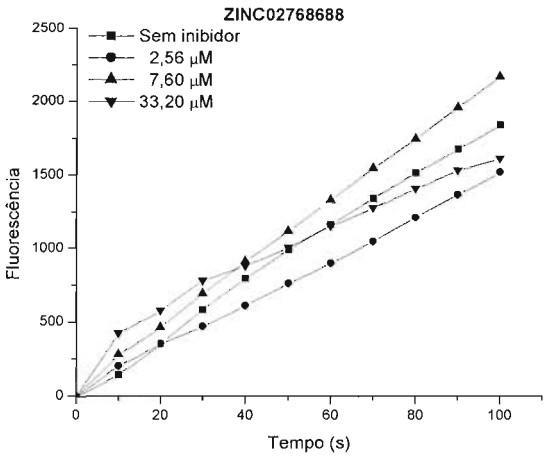


Figura 60. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC02768688** (*Vitas-M STK008910*).

[a] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 μL de Bz-FR-MCA (170 μM).

[c] 0,3 μL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 μL de DTT

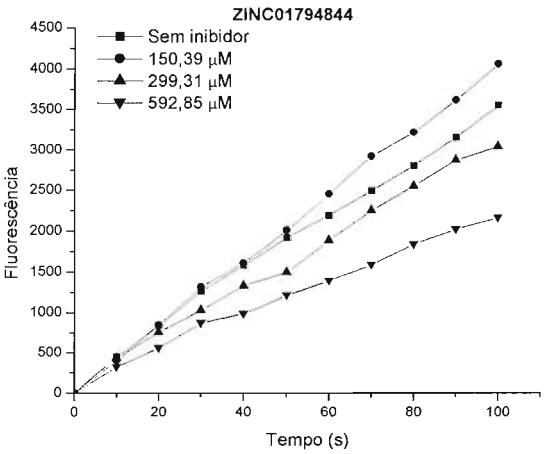


Figura 61. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC01794844** (*LifeChemicals F3145-3436*).

[a] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 μL de Bz-FR-MCA (170 μM).

[c] 0,3 μL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 μL de DTT

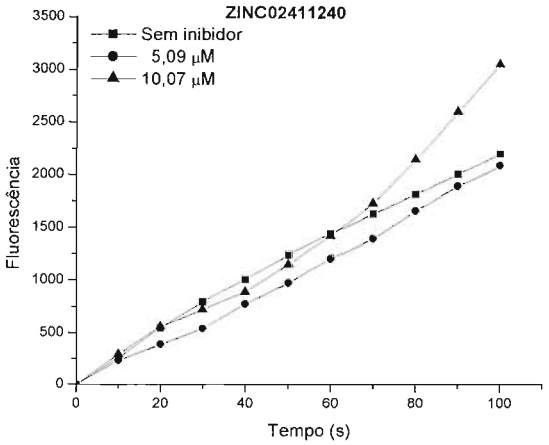


Figura 62. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC02411240** (*IBScreen, STOCK4S-44363*).

[a] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 μL de Bz-FR-MCA (170 μM).

[c] 0,3 μL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 μL de DTT

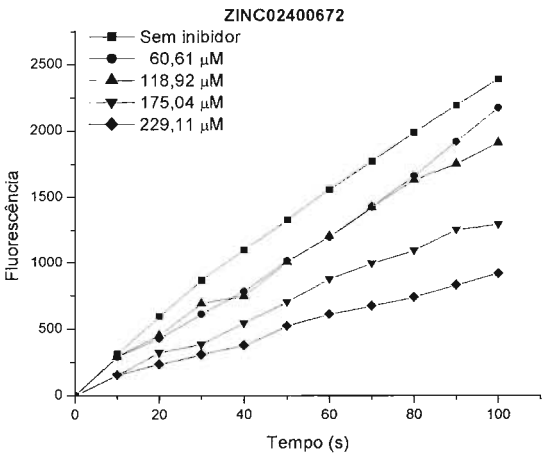


Figura 63. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC02400672** (*IBScreen, STOCK4S-33348*).

[a] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 μL de Bz-FR-MCA (170 μM).

[c] 0,3 μL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 μL de DTT

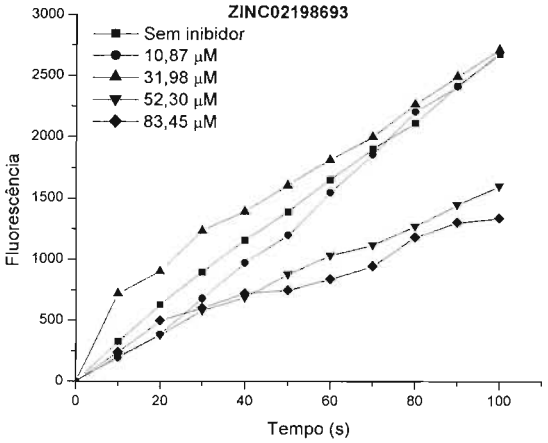


Figura 64. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC02198693** (*IBScreen, STOCK2S-31029*).

[a] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 μL de Bz-FR-MCA (170 μM).

[c] 0,3 μL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 μL de DTT

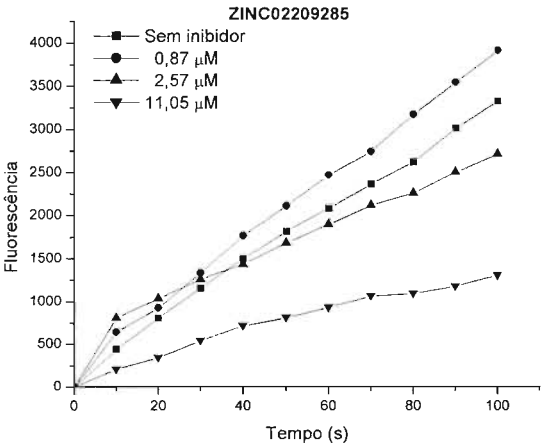


Figura 65. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC02209285** (*IBScreen, STOCK2S-31935*).

[a] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 μL de Bz-FR-MCA (170 μM).

[c] 0,3 μL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 μL de DTT

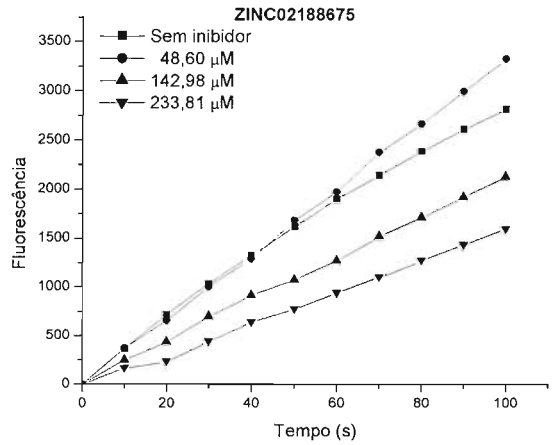


Figura 66. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC02188675** (*Specs, AG-205/37047002*).

[a] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 μL de Bz-FR-MCA (170 μM).

[c] 0,3 μL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 μL de DTT

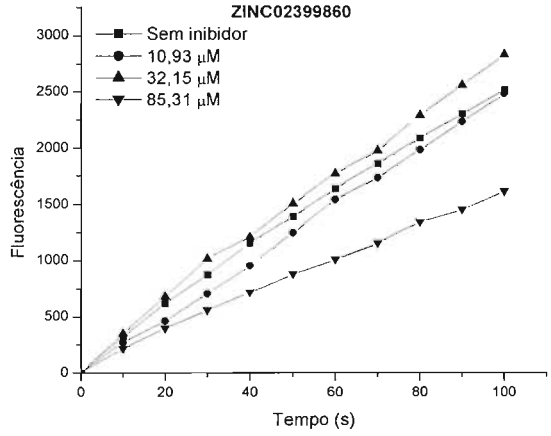


Figura 67. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC02399860** (*IBScreen, STOCK4S-32328*).

[a] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 μL de Bz-FR-MCA (170 μM).

[c] 0,3 μL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 μL de DTT

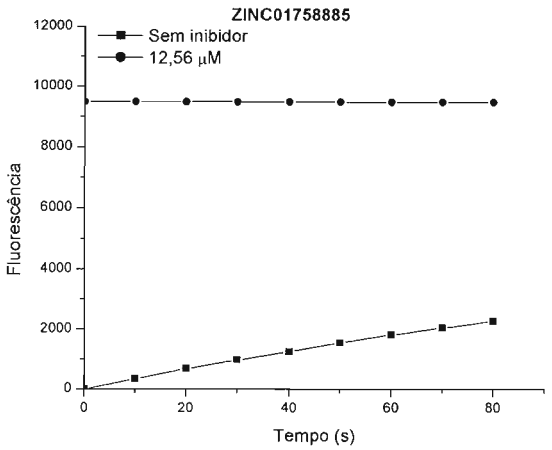


Figura 68. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC01758885** (*IBScreen, STOCK3S-57280*).

[a] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 μL de Bz-FR-MCA (170 μM).

[c] 0,3 μL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 μL de DTT

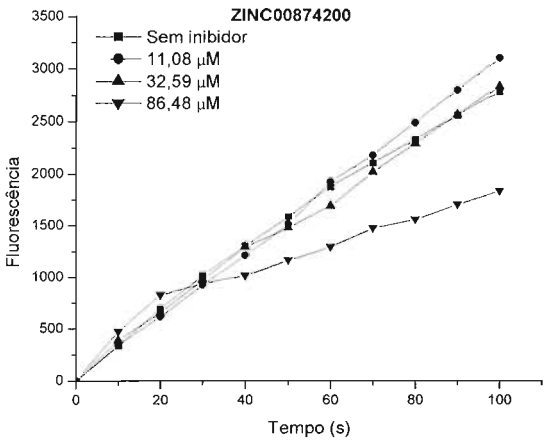


Figura 69. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC00874200** (*IBScreen, STOCKS2S-49269*).

[a] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 μL de Bz-FR-MCA (170 μM).

[c] 0,3 μL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 μL de DTT

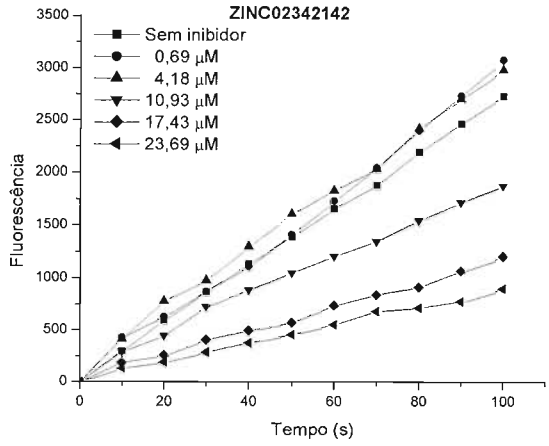


Figura 70. Gráfico dos valores da velocidade [a] de consumo do substrato [b] pela cruzaína [c] em diferentes concentrações do composto **ZINC02342142** (*IBScreen, STOCK2S-99288*).

[a] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[b] 5 μL de Bz-FR-MCA (170 μM).

[c] 0,3 μL de cruzaína (0,18 nM) com 5 min de pré-incubação na presença de 1 μL de DTT

Dentre 19 os compostos selecionados pelo modelo de *virtual screening*, adquiridos e testados experimentalmente como inibidores da cruzaína, apenas um apresentou inibição significativa, dentro do erro experimental, enquanto, que os demais compostos não foram capazes de inibir a enzima até concentrações em que foi observada a perda de solubilidade dos compostos na solução tampão, ou seja, de 4 μM a 592 μM. O composto **ZINC01794422** (*LifeChemicals F3145-4606*) (**Figura 71**) inibiu a cruzaína apresentando uma constante de inibição no valor de K_i = 21 μM. Este valor foi obtido a partir dos valores de velocidade de consumo do substrato em função da concentração do inibidor, ajustados pela equação de *tight binding*.[Gillmor *et al.*, 1997] A **Figura 72** apresenta gráfico do ajuste dos parâmetros da equação de *tight-binding* aos valores de velocidade de consumo do substrato pela cruzaína em função da concentração do composto **ZINC01794422** (*LifeChemicals F3145-4606*).

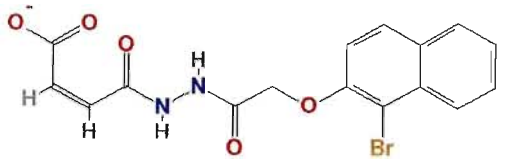


Figura 71. Representação estrutural do composto **ZINC01794422** (*LifeChemicals F3145-4606*) identificado como inibidor da cruzaína, através **Modelo II** de *virtual screening* aplicado ao banco de *ZINC*

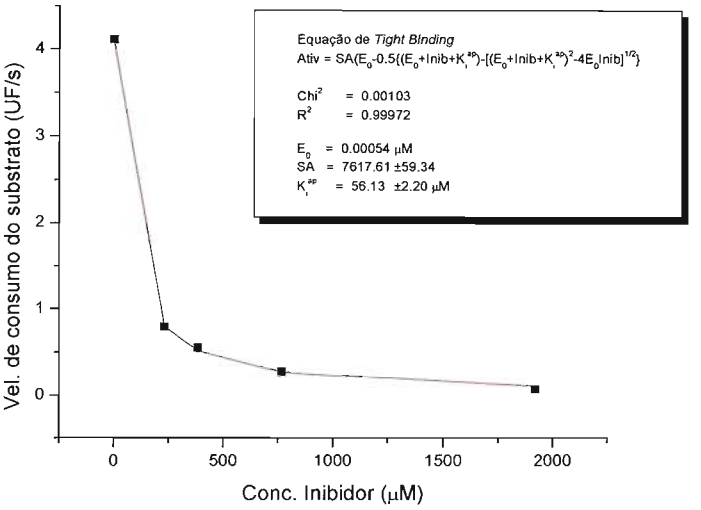


Figura 72. Gráfico do ajuste[a] (—) dos parâmetros da equação de *tight-binding*[b] aos valores (■) de velocidade[c] de consumo do substrato[d] pela cruzaína[e] em função da concentração do composto **ZINC01794422** (*LifeChemicals F3145-4606*).

[a.] Utilizando o programa Origin6.0 [Originlab®, 2002].

[b.] Equação 2 (item 2.6., *Materiais e Métodos*, página 87)

[c] Valores determinados fluorimetricamente, em 1000 μL de solução tampão fosfato dissódico (50 mM) e EDTA (2,5 mM), pH=6,5 e 1% (v/v) de *Triton X-100, sob agitação* magnética e temperatura controlada e ajustada para (37,0 ± 0,1 °C) em um Fluorímetro HITACHI4500 (Ex: 380 nm; Em: 460 nm).

[d.] 5 μL de Bz-FR-MCA (170 μM).

[e.] 0,3 μL de cruzaína (0,18nM) com 5 min de pré-incubação na presença de 1 μL de DTT

A partir do resultado experimental de inibição observado e considerando-se que o composto **ZINC01794422** (*LifeChemicals F3145-4606*) é um inibidor não-peptídico, contendo o grupamento ácido-carboxílico α,β-insaturado, pode-se propor um mecanismo de inibição envolvendo o ataque nucleofílico da *Cys25* na dupla ligação ativada pelo grupamento retirador de elétrons carboxilato e a conseqüente alquilação irreversível da enzima. Este mecanismo é, ainda, uma hipótese, e deve, portanto, ser comprovado por

outros experimentos, incluindo-se a formação e determinação da estrutura do complexo por cristalografia de raio-X. A inibição de cisteíno-proteases por sistemas carboxílicos α,β-insaturados já foi descrita na literatura,[Barrett *et al.*, 1982] entretanto, estes grupos funcionais sempre foram associados a inibidores peptídicos.

De acordo com os resultados do procedimento de *docking* do composto **ZINC01794422** (*LifeChemicals F3145-4606*) sobre a cavidade do sítio ativo da cruzaína, o anel naftalênico do composto se posiciona na cavidade *S2* da enzima, enquanto que os átomo de nitrogênio da hidrazida e o átomo de oxigênio carbonílico estabelecem ligações de hidrogênio com os resíduos *Gly66* e *Asp158* da cruzaína. A porção ácido-carboxílico terminal se posiciona na cavidade oxiânion da cruzaína e deve contribuir para a solubilidade (e provavelmente, a biodisponibilidade) do composto. A **Figura 73** mostra a pose do composto **ZINC01794422** (*LifeChemicals F3145-4606*), conforme docada pelo programa *GOLD*, no sítio ativo da cruzaína. São indicados os valores das distâncias das ligações de hidrogênio entre os átomos dos ligantes e os resíduos *Gly66* e *Asp158* da cavidade do sítio ativo da cruzaína.

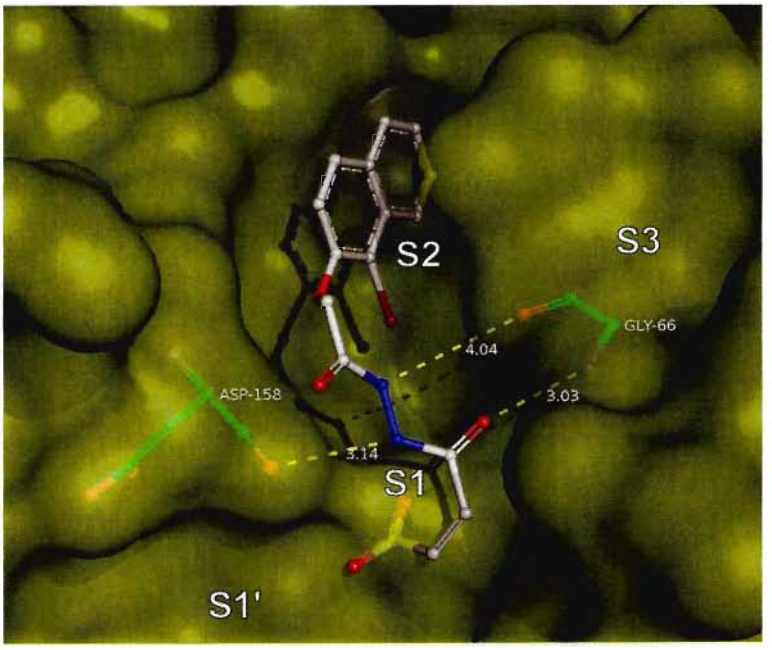


Figura 73. Pose de maior valor de *Fitness* identificadas pelo programa *GOLD*, após o procedimento de *docking* do composto **ZINC01794422** (*LifeChemicals F3145-4606*) posicionado dentro da cavidade do sítio ativo da cruzaína. São indicados os valores das distâncias das ligações de hidrogênio entre os átomos dos ligantes e os resíduos Gly66 e Asp158 da cavidade do sítio ativo da cruzaína. As regiões assinaladas *S3*, *S2*, *S1*, *S1'* correspondem às respectivas cavidades do sítio ativo.

A descoberta do composto **ZINC01794422** (*LifeChemicals F3145-4606*) como um inibidor da cruzaína (K_i = 21μM) é um resultado positivo que demonstra o sucesso do procedimento de *virtual screening*, proposto nesta tese de doutorado, na busca de ligantes específicos partindo-se de informações tanto da enzima alvo quanto de inibidores conhecidos. Considerando-se a proporção do número de compostos testados em relação ao número de compostos ativos identificados, este modelo de *virtual screening* encontrou 1 em 19, ou seja, uma taxa de acerto de 5,3%. Este valor está de acordo com as taxas de acerto encontradas na literatura que variam entre 1 a 10% dos compostos sugeridos pelos modelos de *virtual screening*.[Rester, 2006]

Como perspectivas futuras que podem ser levantadas a partir das abordagens utilizadas e dos resultados obtidos nesta tese de doutorado, deve-se ressaltar que, o composto identificado como inibidor da cruzaína é apenas um ligante micromolar (21μM) da enzima e que pode ser usado como ponto de partida para o desenvolvimento de uma molécula bioativa, com possível atividade na terapia da doença de Chagas. Assim, para que este *Hit* (achado) se torne um *Lead* (estrutura-líder), outros estudos se fazem necessários, como por exemplo:

1. Propor modelos de QSAR de uma série de análogos para obter uma otimizar a estrutura líder identificada, quanto às suas propriedades de potência, biodisponibilidade e toxicidade. Nesta série de análogos os substituintes devem contemplar as seguintes exigências estruturais: (*i*) ocupar a cavidade S3, e/ou a cavidade S1`, (*ii*) aumentar o caráter eletrofílico do grupamento que sofre o ataque da *Cys25* cisteína catalítica.
2. Verificar a importância do átomo de bromo para a interação do ligante com a enzima e considerar a sua potencial toxicidade.
3. Avaliar a estabilidade do complexo inibidor-cruzaína através de estudos de dinâmica molecular.
4. Validação, por cristalografia de raio-X, do modo de ligação do composto ativo sugerido no procedimento de *docking*.
5. Avaliar a especificidade do composto frente a outras cisteíno-protease.
6. Realizar os testes de inibição *in vivo* do composto identificado

4. Conclusões

A aplicação de dois modelos de *virtual screening* (**Modelos I** e **II**) de inibidores da cruzaína, uma cisteíno-protease do *Trypanosoma cruzi,* o agente etiológico da doença de Chagas, sobre a biblioteca de compostos orgânicos de baixa massa molecular *ZINC*, resultou na identificação de um inibidor específico da cruzaína (**ZINC01794422**, K_i =21 μM) e na identificação de três inibidores inespecíficos da cruzaína, agindo pelo mecanismo promíscuo de inibição enzimática (**ZINC02663001**, **ZINC01936854**, **ZINC03326243**).

O reconhecimento detalhado das características estruturais da cruzaína e de seu sítio de ligação - após o alinhamento, átomo-a-átomo, dos 13 complexos da cruzaína depositados no *PDB* - realizado através dos seguintes procedimentos: inspeção visual; análise dos campos de interação molecular gerados pelo programa *GRID*;[Goodford, 1985] identificação das propriedades de interação molecular na superfície da cavidade, geradas pelo programa *CAVBASE*[Sotriffer & Klebe, 2002] e simulações de dinâmica molecular; permitiu as seguintes constatações:

1. Os treze complexos da cruzaína depositados no banco de dados *PDB* (resolução de 1,2 Å à 2,4 Å) são estruturalmente semelhantes em relação às coordenadas dos átomos da enzima (**Tabela 7**).
2. Na estrutura cristalina dos treze complexos da cruzaína analisados, não foram observadas moléculas de água conservadas, nem moléculas de água intermediando as ligações entre cada inibidor e as correspondentes estruturas da cruzaína (item 3.2.2.).
3. Regiões favoráveis para posicionar ligantes que apresentam, respectivamente, interações hidrofóbicas e ligações de hidrogênio doadora e aceptora com os correspondentes resíduos do sítio de ligação da cruzaína (**Figura 31**) foram reveladas através dos campos de interação molecular, calculados pelo programa *GRID*. O reconhecimento destes pontos de interação favorável contribui para a construção do modelo farmacofórico de inibidores da cruzaína.
4. Vinte complexos *PDB* com cavidades similares ao sítio de ligação da cruzaína (**Tabela 8**) foram identificados a partir das propriedades de reconhecimento molecular (pseudocentros), expostas na superfície da cavidade da cruzaína

utilizando-se o programa *CAVBASE* (**Figuras 32** e **33**). Todas as estruturas selecionadas eram de cisteíno-proteases, diferentes da cruzaína e, apresentavam inibidores covalentemente ligados ao enxofre da cisteína catalítica.

5. A cruzaína apresenta significativa estabilidade de sua estrutura terciária e enovelamento globular compacto verificada através de estudos de simulação de dinâmica molecular (**Tabela 9**). Verificou-se que o sítio ativo da enzima, com ou sem o ligante, não sofre grandes mudanças conformacionais. Entretanto, a flexibilidade observada nas cadeias laterais de alguns resíduos do sítio ativo sugere a ocorrência de ajuste induzido na conformação da enzima para a acomodação do ligante (**Tabela 10**).

Ao se aplicar o **Modelo I** de *virtual screening* de inibidores da cruzaína - obtido a partir do reconhecimento das estruturas dos 13 complexos da cruzaína depositadas no *PDB* - sobre o banco de dados *ZINC* (3.294.714 compostos), constatamos que:

6. Seis compostos (6/3.294.714), sugeridos pelo **Modelo I** como potenciais inibidores da cruzaína (**Tabela 13**), foram adquiridos e submetidos ao teste de inibição enzimática da cruzaína, de acordo com o método de *tight-binding* modificado (item 2.6.), para validação experimental do modelo. Observou-se que três compostos (**ZINC02470662**, **ZINC02682879** e **ZINC03192044,** respectivamente) não mostraram inibição significativa da cruzaína, nas condições experimentais utilizadas, até a concentração de 670μM (**Figura 42**), enquanto que os três restantes (compostos **ZINC02663001**, **ZINC01936854** e **ZINC03326243**, respectivamente) apresentaram inibição enzimática inespecífica, sugerindo que estes últimos agem segundo o mecanismo promíscuo. (**Figura 42**)
7. O mecanismo promíscuo de inibição enzimática dos compostos **ZINC02663001**, **ZINC01936854** e **ZINC03326243**, respectivamente, foi verificado pela adição de *Triton X-100* (0,1%, v/v) no ensaio enzimático, observando-se a correspondente perda de atividade de inibição da cruzaína. (**Figura 42**). Para cada composto, a confirmação deste mecanismo foi realizada através de teste diferente, verificando-se o efeito do aumento da concentração da enzima em dez vezes sobre a inibição.

Observou-se a perda de atividade de inibição da cruzaína, para os três compostos. (**Figura 43**)

8. Apesar do **Modelo I** de *virtual screening* ter sido desenvolvido para selecionar inibidores específicos da cruzaína, os seis compostos selecionados e testados se mostraram inativos frente à inibição da cruzaína (compostos **ZINC02663001**, **ZINC01936854** e **ZINC03326243**, respectivamente), ou promoveram resultados falso-positivos nos testes enzimáticos *in vitro* (compostos **ZINC02663001**, **ZINC01936854** e **ZINC03326243**, respectivamente).

A análise crítica dos resultados obtidos na preparação e na aplicação do **Modelo I**, sobre o banco de dados *ZINC*, motivou as mudanças introduzidas na obtenção e na validação de um outro modelo (**Modelo II**) de *virtual screening* de inibidores da cruzaína. Ao se aplicar o **Modelo II** de *virtual screening* de inibidores da cruzaína - obtido a partir do reconhecimento das estruturas de 33 complexos de cisteíno-proteases depositadas no *PDB*, a saber, os 13 complexos da cruzaína e os 20 complexos de outras cisteíno-proteases, identificadas na busca por cavidades similares à cruzaína - sobre o banco de dados *ZINC* (3.294.714 compostos), constatamos que:

9. 19 compostos (19/3.294.714), sugeridos pelo **Modelo II** como potenciais inibidores da cruzaína (**Tabela 18**), foram adquiridos e submetidos ao teste de inibição enzimática da cruzaína, de acordo com o método de *tight-binding* modificado (item 2.8.), para validação experimental do modelo. Observou-se que um composto (**ZINC01794422 – Figura 71**) apresentou inibição específica da enzima com constante de inibição no valor de K_i = 21 μM (**Figura 57**), enquanto os demais dezoito compostos não mostraram inibição significativa, nas condições experimentais utilizadas, até a concentração de 592 μM (**Figuras 52-56** e **58-70**).
10. Para os 19 compostos selecionados pelo **Modelo II** *de virtual screening*, o mecanismo promíscuo de inibição enzimática não foi observado, uma vez que todos os testes foram realizados na presença de 0,1% de *Triton X-100*.
11. O **Modelo II** de *virtual screening*, identificou no banco de dados *ZINC* dois compostos (**ZINC04899534** e **ZINC01547017**; **Figura 50**) descritos na literatura como inibidores da cruzaína (IC_{50} de 2,2μM.[Harth *et al.*, 1993; Crawford *et al.*,

1988]), embora estes não tenham sido incluídos na construção do modelo. Estes compostos não foram adquiridos.

12. Considerando-se apenas o novo inibidor identificado, o **Modelo II** de *virtual screening*, encontrou 1 composto ativo em 19 testados, ou seja, uma taxa de acerto de 5,3%. Este valor esta de acordo com as taxas de acerto encontradas na literatura que variam entre 1 a 50% .[Rester, 2006]

5. Referencias

[Anon], Nomenclature And Symbolism For Amino-Acids And Peptides, *Pure and Applied Chemistry*, 56, 595-624, **1984**.

Afzelius, L., Masimirembwa, C.M., Karlen, A., Andersson, T.B., Zamora, I., Discriminant and quantitative PLS analysis of competitive CYP2C9 inhibitors versus non-inhibitors using alignment independent GRIND descriptors, *Journal of Computer-Aided Molecular Design*, 16, 443-458, **2002**.

Ajay, Murcko, M.A., Computational methods to predict binding free energy in ligand-receptor complexes, *Journal of Medicinal Chemistry*, 38, 4953-4967, **1995**.

Ajay, Walters, W.P., Murcko, M.A., Can we learn to distinguish between "drug-like" and "nondrug-like" molecules?, *Journal of Medicinal Chemistry*, 41, 3314-3324, **1998**.

Alves, L.C., Melo, R.L., Cezari, M.H.S., Sanderson, S.J., Mottram, J.C., Coombs, G.H., Juliano, L., Juliano, M.A., Analysis of the S-2 subsite specificities of the recombinant cysteine proteinases CPB of Leishmania mexicana, and cruzain of Trypanosoma cruzi, using fluorescent substrates containing non-natural basic amino acids, *Molecular and Biochemical Parasitology*, 117, 137-143, **2001a**.

Alves, L.C., Melo, R.L., Sanderson, S.J., Mottram, J.C., Coombs, G.H., Caliendo, G., Santagada, V., Juliano, L., Juliano, M.A., S-1 subsite specificity of a recombinant cysteine proteinase, CPB, of Leishmania mexicana compared with cruzain, human cathepsin L and papain using substrates containing non-natural basic amino acids, *European Journal of Biochemistry*, 268, 1206-1212, **2001b**.

Andrews, P.R., Craik, D.J., Martin, J.L., Functional-group contributions to drug receptor interactions, *Journal of Medicinal Chemistry*, 27, 1648-1657, **1984**.

Asboth, B., polgar, L., Transition-state stabilization at the oxyanion binding-sites of serine and thiol proteinases - hydrolyses of thiono and oxygen esters, *Biochemistry*, 22, 117-122, **1983**.

Asboth, B., Stokum, E., Khan, I.U., Polgar, L., Mechanism of action of cysteine proteinases - oxyanion binding- site is not essential in the hydrolysis of specific substrates, *Biochemistry*, 24, 606-609, **1985**.

Ashall, F., Angliker, H., Shaw, E., Lysis of Trypanosomes by Peptidyl fluoromethyl Ketones, *Biochemical and Biophysical Research Communications*, 170, 923-929, **1990**.

Babine, R.E., Bender, S.L., Molecular recognition of protein-ligand complexes: Applications to drug design, *Chemical Reviews* , 97, 1359-1472, **1997**.

Baroni, R., Aspectos estruturais, mecanismo de formaçào e atividade antimicrobiana de benzilidenotiossemocarbazonas 2- ou 4-, potencialmente quimioterápicas, Tese de Mestrado, FCF-USP, **1987**.

Barrett, A.J., Kembhavi, A.A., Brown, M.A., Kirschke, H., Knight, C.G., Tamai, M., Hanada, K., l-Trans-epoxysuccinyl-leucylamido(4-guanidino)butane (e-64) and its analogs as inhibitors of cysteine proteinases including cathepsins b, h and L, *Biochemical Journal*, 201, 189-198, **1982**.

Berendsen, H.J.C., Postma, J.P.M., Vangunsteren, W.F., Dinola, A., Haak, J.R., Molecular-Dynamics with Coupling to an External Bath, *Journal of Chemical Physics*, 81, 3684-3690, **1984**.

Berendsen, H.J.C., Vanderspoel, D., Vandrunen, R., GROMACS - A message-passing parallel molecular-dynamics implementation, *Computer Physics Communications*, 91, 43-56, **1995**.

Bhattacharjee, A.K., Geyer, J.A., Woodard, C.L., Kathcart, A.K., Nichols, D.A., Prigge, S.T., Li, Z.Y., Mott, B.T., Waters, N.C., A three-dimensional in silico pharmacophore model for inhibition of Plasmodium falciparum cyclin-dependent kinases and discovery of different classes of novel Pfmrk specific inhibitors, *Journal of Medicinal Chemistry*, 47, 5418-5426, **2004**.

Biobyte Corp., CLOGP, **1995**

Bissantz, C., Folkers, G., Rognan, D., Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations, *Journal of Medicinal Chemistry*, 43, 4759-4767, **2000**.

Bleicher, K.H., Bohm, H.J., Muller, K., Alanine, A.I., Hit and lead generation: Beyond high-throughput screening, *Nature Reviews Drug Discovery*, 2, 369-378, **2003**.

Bleicher, K.H., Wuthrich, Y., Adam, G., Hoffmann, T., Sleight, A.J., Parallel solution- and solid-phase synthesis of spiropyrrolo-pyrroles as novel neurokinin receptor ligands, *Bioorganic & Medicinal Chemistry Letters*, 12, 3073-3076, **2002a**.

Bleicher, K.H., Wuthrich, Y., De Boni, M., Kolczewski, S., Hoffmann, T., Sleight, A.J., Parallel solution- and solid-phase synthesis of spirohydantoin derivatives as neurokinin-1 receptor ligands, *Bioorganic & Medicinal Chemistry Letters*, 12, 2519-2522, **2002b**.

Bohm, H. J., Prediction of Non-bonded Interactions in Drug Design, in *Protein-Ligand Interactions*, edited by H. J. Bohm and G. Schneider, pp. 3-20, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim,**2003**.

Bohm, H.J., Klebe, G., What can we learn from molecular recognition in protein-ligand complexes for the design of new drugs?, *Angewandte Chemie-International Edition In English*, 35, 2589-2614, **1996**.

Bonaldo, M.C., Descoffier, L.N., Salles, J.M., Goldenberg, S., Characterization and expression of proteases during trypanosoma-cruzi metacyclogenesis, *Experimental Parasitology*, 73, 44-51, **1991**.

Bontempi, E., Decazzulo, B.M., Ruiz, A.M., Cazzulo, J.J., Purification and some properties of an acidic protease from epimastigotes of trypanosoma-cruzi, *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology*, 77, 599-604, **1984**.

Boobbyer, D.N.A., Goodford, P.J., Mcwhinnie, P.M., Wade, R.C., New hydrogen-bond potentials for use in determining energetically favorable binding-sites on molecules of known structure, *Journal of Medicinal Chemistry*, 32, 1083-1094, **1989**.

Borman, S., Combinatorial Chemistry, *Chemical & Engineering News*, 76, 47-67, **2006**.

Brinen, L.S., Hansell, E., Cheng, J.M., Roush, W.R., McKerrow, J.H., Fletterick, R.J., A target within the target: probing Cruzain's P1 ' site to define structural determinants for the Chagas' disease protease, *Structure with Folding & Design*, 8, 831-840, **2000**.

Caffrey, C.R., Schanz, M., Nkemgu-Njinkeng, J., Brush, M., Hansell, E., Cohen, F.E., Flaherty, T.M., McKerrow, J.H., Steverding, D., Screening of acyl hydrazide proteinase inhibitors for antiparasitic activity against Trypanosoma brucei, *International Journal of Antimicrobial Agents*, 19, 227-231, **2002**.

Carlson, H.A., Masukawa, K.M., Rubins, K., Bushman, F.D., Jorgensen, W.L., Lins, R.D., Briggs, J.M., Mccammon, J.A., Developing a dynamic pharmacophore model for HIV-1 integrase, *Journal of Medicinal Chemistry*, 43, 2100-2114, **2000**.

Catalyst, © Accelrys Software Inc. San Diego, CA, USA, **2005**

Cazzulo, J.J., Cruzipain, major cysteine proteinase of Trypanosoma cruzi: sequence and genomic organization of the codifying genes, *Medicina-Buenos Aires*, 59, 7-10, **1999**.

Cazzulo, J.J., Cazzulo Franke, M.C., Martinez, J., Franke de Cazzulo, B.M., Some kinetic properties of a cysteine proteinase (cruzipain) from Trypanosoma cruzi, *Biochimica et Biophysica Acta*, 1037, 186-191, **1990**.

Cazzulo, J.J., Stoka, V., Turk, V., Cruzipain, the major cysteine proteinase from the protozoan parasite Trypanosoma cruzi, *Biological Chemistry*, 378, 1-10, **1997**.

Cazzulo, J.J., Stoka, V., Turk, V., The major cysteine proteinase of Trypanosoma cruzi: A valid target for chemotherapy of Chagas disease, *Current Pharmaceutical Design*, 7, 1143-1156, **2001**.

CCDC, Copyright © 2004 The Cambridge Crystallographic Data Centre, **2006**

Charifson, P.S., Corkery, J.J., Murcko, M.A., Walters, W.P., Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins, *Journal of Medicinal Chemistry*, 42, 5100-5109, **1999**.

Charifson, P.S., Walters, W.P., Filtering databases and chemical libraries, *Journal of Computer-Aided Molecular Design*, 16, 311-323, **2002**.

Chibale, K., Musonda, C.C., The synthesis of parasitic cysteine protease and trypanothione reductase inhibitors, *Current Medicinal Chemistry*, 10, 1863-1889, **2003**.

Chiyanzu, I., Hansell, E., Gut, J., Rosenthal, P.J., McKerrow, J.H., Chibale, K., Synthesis and evaluation of isatins and thiosemicarbazone derivatives against cruzain, falcipain-2 and rhodesain, *Bioorganic & Medicinal Chemistry Letters*, 13, 3527-3530, **2003**.

Choe, Y., Brinen, L.S., Price, M.S., Engel, J.C., Lange, M., Grisostomi, C., Weston, S.G., Pallai, P.V., Cheng, H., Hardy, L.W., Development of [alpha]-keto-based inhibitors of cruzain, a cysteine protease implicated in Chagas disease, *Bioorganic & Medicinal Chemistry*, 13, 2141-2156, **2005**.

Clark, R.D., Strizhev, A., Leonard, J.M., Blake, J.F., Matthew, J.B., Consensus scoring for ligand/protein interactions, *Journal of Molecular Graphics & Modelling*, 20, 281-295, **2002**.

Colli, W., Alves, M.J.M., Relevant glycoconjugates on the Surface of Trypanosoma cruzi, *Memorias do Instituto Oswaldo Cruz*, 94, 37-49, **1999**.

Coura, J.R., de Castro, S.L., A critical review on Chagas disease chemotherapy, *Memorias do Instituto Oswaldo Cruz*, 97, 3-24, **2002**.

Cramer, R.D., Patterson, D.E., Bunce, J.D., Comparative Molecular-Field Analysis (CoMFA) .1. Effect of shape on binding of steroids to carrier proteins, *Journal of the American Chemical Society*, 110, 5959-5967, **1988**.

Crawford, C., Mason, R.W., Wikstrom, P., Shaw, E., The design of peptidyldiazomethane inhibitors to distinguish between the cysteine proteinases calpain-ii, cathepsin-l and cathepsin-b, *Biochemical Journal*, 253, 751-758, **1988**.

Davis, A.M., Teague, S.J., Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis, *Angewandte Chemie-International Edition*, 38, 737-749, **1999**.

Decazzulo, B.M.F., Martinez, J., north, M.J., Coombs, G.H., Cazzulo, J.J., Effects of proteinase-inhibitors on the growth and differentiation of trypanosoma-cruzi, *Fems Microbiology Letters*, 124, 81-86, **1994**.

DelNery, E., Juliano, M.A., Lima, A.P.C.A., Scharfstein, J., Juliano, L., Kininogenase activity by the major cysteinyl proteinase (Cruzipain) from Trypanosoma cruzi, *Journal of Biological Chemistry*, 272, 25713-25718, **1997a**.

DelNery, E., Juliano, M.A., Meldal, M., Svendsen, I., Scharfstein, J., Walmsley, A., Juliano, L., Characterization of the substrate specificity of the major cysteine protease (cruzipain) from Trypanosoma cruzi using a portion-mixing combinatorial library and fluorogenic peptides, *Biochemical Journal*, 323, 427-433, **1997b**.

Di Santo, R., Tafi, A., Costi, R., Botta, M., Artico, M., Corelli, F., Forte, M., Caporuscio, F., Angiolella, L., Palamara, A.T., Antifungal agents. 11. N-substituted derivatives of 1-[(aryl)(4-aryl-1H-pyrrol-3-yl)methyl]-1H-imidazole: Synthesis, anti-Candida activity, and QSAR studies, *Journal of Medicinal Chemistry*, 48, 5140-5153, **2005**.

Dias, R.L.A., Correa, A.G., Applications of combinatorial chemistry to drug development, *Quimica Nova*, 24, 236-242, **2001**.

do Amaral, A.T., Oliveira, A.C., Neidlein, R., Gallacci, M., Caprara, L., Miyazaki, Y., Physicochemical parameters involved in the lethal toxicity of N,N-[(dimethylamino)ethyl]-4-substituted benzoate hydrochlorides: a QSAR study, *European Journal of Medicinal Chemistry*, 32, 433-443, **1997**.

Docampo, R., Sensitivity of parasites to free-radical damage by antiparasitic drugs, *Chemico-Biological Interactions*, 73, 1-27, **1990**.

Docampo, R., Moreno, S.N.J., Free-radical metabolites in the mode of action of chemotherapeutic-agents and phagocytic-cells on trypanosoma-cruzi, *Reviews of Infectious Diseases*, 6, 223-238, **1984**.

Doman, T.N., McGovern, S.L., Witherbee, B.J., Kasten, T.P., Kurumbail, R., Stallings, W.C., Connolly, D.T., Shoichet, B.K., Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B, *Journal of Medicinal Chemistry*, 45, 2213-2221, **2002**.

Doyle, P.S., Zhou, Y.M., Engel, J.C., McKerrow, J.H., A Cysteine Protease Inhibitor Cures Chagas' Disease in an Immunodeficient-Mouse Model of Infection, *Antimicrob.Agents Chemother.*, 51, 3932-3939, **2007**.

Drews, J., Drug discovery: A historical perspective, *Science*, 287, 1960-1964, **2000**.

Du, X.H., Guo, C., Hansell, E., Doyle, P.S., Caffrey, C.R., Holler, T.P., McKerrow, J.H., Cohen, F.E., Synthesis and structure-activity relationship study of potent trypanocidal thio semicarbazone inhibitors of the trypanosomal cysteine protease cruzain, *Journal of Medicinal Chemistry*, 45, 2695-2707, **2002**.

Du, X.H., Hansell, E., Engel, J.C., Caffrey, C.R., Cohen, F.E., McKerrow, J.H., Aryl ureas represent a new class of anti-trypanosomal agents, *Chemistry & Biology*, 7, 733-742, **2000**.

Dunn, D.A., Feygin, I., Challenges and solutions to ultra-high-throughput screening assay miniaturization: submicroliter fluid handling, *Drug Discovery Today*, 5, S84-S91, **2000**.

Eakin, A.E., Mcgrath, M.E., McKerrow, J.H., Fletterick, R.J., Craik, C.S., Production of crystallizable cruzain, the major cysteine protease from Trypanosoma-cruzi, *Journal of Biological Chemistry*, 268, 6115-6118, **1993**.

Eakin, A.E., Mills, A.A., Harth, G., McKerrow, J.H., Craik, C.S., The sequence, organization, and expression of the major cysteine protease (cruzain) from Trypanosoma-cruzi, *Journal of Biological Chemistry*, 267, 7411-7420, **1992**.

Eddershaw, P.J., Beresford, A.P., Bayliss, M.K., ADME/PK as part of a rational approach to drug discovery, *Drug Discovery Today*, 5, 409-414, **2000**.

Elhalem, E., Bailey, B.N., Docampo, R., Ujvary, I., Szajnman, S.H., Rodriguez, J.B., Design, synthesis, and biological evaluation of aryloxyethyl thiocyanate derivatives against Trypanosoma cruzi, *Journal of Medicinal Chemistry*, 45, 3984-3999, **2002**.

Engel, J.C., Doyle, P.S., Hsieh, I., McKerrow, J.H., Cysteine protease inhibitors cure an experimental Trypanosoma cruzi infection, *Journal of Experimental Medicine*, 188, 725-734, **1998**.

Ewing, T.J.A., Kuntz, I.D., Critical evaluation of search algorithms for automated molecular docking and database screening, *Journal of Computational Chemistry*, 18, 1175-1189, **1997**.

Fairlamb, A.H., Blackburn, P., Ulrich, P., Chait, B.T., Cerami, A., Trypanothione - A Novel Bis(Glutathionyl)Spermidine Cofactor For Glutathione-Reductase In Trypanosomatids, *Science*, 227, 1485-1487, **1985**.

Fairlamb, A.H., Cerami, A., Identification of a novel, thiol-containing cofactor essential for glutathione-reductase enzyme-activity in trypanosomatids, *Molecular and Biochemical Parasitology*, 14, 187-198, **1985**.

Feng, B.Y., Shelat, A., Doman, T.N., Guy, R.K., Shoichet, B.K., High-throughput assays for promiscuous inhibitors, *Nature Chemical Biology*, 1, 146-148, **2005**.

Feng, B.Y., Shoichet, B.K., A Detergent-Based Assay for the Detection of Promiscuous Inhibitors, *Nat.Protoc.PMC*, 1, 550-553, **2006**.

Filardi, L.S., Brener, Z., Susceptibility and natural-resistance of trypanosoma-cruzi strains to drugs used clinically in chagas-disease, *Transactions of the Royal Society of Tropical Medicine and Hygiene* , 81, 755-759, **1987**.

Freitas, R.F., Oprea, T.I., Montanari, C.A., 2D QSAR and similarity studies on cruzain inhibitors aimed at improving selectivity over cathepsin L, *Bioorganic & Medicinal Chemistry*, In Press, Corrected Proof, **2007**.

Freymann, D.M., Wenck, M.A., Engel, J.C., Feng, J., Focia, P.J., EAKIN, A.E., Craig, S.P., Efficient identification of inhibitors targeting the closed active site conformation of the HPRT from Trypanosoma cruzi, *Chemistry & Biology*, 7, 957-968, **2000**.

Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K., Shaw, D.E., Francis, P., Shenkin, P.S., Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *Journal of Medicinal Chemistry*, 47, 1739-1749, **2004**.

Fujii, N., Mallari, J.P., Hansell, E.J., Mackey, Z., Doyle, P., Zhou, Y.M., Gut, J., Rosenthal, P.J., McKerrow, J.H., Guy, R.K., Discovery of potent thiosemicarbazone inhibitors of rhodesain and cruzain, *Bioorganic & Medicinal Chemistry Letters*, 15, 121-123, **2005**.

Furka, A., Sebestyen, F., Asgedom, M., Dibo, G., General-method for rapid synthesis of multicomponent peptide mixtures, *International Journal of Peptide and Protein Research*, 37, 487-493, **1991**.

Gelb, M.H., Hol, W.G.J., Parasitology - Drugs to combat tropical protozoan parasites, *Science*, 297, 343-344, **2002**.

Gilbert, I.H., Inhibitors of dihydrofolate reductase in leishmania and trypanosomes, *Biochimica et Biophysica Acta-Molecular Basis of Disease*, 1587, 249-257, **2002**.

Gillmor, S.A., Craik, C.S., Fletterick, R.J., Structural determinants of specificity in the cysteine protease cruzain, *Protein Science*, 6, 1603-1611, **1997**.

Gohlke, H., Hendlich, M., Klebe, G., Knowledge-based scoring function to predict protein-ligand interactions, *Journal of Molecular Biology*, 295, 337-356, **2000a**.

Gohlke, H., Hendlich, M., Klebe, G., Predicting binding modes, binding affinities and 'hot spots' for protein-ligand complexes using a knowledge-based scoring function, *Perspectives in Drug Discovery and Design*, 20, 115-144, **2000b**.

Gonzalez, F.V., Izquierdo, J., Rodri'guez, S., McKerrow, J.H., Hansell, E., Dipeptidyl-[alpha],[beta]-epoxyesters as potent irreversible inhibitors of the cysteine proteases cruzain and rhodesain, *Bioorganic & Medicinal Chemistry Letters*, 17, 6697-6700, **2007** .

Goodford, P.J., A computational-procedure for determining energetically favorable binding-sites on biologically important macromolecules, *Journal of Medicinal Chemistry*, 28, 849-857, **1985**.

Greenbaum, D.C., Mackey, Z., Hansell, E., Doyle, P., Gut, J., Caffrey, C.R., Lehrman, J., Rosenthal, P.J., McKerrow, J.H., Chibale, K., Synthesis and structure-activity relationships of parasiticidal thiosemicarbazone cysteine protease inhibitors against Plasmodium falciparum, Trypanosoma brucei, and Trypanosoma cruzi, *Journal of Medicinal Chemistry*, 47, 3212-3219, **2004**.

Gruneberg, S., Stubbs, M.T., Klebe, G., Successful virtual screening for novel inhibitors of human carbonic anhydrase: Strategy and experimental confirmation, *Journal of Medicinal Chemistry*, 45, 3588-3602, **2002**.

Gruneberg, S., Wendt, B., Klebe, G., Subnanomolar inhibitors from computer screening: A model study using human carbonic anhydrase II, *Angewandte Chemie-International Edition*, 40, 389-393, **2001**.

Guex, N., Peitsch, M.C., SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling, *Electrophoresis*, 18, 2714-2723, **1997**.

Guillier, F., Orain, D., Bradley, M., Linkers and cleavage strategies in solid-phase organic synthesis and combinatorial chemistry (vol 100, pg 2091, 2000), *Chemical Reviews*, 100, 3859, **2000**.

Hajduk, P.J., Huth, J.R., Fesik, S.W., Druggability indices for protein targets derived from NMR-based screening data, *Journal of Medicinal Chemistry*, 48, 2518-2525, **2005a**.

Hajduk, P.J., Huth, J.R., Tse, C., Predicting protein druggability, *Drug Discovery Today*, 10, 1675-1682, **2005b**.

Hann, M.M., Leach, A.R., Harper, G., Molecular complexity and its impact on the probability of finding leads for drug discovery, *Journal of Chemical Information and Computer Sciences*, 41, 856-864, **2001**.

Hansson, T., Oostenbrink, C., van Gunsteren, W.F., Molecular dynamics simulations, *Current Opinion in Structural Biology*, 12, 190-196, **2002**.

Harris, J.L., Backes, B.J., Leonetti, F., Mahrus, S., Ellman, J.A., Craik, C.S., Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries, *Proceedings of the National Academy of Sciences of the United States of America*, 97, 7754-7759, **2000**.

Harrison, M.J., Burton, N.A., Hillier, I.H., Catalytic mechanism of the enzyme papain: Predictions with a hybrid quantum mechanical molecular mechanical potential, *Journal of the American Chemical Society* , 119, 12285-12291, **1997**.

Harth, G., Andrews, N., Mills, A.A., Engel, J.C., Smith, R., McKerrow, J.H., Peptide-fluoromethyl ketones arrest intracellular replication and intercellular transmission of Trypanosoma-cruzi, *Molecular and Biochemical Parasitology*, 58, 17-24, **1993**.

Hartshorn, M.J.; Murray, C.W.; Cleasby, A.; Frederickson, M.; Tickle, I.J.; Jhoti, H., Fragment-Based Lead Discovery Using X-Ray Crystallography. *Journal of Medicinal Chemistry*, 48, 403-413, **2005**.

Hendlich, M., Rippmann, F., Barnickel, G., LIGSITE: Automatic and efficient detection of potential small molecule-binding sites in proteins, *Journal of Molecular Graphics & Modelling*, 15, 359-+, **1997**.

Hibert, M.F., Gittos, M.W., Middlemiss, D.N., Mir, A.K., Fozard, J.R., Graphics computer-aided receptor mapping as a predictive tool for drug design - development of potent, selective, and stereospecific ligands for the 5-ht1a receptor, *Journal of Medicinal Chemistry*, 31, 1087-1093, **1988**.

Hodgson, J., ADMET - turning chemicals into drugs, *Nature Biotechnology*, 19, 722-726, **2001**.

Homans, S.W., Water, water everywhere - except where it matters?, *Drug Discovery Today*, 12, 534-539, **2007**.

Hopkins, A.L., Groom, C.R., The druggable genome, *Nature Reviews Drug Discovery*, 1, 727-730, **2002**.

Hopkins, A.L.; Groom, C.R.; Alex, A., Ligand Efficiency: a Useful Metric for Lead Selection. *Drug Discovery Today*, 9, 430-431, **2004**.

Horrobin, D.F., Modern biomedical research: an internally self-consistent universe with little contact with medical reality?, *Nature Reviews Drug Discovery*, 2, 151-154, **2003**.

Höfliger, M. M. and A. G. Beck-Sickinger, Receptor-Ligand Interaction, in *Protein-Ligand Interactions*, edited by H. J. Bohm and G. Schneider, pp. 107-136, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim,**2003**.

Huang, L., Brinen, L.S., Ellman, J.A., Crystal structures of reversible ketone-based inhibitors of the cysteine protease cruzain, *Bioorganic & Medicinal Chemistry*, 11, 21-29, **2003**.

Huang, L., Lee, A., Ellman, J.A., Identification of potent and selective mechanism-based inhibitors of the cysteine protease cruzain using solid-phase parallel synthesis, *Journal of Medicinal Chemistry*, 45, 676-684, **2002**.

Ifa, D.R., Rodrigues, C.R., de Alencastro, R.B., Fraga, C.A.M., Barreiro, E.J., A possible molecular mechanism for the inhibition of cysteine proteases by salicylaldehyde N-acylhydrazones and related compounds, *Journal of Molecular Structure-Theochem*, 505, 11-17, **2000**.

Irwin, J.J., Shoichet, B.K., ZINC - A free database of commercially available compounds for virtual screening, *Journal of Chemical Information and Modeling*, 45, 177-182, **2005**.

Ishiki, H. M., Relações Quantitativas estrutura-atividade QSAR/QSAR-3D de dipiridodiazepinonas, inibidoras da transcriptase reversa do virus da imunodeficiência humana do tipo I (HIV-1), Tese de Mestrado, FFCLRP-USP, **1999**.

Ishiki, H. M., Quantitativas entre Estrutura Química e Atividade Biológica (QSAR/QSAR-3D) de compostos com potencial atividade antituberculose, IQ - USP, **2005**.

Ishiki, H. M., Galembeck, S. E, and do Amaral, A. T., Application of two alignments procedures in CoMFA analysis of nevirapine derivative. *Proceedings of 13th European Symposium on Quantitative Structure-Activity Relationships: Rational Approaches to Drug Design*, Amsterdam, **2001**

Jacobsen, W., Christians, U., Benet, L.Z., In vitro evaluation of the disposition of a novel cysteine protease inhibitor, *Drug Metabolism and Disposition*, 28, 1343-1351, **2000**.

Jacobsson, M., Liden, P., Stjernschantz, E., Bostrom, H., Norinder, U., Improving structure-based virtual screening by multivariate analysis of scoring data, *Journal of Medicinal Chemistry*, 46, 5781-5789, **2003**.

Jain, A.N., Surflex-Dock 2.1: Robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search, *Journal of Computer-Aided Molecular Design*, 21, 281-306, **2007**.

Jaishankar, P., Hansell, E., Zhao, D.M., Doyle, P.S., McKerrow, J.H., Renslo, A.R., Potency and selectivity of P2/P3-modified inhibitors of cysteine proteases from trypanosomes, *Bioorganic & Medicinal Chemistry Letters*, In Press, Corrected Proof, **2007**.

Jannin, J., Villa, L., An overview of Chagas disease treatment, *Memórias do Instituto Oswaldo Cruz*, 102, 95-98, **2007**.

Jenkins, J.L., Kao, R.Y.T., Shapiro, R., Virtual screening to enrich hit lists from high-throughput screening: A case study on small-molecule inhibitors of angiogenin, *Proteins-Structure Function and Genetics*, 50, 81-93, **2003**.

Jockersscherubl, M.C., Schirmer, R.H., Krauthsiegel, R.L., Trypanothione reductase from trypanosoma-cruzi - catalytic properties of the enzyme and inhibition studies with trypanocidal compounds, *European Journal of Biochemistry*, 180, 267-272, **1989**.

Jonckers, T.H.M., van Miert, S., Cimanga, K., Bailly, C., Colson, P., Pauw-Gillet, M.C., van den Heuvel, H., Claeys, M., Lemiere, F., Esmans, E.L., Rozenski, J., Quirijnen, L., Maes, L., Dommisse, R., Lemiere, G.L.F., Vlietinck, A., Pieters, L., Synthesis, cytotoxicity, and antiplasmodial and antitrypanosomal activity of new neocryptolepine derivatives, *Journal of Medicinal Chemistry*, 45, 3497-3508, **2002**.

Judice, W.A.S., Cezari, M.H.S., Lima, A.P.C.A., Scharfstein, J., Chagas, J.R., Tersariol, I.L.S., Juliano, M.A., Juliano, L., Comparison of the specificity, stability and individual rate constants with respective activation parameters for the peptidase activity of cruzipain and its recombinant form, cruzain, from Trypanosoma cruzi, *European Journal of Biochemistry*, 268, 6578-6586, **2001**.

Kamath, L., ADMETox improves drug discovery rates - Combining with in silico modeling, in vitro testing, automation, and high-throughput screening, *Genetic Engineering News*, 26, 1-+, **2006**.

Karnachi, P. and A. Kulkarni, Application of Pharmacophore Fingerprints to Structure-based Design and Data Mining , in *Pharmacophores and Pharmacophore Searches*, edited by T. Langer and R. D. Hoffmann, pp. 193-204, WILEY-VCH, Weinheim, Germany **2006**.

Karplus, M., Petsko, G.A., Molecular-dynamics simulations in biology, *Nature*, 347, 631-639, **1990**.

Kastenholz, M.A., Pastor, M., Cruciani, G., Haaksma, E.E.J., Fox, T., GRID/CPCA: A new computational tool to design selective ligands, *Journal of Medicinal Chemistry*, 43, 3033-3044, **2000**.

Kellenberger, E., Rodrigo, J., Muller, P., Rognan, D., Comparative evaluation of eight docking tools for docking and virtual screening accuracy, *Proteins-Structure Function and Bioinformatics*, 57, 225-242, **2004**.

Kitchen, D.B., Decornez, H., Furr, J.R., Bajorath, J., Docking and scoring in virtual screening for drug discovery: Methods and applications, *Nature Reviews Drug Discovery*, 3, 935-949, **2004**.

Klebe, G., Recent developments in structure-based drug design, *Journal of Molecular Medicine-Jmm*, 78, 269-281, **2000**.

Klebe, G., Virtual ligand screening: strategies, perspectives and limitations, *Drug Discovery Today*, 11, 580-594, **2006**.

Kontoyianni, M., McClellan, L.M., Sokol, G.S., Evaluation of docking performance: Comparative data on docking algorithms, *Journal of Medicinal Chemistry*, 47, 558-565, **2004**.

Kontoyianni, M., Sokol, G.S., McClellan, L.M., Evaluation of library ranking efficacy in virtual screening, *Journal of Computational Chemistry*, 26, 11-22, **2005**.

Koppitz, M., Eis, K., Automated medicinal chemistry, *Drug Discovery Today*, 11, 561-568, **2006**.

Kubinyi, H., Strategies and recent technologies in drug discovery, *Pharmazie*, 50, 647-662, **1995**.

Kubinyi, H., Combinatorial and computational approaches in structure-based drug design, *Current Opinion in Drug Discovery and Development*, 1, 16-27, **1998a**.

Kubinyi, H., Similarity and dissimilarity: A medicinal chemist's view, *Perspectives in Drug Discovery and Design*, 9-11, 225-252, **1998b**.

Kubinyi, H., Structure-based design of enzyme inhibitors and receptor ligands, *Current Opinion in Drug Discovery and Development*, 1, 4-15, **1998c**.

Kubinyi, H., Drug research: myths, hype and reality, *Nature Reviews Drug Discovery*, 2, 665-668, **2003**.

Kubinyi, H., Combinatorial Chemistry in the Life Sciences, General Approuches in Drug Design, **2004**

Kubinyi, H., New Approaches in Drug Design, **16-11-2005**

Kubinyi, H., Success Stories of Computer-Aided Design, in *Computer applications in pharmaceutical research and development*, pp. 377-424, Wiley, Hoboken,NJ, Estados Unidos **2006**.

Lahana, R., How many leads from HTS?, *Drug Discovery Today*, 4, 447-448, **1999**.

Lecaille, F., Authie, E., Moreau, T., Serveau, C., Gauthier, F., Lalmanach, G., Subsite specificity of trypanosomal cathepsin L-like cysteine proteases - Probing the S2 pocket with phenylalanine-derived amino acids, *European Journal of Biochemistry*, 268, 2733-2741, **2001**.

Lecaille, F., Kaleta, J., Bromme, D., Human and parasitic papain-like cysteine proteases: Their role in physiology and pathology and recent developments in inhibitor design, *Chemical Reviews*, 102, 4459-4488, **2002**.

Leite, A.C.L., de Lima, R.S., Moreira, D.R.D., Cardoso, V.D.O., de Brito, A.C.G., dos Santos, L.M.F., Hernandes, M.Z., Kiperstok, A.C., de Lima, R.S., Soares, M.B.P., Synthesis, docking, and in vitro activity of thiosemicarbazones, aminoacyl-thiosemicarbazides and acyl-thiazolidones against Trypanosoma cruzi, *Bioorganic & Medicinal Chemistry*, 14, 3749-3757, **2006**.

Leite, A.C.L., Moreira, D.R.D., Cardoso, M.V.D., Hernandes, M.Z., Pereira, V.R.A., Silva, R.O., Kiperstok, A.C., Lima, M.D., Soares, M.B.P., Synthesis, cruzain docking, and in vitro studies of aryl-4-oxothiazolylhydrazones against Trypanosoma cruzi, *Chemmedchem*, 2, 1339-1345, **2007**.

Leung, D., Abbenante, G., Fairlie, D.P., Protease inhibitors: Current status and future prospects, *Journal of Medicinal Chemistry*, 43, 305-341, **2000**.

Li, R.B., Sirawaraporn, R., Chitnumsub, P., Sirawaraporn, W., Wooden, J., Athappilly, F., Turley, S., Hol, W.G.J., Three-dimensional structure of M-tuberculosis dihydrofolate reductase reveals opportunities for the design of novel tuberculosis drugs, *Journal of Molecular Biology*, 295, 307-323, **2000**.

Li, R.S., Chen, X.W., Gong, B.Q., Selzer, P.M., Li, Z., Davidson, E., Kurzban, G., Miller, R.E., Nuzum, E.O., McKerrow, J.H., Fletterick, R.J., Gillmor, S.A., Craik, C.S., Kuntz, I.D., Cohen, F.E., Kenyon, G.L., Structure-based design of parasitic protease inhibitors, *Bioorganic & Medicinal Chemistry*, 4, 1421-1427, **1996**.

Li, R.S., Kenyon, G.L., Cohen, F.E., Chen, X.W., Gong, B.Q., Dominguez, J.N., Davidson, E., Kurzban, G., Miller, R.E., Nuzum, E.O., Rosenthal, P.J., McKerrow, J.H., In vitro antimalarial activity of chalcones and their derivatives, *Journal of Medicinal Chemistry*, 38, 5031-5037, **1995**.

Lima, A.P.C.A., dos Reis, F.C.G., Serveau, C., Lalmanach, G., Juliano, L., Menard, R., Vernet, T., Thomas, D.Y., Storer, A.C., Scharfstein, J., Cysteine protease isoforms from Trypanosoma cruzi, cruzipain 2 and cruzain, present different substrate preference and susceptibility to inhibitors, *Molecular and Biochemical Parasitology*, 114, 41-52, **2001**.

Lindahl, E., Hess, B., van der Spoel, D., GROMACS 3.0: a package for molecular simulation and trajectory analysis, *Journal of Molecular Modeling*, 7, 306-317, **2001**.

Lipinski, C.A., Drug-like properties and the causes of poor solubility and poor permeability, *Journal of Pharmacological and Toxicological Methods*, 44, 235-249, **2000**.

Lipinski, C.A., Hopkins, A., Navigating chemical space for biology and medicine, *Nature*, 432, 855-861, **2004**.

Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Advanced Drug Delivery Reviews*, 23, 3-25, **1997**.

Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Advanced Drug Delivery Reviews*, 46, 3-26, **2001**.

Liu, M., Wilairat, P., Croft, S.L., Tan, A.L.C., Go, M.L., Structure-activity relationships of antileishmanial and antimalarial chalcones, *Bioorganic & Medicinal Chemistry*, 11, 2729-2738, **2003**.

Lloyd, D.G., Golfis, G., Knox, A.J.S., Fayne, D., Meegan, M.J., Oprea, T.I., Oncology exploration: charting cancer medicinal chemistry space, *Drug Discovery Today*, 11, 149-159, **2006**.

Lopez, J.A., Carvalho, T.U., de Souza, W., Flohe, L., Guerrero, S.A., Montemartini, M., Kalisz, H.M., Nogoceke, E., Singh, M., Alves, M.J.M., Colli, W., Evidence for a trypanothione-dependent peroxidase system in Trypanosoma cruzi, *Free Radical Biology and Medicine*, 28, 767-772, **2000**.

Lunardi, F., Guzela, M., Rodrigues, A.T., Correa, R., Eger-Mangrich, I., Steindel, M., Grisard, E.C., Assreuy, J., Calixto, J.B., Santos, A.R.S., Trypanocidal and leishmanicidal properties of substitution-containing chalcones, *Antimicrob.Agents Chemother.*, 47, 1449-1451, **2003**.

Luque, I., Freire, E., Structural stability of binding sites: Consequences for binding affinity and allosteric effects, *Proteins-Structure Function and Genetics*, 63-71, **2000**.

Mack, H., Pfeiffer, T., Hornberger, W., bohm, H.J., Hoffken, H.W., Design, synthesis and biological-activity of novel rigid amidino-phenylalanine derivatives as inhibitors of thrombin, *Journal of Enzyme Inhibition*, 9, 73-&, **1995**.

MacKerell, A. D. Jr., Atomistic Models and Force Filds, in *Computational Biochemistry and Biophysics*, edited by B. Roux, M. Watanabe, and O. M. Becker, pp. 7-38, Marcel Dekker, New York, **2001**.

Malvezzi, A., Estudo comparativo de parâmetros hidrofóbicos e, relacionados à ionização, de série de derivados da procaína com atividade bloqueadora neuromuscular, Tese de Mestrado, IQ-USP, **2000**.

Malvezzi, A., de Rezende, L., Izidoro, M.A., Cezari, M.H.S., Juliano, L., Amaral, A.T., Uncovering false positives on a virtual screening search for cruzain inhibitors, *Bioorganic & Medicinal Chemistry Letters*, 18, 350-354, **2008**.

Malvezzi, A., do Amaral, A. T., and Rezende, L, Busca de Moléculas Candidatas a Inibidores da Cruzaína com o Auxílio de Computador, painel apresentado na 27a. Reunião Anual da Sociedade Brasileira de Química, Salvador, BA, 30 de Maio - 02 de Junho, **2004**

Malvezzi, A., do Amaral, A. T., and Rezende, L, Uso de Consensus Scoring na Busca de Moléculas Candidatas a Inibidores da Cruzaína", painel apresentado na 28a. Reunião Anual da Sociedade Brasileira de Química, Poços de Caldas, MG, 30 de maio - 02 de junho, **2005**

Malvezzi, A., do Amaral, A. T., and Rezende, L, Virtual Screening de Inibidores da Cruzaína, painel apresentado na 29a. Reunião Anual da Sociedade Brasileira de Química, Águas de Lindóia, SP, 19 - 22 de maio, **2006**

Martin, Y.C., Willett, P., Lajiness, M., Johnson, M., Maggiora, G., Martin, E., Bures, M.G., GASTEIGER, J., Cramer, R.D., Pearlman, R.S., Mason, J.S., Diverse viewpoints on computational aspects of molecular diversity, *Journal of Combinatorial Chemistry*, 3, 231-250, **2001**.

Mattos, C. and D. Ringe, Multiple binding modes, in *3D QSAR in drug design: Theory, methods and applications*, edited by H. Kubinyi, p. 226, ESCOM, Leiden, The Netherlands **1993**.

Maybridge Chemicals Ltd., Trevillet, Tintangel, Cornwall PL34 OHW, Inglaterra,

McGovern, S.L., Caselli, E., Grigorieff, N., Shoichet, B.K., A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening, *Journal of Medicinal Chemistry*, 45, 1712-1722, **2002**.

McGovern, S.L., Helfand, B.T., Feng, B., Shoichet, B.K., A specific mechanism of nonspecific inhibition, *Journal of Medicinal Chemistry*, 46, 4265-4272, **2003**.

McGovern, S.L., Shoichet, B.K., Kinase inhibitors: Not just for kinases anymore, *Journal of Medicinal Chemistry*, 46, 1478-1483, **2003**.

Mcgrath, M.E., Eakin, A.E., Engel, J.C., McKerrow, J.H., Craik, C.S., Fletterick, R.J., The crystal-structure of cruzain - A therapeutic target for chagas-disease, *Journal of Molecular Biology*, 247, 251-259, **1995**.

McGregor, M.J., Muskal, S.M., Pharmacophore fingerprinting. 1. Application to QSAR and focused library design, *Journal of Chemical Information and Computer Sciences*, 39, 569-574, **1999**.

McKerrow, J.H., Development of cysteine protease inhibitors as chemotherapy for parasitic diseases: insights on safety, target validation, and mechanism of action, *International Journal for Parasitology*, 29, 833-837, **1999**.

Meirelles, M.N.L., Juliano, L., Carmona, E., Silva, S.G., Costa, E.M., Murta, A.C.M., Scharfstein, J., Inhibitors of the major cysteinyl proteinase (GP57/51) impair host-cell invasion and arrest the intracellular development of Trypanosoma-cruzi invitro, *Molecular and Biochemical Parasitology*, 52, 175-184, **1992**.

Merrifield, R.B., Solid Phase Peptide Synthesis .1. Synthesis Of A Tetrapeptide, *Journal of the American Chemical Society*, 85, 2149-&, **1963**.

Millet, R., Maes, L., Landry, V., Sergheraert, C., Davioud-Charvet, E., Antitrypanosomal activities and cytotoxicity of 5-nitro-2- furancarbohydrazides, *Bioorganic & Medicinal Chemistry Letters*, 12, 3601-3604, **2002**.

Miyamoto, S., Kollman, P.A., What determines the strength of noncovalent association of ligands to proteins in aqueous-solution, *Proceedings of the National Academy of Sciences of the United States of America*, 90, 8402-8406, **1993**.

Moreno, S.N.J., Mason, R.P., Muniz, R.P.A., Cruz, F.S., Docampo, R., Generation of free-radicals from metronidazole and other nitroimidazoles by tritrichomonas-foetus, *Journal of Biological Chemistry*, 258, 4051-4054, **1983**.

Moro, S., Braiuca, P., Deflorian, F., Ferrari, C., Pastorin, G., Cacciari, B., Baraldi, P.G., Varani, K., Borea, P.A., Spalluto, G., Combined target-based and ligand-based drug design approach as a tool to define a novel 3D-pharmacophore model of human A(3) adenosine receptor antagonists: Pyrazolo[4,3-e]1,2,4-triazolo[1,5-c]pyrimidine derivatives as a key study, *Journal of Medicinal Chemistry*, 48, 152-162, **2005**.

Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., Olson, A.J., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *Journal of Computational Chemistry*, 19, 1639-1662, **1998**.

Murray, C.W., Baxter, C.A., Frenkel, A.D., The sensitivity of the results of molecular docking to induced fit effects: Application to thrombin, thermolysin and neuraminidase, *Journal of Computer-Aided Molecular Design*, 13, 547-562, **1999**.

NCI-NIH, Developmental Therapeutics Program, **2005**

Nicolaou, K.C., Winssinger, N., Vourloumis, D., Ohshima, T., Kim, S., Pfefferkorn, J., Xu, J.Y., Li, T., Solid and solution phase synthesis and biological evaluation of combinatorial sarcodictyin libraries, *Journal of the American Chemical Society*, 120, 10814-10826, **1998**.

Nielsen, S.F., Christensen, S.B., Cruciani, G., Kharazmi, A., Liljefors, T., Antileishmanial chalcones: Statistical design, synthesis, and three-dimensional quantitative structure-activity relationship analysis, *Journal of Medicinal Chemistry*, 41, 4819-4832, **1998**.

Oprea, T.I., Property distribution of drug-related chemical databases, *Journal of Computer-Aided Molecular Design*, 14, 251-264, **2000**.

Oprea, T.I., Virtual screening in lead discovery: A viewpoint, *Molecules*, 7, 51-62, **2002**.

Oprea, T.I., Davis, A.M., Teague, S.J., Leeson, P.D., Is there a difference between leads and drugs? A historical perspective, *Journal of Chemical Information and Computer Sciences*, 41, 1308-1315, **2001**.

Oprea, T.I., Gottfries, J., Chemography: The art of navigating in chemical space, *Journal of Combinatorial Chemistry*, 3, 157-166, **2001**.

Originlab®, Corporation ©, *Microcal Software*, Inc. One Roundhouse Plasa. Northampton MA, USA, **2002**

Ortholand, J.Y., Ganesan, A., Natural products and combinatorial chemistry: back to the future, *Current Opinion in Chemical Biology*, 8, 271-280, **2004**.

Otto, H.H., Schirmeister, T., Cysteine proteases and their inhibitors, *Chemical Reviews*, 97, 133-171, **1997**.

Palmer, J.T., Rasnick, D., Klaus, J.L., Bromme, D., Vinyl sulfones as mechanism-based cysteine protease inhibitors, *Journal of Medicinal Chemistry*, 38, 3193-3196, **1995**.

Parker, C.N., Bajorath, J., Towards unified compound screening strategies: A critical evaluation of error sources in experimental and virtual high-throughput screening, *Qsar & Combinatorial Science*, 25, 1153-1161, **2006**.

Pastor, M., Cruciani, G., McLay, I., Pickett, S., Clementi, S., GRid-INdependent descriptors (GRIND): A novel class of alignment-independent three-dimensional molecular descriptors, *Journal of Medicinal Chemistry*, 43, 3233-3243, **2000**.

Patel, Y., Gillet, V.J., Bravi, G., Leach, A.R., A comparison of the pharmacophore identification programs: Catalyst, DISCO and GASP, *Journal of Computer-Aided Molecular Design*, 16, 653-681, **2002**.

Paul, N., Rognan, D., ConsDock: A new program for the consensus analysis of protein-ligand interactions, *Proteins-Structure Function and Genetics*, 47, 521-533, **2002**.

Pavao, F., Castilho, M.S., Pupo, M.T., Dias, R.L.A., Correa, A.G., Fernandes, J.B., da Silva, M.F.G.F., Mafezoli, J., Vieira, P.C., Oliva, G., Structure of Trypanosoma cruzi glycosomal glyceraldehyde-3-phosphate dehydrogenase complexed with chalepin, a natural product inhibitor, at 1.95 angstrom resolution, *Febs Letters*, 520, 13-17, **2002**.

Pecoul, B., Chirac, P., Trouiller, P., Pinel, J., Access to essential drugs in poor countries - A lost battle?, *Journal of the American Medical Association*, 281, 361-367, **1999**.

Peng, H., Huang, N., Qi, J., Xie, P., Xu, C., Wang, J.X., Yang, C.Z., Identification of novel inhibitors of BCR-ABL tyrosine kinase via virtual screening, *Bioorganic & Medicinal Chemistry Letters*, 13, 3693-3699, **2003**.

Perola, E., Charifson, P.S., Conformational analysis of drug-like molecules bound to proteins: An extensive study of ligand reorganization upon binding, *Journal of Medicinal Chemistry*, 47, 2499-2510, **2004**.

Perola, E., Walters, W.P., Charifson, P.S., A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance, *Proteins-Structure Function and Bioinformatics*, 56, 235-249, **2004**.

Pires, J. R., Relações Quantitativas entre a Estrutura Química de Derivados Nitrofurânicos e a Atividade Antimicrobiana contra Caulobacter crescentus e Staphylococcus aureus., Tese de Doutorado, IQ-USP, **1998**.

Pires, J.R., Saito, C., Gomes, S.L., Giesbrecht, A.M., Amaral, A.T.D., Investigation of 5-nitrofuran derivatives: Synthesis, antibacterial activity, and quantitative structure-activity relationships, *Journal of Medicinal Chemistry*, 44, 3673-3681, **2001**.

Poptodorov, K., T. Luu, and R. D. Hoffmann, Pharmacophore Model Generation Software Tools, in *Pharmacophores and Pharmacophore Searches*, edited by T. Langer and R. D. Hoffmann, pp. 17-44, Wiley-VCH, Weinheim, Germany **2006**.

Powers, J.C., Asgian, J.L., Ekici, O.D., James, K.E., Irreversible inhibitors of serine, cysteine, and threonine proteases, *Chemical Reviews*, 102, 4639-4750, **2002**.

Raag, R., Poulos, T.L., Crystal-structures of cytochrome-p-450cam complexed with camphane, thiocamphor, and adamantane - factors controlling p-450 substrate hydroxylation, *Biochemistry*, 30, 2674-2684, **1991**.

Rahman, A.A., Daoud, M.K., Dukat, M., Herrick-Davis, K., Purohit, A., Teitler, M., do Amaral, A.T., Malvezzi, A., Glennon, R.A., Conformationally-restricted analogues and partition coefficients of the 5-HT3 serotonin receptor ligands meta-chlorophenylbiguanide (mCPBG) and meta-chlorophenylguanidine (mCPG), *Bioorganic & Medicinal Chemistry Letters*, 13, 1119-1123, **2003**.

Raminelli, C., Estudo da relações quantitativa entre a estrutura química e atividade citotóxica de séries de derivads de bases de Mannich, Tese de Mestrado, IQ-USP, **2001**.

Rao, B.G., Tilton, R.F., Singh, U.C., Free-energy perturbation studies on inhibitor binding to hiv-1 proteinase, *Journal of the American Chemical Society*, 114, 4447-4452, **1992**.

Rarey, M., Dixon, J.S., Feature trees: A new molecular similarity measure based on tree matching, *Journal of Computer-Aided Molecular Design*, 12, 471-490, **1998**.

Rarey, M., P. Fricker, S. Hindle, G. Metz, C. Rummey, and M. Zimmermann, Feature Trees: Theory and Applications from Large-scale Virtual Screening to Data Analysis, in *Pharmacophores and Pharmacophore Searches*, edited by T. Langer and R. D. Hoffmann, pp. 81-114, Wiley-VCH, Weinheim,**2006**.

Rarey, M., Kramer, B., Lengauer, T., Multiple automatic base selection: Protein-ligand docking based on incremental construction without manual intervention, *Journal of Computer-Aided Molecular Design*, 11, 369-384, **1997**.

Reddy, M.R., Erion, M.D., Relative binding affinities of fructose-1,6-bisphosphatase inhibitors calculated using a quantum mechanics-based free energy perturbation method, *Journal of the American Chemical Society*, 129, 9296-+, **2007**.

Reddy, M.R., Viswanadhan, V.N., Weinstein, J.N., Relative differences in the binding free-energies of human immunodeficiency virus-1 protease inhibitors - a thermodynamic cycle-perturbation approach, *Proceedings of the National Academy of Sciences of the United States of America*, 88, 10287-10291, **1991**.

Rester, U., Dock around the clock - Current status of small molecule docking and scoring, *Qsar & Combinatorial Science*, 25, 605-615, **2006**.

Ring, C.S., Sun, E., McKerrow, J.H., Lee, G.K., Rosenthal, P.J., Kuntz, I.D., Cohen, F.E., Structure-based inhibitor design by using protein models for the development of antiparasitic agents, *Proceedings of the National Academy of Sciences of the United States of America*, 90, 3583-3587, **1993**.

Rishton, G.M., Reactive compounds and in vitro false positives in HTS, *Drug Discovery Today*, 2, 382-384, **1997**.

Rishton, G.M., Nonleadlikeness and leadlikeness in biochemical screening, *Drug Discovery Today*, 8, 86-96, **2003**.

Ritonja, A., Coetzer, T.H.T., Pike, R.N., Dennison, C., The amino acid sequences, structure comparisons and inhibition kinetics of sheep cathepsin L and sheep stefin B, *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology*, 114, 193-198, **1996**.

Roche, O., Schneider, P., Zuegge, J., Guba, W., Kansy, M., Alanine, A., Bleicher, K., Danel, F., Gutknecht, E.M., Rogers-Evans, M., Neidhart, W., Stalder, H., Dillon, M., Sjogren, E., Fotouhi, N., Gillespie, P., Goodnow, R., Harris, W., Jones, P., Taniguchi, M., Tsujii, S., der Saal, W., Zimmermann, G., Schneider, G., Development of a virtual screening method for identification of "frequent hitters" in compound libraries, *Journal of Medicinal Chemistry*, 45, 137-142, **2002**.

Rodrigues, C.R., Flaherty, T.M., Springer, C., McKerrow, J.H., Cohen, F.E., CoMFA and HQSAR of acylhydrazide cruzain inhibitors, *Bioorganic & Medicinal Chemistry Letters*, 12, 1537-1541, **2002**.

Roush, W.R., Gonzalez, F.V., McKerrow, J.H., Hansell, E., Design and synthesis of dipeptidyl alpha ',beta '-epoxy ketones, potent irreversible inhibitors of the cysteine protease cruzain, *Bioorganic & Medicinal Chemistry Letters*, 8, 2809-2812, **1998a**.

Roush, W.R., Gwaltney, S.L., Cheng, J.M., Scheidt, K.A., McKerrow, J.H., Hansell, E., Vinyl sulfonate esters and vinyl sulfonamides: Potent, irreversible inhibitors of cysteine proteases, *Journal of the American Chemical Society*, 120, 10994-10995, **1998b**.

Roush, W.R., Hernandez, A.A., McKerrow, J.H., Selzer, P.M., Hansell, E., Engel, J.C., Design, synthesis and evaluation of D-homophenylalanyl epoxysuccinate inhibitors of the trypanosomal cysteine protease cruzain, *Tetrahedron*, 56, 9747-9762, **2000**.

Ryan, A.J., Gray, N.M., Lowe, P.N., Chung, C.W., Effect of detergent on "promiscuous" inhibitors, *Journal of Medicinal Chemistry*, 46, 3448-3451, **2003**.

Ryckaert, J.P., Ciccotti, G., Berendsen, H.J.C., Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of n-alkanes, *Journal of Computational Physics*, 23, 327-341, **1977**.

Sadowski, J., Gasteiger, J., From atoms and bonds to 3-dimensional atomic coordinates - Automatic model builders, *Chemical Reviews*, 93, 2567-2581, **1993**.

Sadowski, J., Kubinyi, H., A scoring scheme for discriminating between drugs and nondrugs, *Journal of Medicinal Chemistry*, 41, 3325-3329, **1998**.

Scharfstein, J., Schechter, M., Senna, M., Peralta, J.M., Mendoncapreviato, L., Miles, M.A., Trypanosoma-cruzi - characterization and isolation of a 57/51,000 MW surface glycoprotein (GP57/51) expressed by epimastigotes and blood-stream trypomastigotes, *Journal of Immunology*, 137, 1336-1341, **1986**.

Scheidt, K.A., Roush, W.R., McKerrow, J.H., Selzer, P.M., Hansell, E., Rosenthal, P.J., Structure-based design, synthesis and evaluation of conformationally constrained cysteine protease inhibitors, *Bioorganic & Medicinal Chemistry*, 6, 2477-2494, **1998**.

Schmitt, S., Hendlich, M., Klebe, G., From structure to function: A new approach to detect functional similarity among proteins independent from sequence and fold homology, *Angewandte Chemie-International Edition*, 40, 3141-+, **2001**.

Schmitt, S., Kuhn, D., Klebe, G., A new method to detect related function among proteins independent of sequence and fold homology, *Journal of Molecular Biology*, 323, 387-406, **2002**.

Schulz-Gasch, T., Stahl, M., Binding site characteristics in structure-based virtual screening: evaluation of current docking tools, *Journal of Molecular Modeling*, 9, 47-57, **2003** .

Scotti, M. T., Estudo de propriedades físico-químicas e de critérios para obtenção e validação de modelos QSAR para séries de análogos de semicarbazonas com atividade antichagásica, retiradas da literatura, IQ - USP, **2005**.

Seidler, J., McGovern, S.L., Doman, T.N., Shoichet, B.K., Identification and prediction of promiscuous aggregating inhibitors among known drugs, *Journal of Medicinal Chemistry*, 46, 4477-4486, **2003**.

Selzer, P.M., Chen, X.W., Chan, V.J., Cheng, M.S., Kenyon, G.L., Kuntz, I.D., Sakanari, J.A., Cohen, F.E., McKerrow, J.H., Leishmania major: Molecular modeling of cysteine proteases and prediction of new nonpeptide inhibitors, *Experimental Parasitology*, 87, 212-221, **1997**.

Serveau, C., Lalmanach, G., Hirata, I., Scharfstein, J., Juliano, M.A., Gauthier, F., Discrimination of cruzipain, the major cysteine proteinase of Trypanosoma cruzi, and mammalian cathepsins B and L, by a pH-inducible fluorogenic substrate of trypanosomal cysteine proteinases, *European Journal of Biochemistry*, 259, 275-280, **1999**.

Serveau, C., Lalmanach, G., Juliano, M.A., Scharfstein, J., Juliano, L., Gauthier, F., Investigation of the substrate specificity of cruzipain, the major cysteine proteinase of Trypanosoma cruzi, through the use of cystatin-derived substrates and inhibitors, *Biochemical Journal*, 313, 951-956, **1996**.

Sherman, W., Day, T., Jacobson, M.P., Friesner, R.A., Farid, R., Novel procedure for modeling ligand/receptor induced fit effects, *Journal of Medicinal Chemistry*, 49, 534-553, **2006**.

Shoichet, B.K., Virtual screening of chemical libraries, *Nature*, 432, 862-865, **2004**.

Shoichet, B.K., Screening in a spirit haunted world, *Drug Discovery Today*, 11, 607-615, **2006**.

Siqueira, L. J. A., Relação quantitativa entre a estrutura química e o bloqueio da transmissào neuromuscular para série de brometos de [2-(4-benzamido)etil)]benzildimetilamônio para-substituídos., Tese de Mestrado, IQ-USP, **2001**.

Sittampalam, G.S., Kahl, S.D., Janzen, W.P., High-throughput screening: advances in assay technologies, *Current Opinion in Chemical Biology*, 1, 384-391, **1997**.

Sotriffer, C.A., Molecular Dynamics Simulations In Drug Design, *Não Publcado*, **2005**.

Sotriffer, C.A., Klebe, G., Identification and mapping of small-molecule binding sites in proteins: computational tools for structure-based drug design, *Farmaco*, 57, 243-251, **2002**.

Sotriffer, C.A., Kramer, O., Klebe, G., Probing flexibility and "induced-fit" phenomena in aldose reductase by comparative crystal structure analysis and molecular dynamics simulations, *Proteins-Structure Function and Bioinformatics*, 56, 52-66, **2004**.

Sousa, W. A., Estudo da relação entre a estrutura química e a toxicidade aguda de série de cloretos de [2-(benzoiloxi) etil]-dimetilamônio meta-substituídas, com atividade anestésica local, Tese de Mestrado, IQ-USP, **1997**.

Souza, D.H.F., Garratt, R.C., Araujo, A.P.U., Guimaraes, B.G., Jesus, W.D.P., Michels, P.A.M., Hannaert, V., Oliva, G., Trypanosoma cruzi glycosomal glyceraldehyde-3-phosphate dehydrogenase: structure, catalytic mechanism and targeted inhibitor design, *Febs Letters*, 424, 131-135, **1998**.

Steuber, H., Zentgraf, M., Gerlach, C., Sotriffer, C.A., Heine, A., Klebe, G., Expect the unexpected or caveat for drug designers: Multiple structure determinations using aldose reductase crystals treated under varying soaking and co-crystallisation conditions, *Journal of Molecular Biology*, 363, 174-187, **2006**.

Sybyl, Version 7.0. Saint Louis (USA): Tripos Inc; Copyright 1991-2004, **2004**

Tavares, L. C., Efeitos de substituintes sobre a polaridade da carbonila, em ésteres de ácidos benzóicos orto-substituídos, com atividade anestésica local, Tese de Mestrado, FCF-USP, **1987**.

Tavares, L. C., Relações quantitativas entre a estrutura química e a atividade antimicrobiana de análogos à nifuroxazida, Tese de Doutorado, FCF-USP, **1993**.

Tavares, L., Amaral, A.T., Efeito eletrônico de substituintes sobre a polaridade do grupo carbonila e avaliação da atividade anestésica local de N-[(dimetilamino)metil]benzamidas 4-substituídas, *Rev Farm Bioquim.Univ.S.Paulo*, 33, 123-129, **1997** .

Taylor, R., Life-science applications of the Cambridge Structural Database, *Acta Crystallographica Section D-Biological Crystallography*, 58, 879-888, **2002**.

Tocris, Cookson Inc. 16144 Westwoods Business Park Ellisville MO 63021 USA, **2005**

Triballeau, N., H. O. Bertrand, and F. Acher, Are You Sure You Have a Good Model?, in *Pharmacophores and Pharmacophore Searches*, edited by T. Langer and R. D. Hoffmann, pp. 325-326, Wiley-VCH, Weinheim,**2006**.

Troeberg, L., Chen, X.W., Flaherty, T.M., Morty, R.E., Cheng, M.S., Hua, H.M., Springer, C., McKerrow, J.H., Kenyon, G.L., Lonsdale-Eccles, J.D., Coetzer, T.H.T., Cohen, F.E., Chalcone, acyl hydrazide, and related amides kill cultured Trypanosoma brucei brucei, *Molecular Medicine*, 6, 660-669, **2000**.

Urbina, J.A., Specific treatment of Chagas disease: current status and new developments, *Current Opinion in Infectious Diseases*, 14, 733-741, **2001**.

Vasilyev, V., Bliznyuk, A., Application of semiempirical quantum chemical methods as a scoring function in docking, *Theoretical Chemistry Accounts*, 112, 313-317, **2004**.

Venhorst, J., ter Laak, A.M., Commandeur, J.N.M., Funae, Y., Hiroi, T., Vermeulen, N.P.E., Homology modeling of rat and human cytochrome P450 2D (CYP2D) isoforms and computational rationalization of experimental ligand-binding specificities, *Journal of Medicinal Chemistry*, 46, 74-86, **2003**.

Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W., Taylor, R.D., Improved protein-ligand docking using GOLD, *Proteins-Structure Function and Genetics*, 52, 609-623, **2003**.

Verli, H., Barreiro, E.J., A medicinal chemistry paradigm: Ligands and receptor flexibility, *Quimica Nova*, 28, 95-102, **2005**.

Vicik, R., Busemann, M., Baumann, K., Schirmeister, T., Inhibitors of cysteine proteases, *Current Topics in Medicinal Chemistry*, 6, 331-353, **2006a**.

Vicik, R., Hoerr, V., Glaser, M., Schultheis, M., Hansell, E., McKerrow, J.H., Holzgrabe, U., Caffrey, C.R., Ponte-Sucre, A., Moll, H., Stich, A., Schirmeister, T., Aziridine-2,3-dicarboxylate inhibitors targeting the major cysteine protease of Trypanosoma brucei as lead trypanocidal agents, *Bioorganic & Medicinal Chemistry Letters*, 16, 2753-2757, **2006b**.

Viegas, C.; Bolzani, V.D.; Barreiro, E.J. The Natural Products and the Modern Medicinal Chemistry. *Quimica Nova*, 29, 326-337, **2006**

von Ahsen, O., Voigtmann, U., Klotz, M., Nifantiev, N., Schottelius, A., Ernst, A., Muller-Tiemann, B., Parczyk, K., A miniaturized high-throughput screening assay for fucosyltransferase VII, *Analytical Biochemistry*, 372, 96-105, **2008**.

Wade, R.C., Clark, K.J., Goodford, P.J., Further development of hydrogen-bond functions for use in determining energetically favorable binding-sites on molecules of known structure .1. Ligand probe groups with the ability to form 2 hydrogen-bonds, *Journal of Medicinal Chemistry*, 36, 140-147, **1993**.

Walters, W.P., Murcko, M.A., Prediction of 'drug-likeness', *Advanced Drug Delivery Reviews*, 54, 255-271, **2002**.

Wang, C.C., Validating targets for antiparasite chemotherapy, *Parasitology*, 114, S31-S44, **1997**.

Wang, R.X., Fang, X.L., Lu, Y.P., Wang, S.M., The PDBbind database: Collection of binding affinities for protein-ligand complexes with known three-dimensional structures, *Journal of Medicinal Chemistry*, 47, 2977-2980, **2004a**.

Wang, R.X., Fu, Y., Lai, L.H., A new atom-additive method for calculating partition coefficients, *Journal of Chemical Information and Computer Sciences*, 37, 615-621, **1997**.

Wang, R.X., Liu, L., Lai, L.H., Tang, Y.Q., SCORE: A new empirical method for estimating the binding affinity of a protein-ligand complex, *Journal of Molecular Modeling*, 4, 379-394, **1998**.

Wang, R.X., Lu, Y.P., Fang, X.L., Wang, S.M., An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes, *Journal of Chemical Information and Computer Sciences* , 44, 2114-2125, **2004b**.

Wang, R.X., Wang, S.M., How does consensus scoring work for virtual library screening? An idealized computer experiment, *Journal of Chemical Information and Computer Sciences*, 41, 1422-1426, **2001**.

Wang, S. H., DBfilter, **1-6-2003**

WEBER, P.C., WENDOLOSKI, J.J., PANTOLIANO, M.W., SALEMME, F.R., CRYSTALLOGRAPHIC AND THERMODYNAMIC COMPARISON OF NATURAL AND SYNTHETIC LIGANDS BOUND TO STREPTAVIDIN, *Journal of the American Chemical Society*, 114, 3197-3200, **1992**.

Werbovetz, K.A., Target-based drug discovery for malaria, leishmaniasis, and trypanosomiasis, *Current Medicinal Chemistry*, 7, 835-860, **2000**.

Wermuth, G., Ganellin, C.R., Lindberg, P., Mitscher, L.A., Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998), *Pure and Applied Chemistry*, 70, 1129-1143, **1998**.

Wess, G., Urmann, M., Sickenberger, B., Medicinal chemistry: Challenges and opportunities, *Angewandte Chemie-International Edition*, 40, 3341-+, **2001**.

Westbrook, J., Feng, Z.K., Chen, L., Yang, H.W., Berman, H.M., The Protein Data Bank and structural genomics, *Nucleic Acids Research*, 31, 489-491, **2003**.

WHO, The World Health Report, **2002**

Williams, D.H., Bardsley, B., Estimating binding constants - The hydrophobic effect and cooperativity, *Perspectives in Drug Discovery and Design*, 17, 43-59, **1999**.

Williard, X., Pop, I., Bourel, L., Horvath, D., Baudelle, R., Melnyk, P., Deprez, B., Tartar, A., Combinatorial chemistry: A rational approach to chemical diversity, *European Journal of Medicinal Chemistry*, 31, 87-98, **1996**.

Wolber, G., Langer, T., LigandScout: 3-d pharmacophores derived from protein-bound Ligands and their use as virtual screening filters, *Journal of Chemical Information and Modeling*, 45, 160-169, **2005**.

Wood, W.J.L., Huang, L., Ellman, J.A., Synthesis of a diverse library of mechanism-based cysteine protease inhibitors, *Journal of Combinatorial Chemistry*, 5, 869-880, **2003**.

Yong, V., Schmitz, V., Vannier-Santos, M.A., Lima, A.P.C.D., Lalmanach, G., Juliano, L., Gauthier, F., Scharfstein, J., Altered expression of cruzipain and a cathepsin B-like target in a Trypanosoma cruzi cell line displaying resistance to synthetic inhibitors of cysteine-proteinases, *Molecular and Biochemical Parasitology*, 109, 47-59, **2000**.

Zhou, Z.G., Madura, J.D., CoMFA 3D-QSAR analysis of HIV-1 RT nonnucleoside inhibitors, TIBO derivatives based on docking conformation and alignment, *Journal of Chemical Information and Computer Sciences*, 44, 2167-2178, **2004**.

Zidek, L., Novotny, M.V., Stone, M.J., Increased protein backbone conformational entropy upon hydrophobic ligand binding, *Nature Structural Biology*, 6, 1118-1121, **1999**.