

**UNIVERSIDADE FEDERAL DE SÃO CARLOS**  
**DEPARTAMENTO DE QUÍMICA**

**Dissertação de Mestrado**

**Planejamento de inibidores da cruzaina baseado em  
fragmentos**

*EMANUELLA MARIA BARRETO FONSECA*

Orientador: Prof. Dr. Carlos Alberto Montanari

**São Carlos - SP**  
**2011**

**UNIVERSIDADE FEDERAL DE SÃO CARLOS**  
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA  
DEPARTAMENTO DE QUÍMICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

**Planejamento de inibidores da cruzaina baseado em  
fragmentos**

EMANUELLA MARIA BARRETO FONSECA\*

Dissertação apresentada como parte  
dos requisitos para obtenção do título  
de MESTRE EM QUÍMICA, área de  
concentração: QUÍMICA ORGÂNICA.

Orientador: Prof. Dr. Carlos Alberto Montanari

\* bolsista CAPES

**São Carlos - SP**  
**2011**

**Ficha catalográfica elaborada pelo DePT da  
Biblioteca Comunitária da UFSCar**

F676pi

Fonseca, Emanuella Maria Barreto.  
Planejamento de inibidores da cruzaina baseado em  
fragmentos / Emanuella Maria Barreto Fonseca. -- São  
Carlos : UFSCar, 2011.  
75 f.

Dissertação (Mestrado) -- Universidade Federal de São  
Carlos, 2011.

1. Química orgânica. 2. Chagas, Doença de. 3.  
Modelagem molecular. 4. Trypanosoma Cruzi. I. Título.

CDD: 547 (20<sup>a</sup>)

**UNIVERSIDADE FEDERAL DE SÃO CARLOS**  
*Centro de Ciências Exatas e de Tecnologia*  
*Departamento de Química*  
**PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**  
**Curso de Mestrado Acadêmico**

---

*Assinaturas dos membros da banca examinadora que avaliaram e aprovaram a defesa de dissertação de mestrado acadêmico da candidata **Emanuella Maria Barreto Fonseca**, realizada em 27 de julho de 2011:*



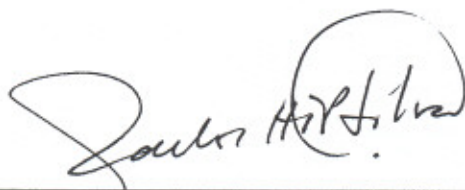
---

Prof. Dr. Carlos Alberto Montanari



---

Prof. Dr. João Batista Fernandes



---

Prof. Dr. Carlos Henrique Tomich de Paula da Silva



## DEDICATÓRIA

A minha mãe Graça Barreto, a minha irmã Raphaella Fonseca e meu  
cunhado Pedro Vasconcelos

*“O sonho encheu a noite  
Extravasou pro meu dia  
Encheu minha vida  
E é dele que eu vou viver  
Porque sonho não morre.”*

*(Adélia Prado)*

## **AGRADECIMENTOS**

A Deus pela vida.

A minha família pela educação, pelas oportunidades, pelas orações, pelo amor que me deram suporte em cada nova etapa da minha vida.

Ao Prof. Dr. Carlos Montanari pela orientação, desafios propostos, estímulo e confiança para o desenvolvimento deste trabalho e sua esposa Dra. Maria Luíza Montanari (Malu) pelas conversas e conselhos.

Ao Dr. Peter Kenny pelas conversas e contribuições quando o desafiante projeto começava a dar seus primeiros passos.

Aos amigos e colegas de grupo pela amizade, alegrias e descontrações durante o período de desenvolvimento da dissertação, discussões construtivas, ajuda no planejamento e execução de experimentos e inúmeras atividades alegres realizadas conjuntamente: Dr. Josmar Rocha, Dr. Helton Wiggers, Geraldo Sartori, Dra. Juliana Cheleski, Prof. Dr. Andrei Leitão, Patrícia Leitão, Igor Prokopzyck, Jean Francisco, Ricardo Goulart, Fabyana Soares, Juliane Sempionato e Fabiana Rossini.

Ao Instituto de Química de São Carlos (IQSC) da Universidade de São Paulo (USP) por proporcionar o desenvolvimento do trabalho em suas dependências.

Aos funcionários da Universidade Federal de São Carlos pelo profissionalismo e competência.

As agências de fomento pelo financiamento dos projetos do NEQUIMED-PN e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa.

## LISTA DE ABREVIATURAS

3D	três dimensões ou tridimensional
Å	angstrom
Ala	resíduo de aminoácido alanina
Asn	resíduo de aminoácido asparagina
Asp	resíduo de aminoácido aspartato
BUW	blocos não escalonados em peso (do inglês, Block Unscaled Weights)
CPCA	análise de componentes principais consensual
Cys	resíduo de aminoácido cisteína
DDT	ditiotreitól
DMSO	dimetilsulfóxido
DNA	ácido desoxirribonucléico
EC	comissão para classificação de enzimas
EDTA	Ácido etilenodiamino tetra-acético
FBDD	planejamento de fármacos baseado em fragmentos
FXa	fator Xa
g	grama
Gln	resíduo de aminoácido glutamina
Glu	resíduo de aminoácido glutamato
Gly	resíduo de aminoácido glicina
GPCRs	receptor acoplado à proteína G
HAC	átomos diferentes de hidrogênio
His	resíduo de aminoácido histidina
HTS	ensaios em larga escala
IC <sub>50</sub>	concentração para inibir 50% da atividade enzimática
K	Kelvin
k <sub>B</sub>	constante de Boltzmann
kcal	quilocaloria
k <sub>cat</sub>	constante catalítica
K <sub>M</sub>	constante de Michaelis Menten
LE	eficiência do ligante
Leu	resíduo de aminoácido leucina
LogP	logaritmo do coeficiente de partição octanol-água
MCA	7-amino-4-metilcumarina

DM	dinâmica molecular
Met	resíduo de aminoácido metionina
MIF	campos de interação molecular
mM	milimolar
MW	massa molecular
nM	nanomolar
NpT	simulação com pressão e temperatura constante
ns	nanossegundos
OPLS	simulações de sistemas líquidos por potenciais otimizados (do inglês, Optimized Potentials for Liquid Simulations)
PC1	primeira componente principal
PC2	segunda componente principal
PCA	análise de componentes principais
PDB	banco de dados de proteínas
Pro	resíduo de aminoácido prolina
ps	picossegundos
PSA	área superficial polar
R	constante dos gases
R <sub>factor</sub>	fator residual
R <sub>free</sub>	fator residual livre
RMN	ressonância magnética nuclear
RMS	desvio quadrático médio
RNA	ácido ribonucleico
Ro3	regra dos três
s	segundos
S1 a S4	sub-sítios integrantes do sítio catalítico da enzima cruzaina
SA	anelamento simulado(do inglês, simulated annealing)
SAR	relação estrutura atividade
Ser	resíduo de aminoácido serina
SMARTS	especificação arbitrária do alvo por notação SMILES(do inglês, SMiles Arbitrary Target Specification)
SMILES	Sistema de entrada de linha de dados moleculares simplificado (do inglês, Simplified Molecular Input Line Entry System)
T	temperatura
TBVS	ensaio virtual baseado na estrutura do receptor (alvo)
Tyr	resíduo de aminoácido tirosina

v/v	volume/volume
Val	resíduo de aminoácido valina
ZFRMCA	N-benzóxicarbonil-L-fenilalanina-L-arginina-7-amido-metil-cumarina
$\Delta G$	variação da energia livre de Gibbs
$\Delta V$	variação do volume
$\mu M$	micromolar

## LISTA DE TABELAS

TABELA 1.1 - Funções biológicas de alguns alvos estudados na busca de novos fármacos para a doença de Chagas.....	6
TABELA 4.2 - Código PDB das estruturas estudadas, parâmetros estruturais e grau de identidade.....	28
TABELA 4.3 - Relação dos aminoácidos correspondentes aos pontos relevantes para a seletividade na cruzaina e os aminoácidos correspondentes nas Catepsinas L e B. ....	33
TABELA 5.1 - Pontuação obtida na docagem molecular ( $\text{kcal mol}^{-1}$ ).....	56
TABELA 5.2 - Pontuação obtida por docagem molecular das moléculas selecionadas .....	60

## LISTA DE FIGURAS

FIGURA 1.1 - Rotas de migração da América Latina e estimativa do número total de pessoas infectadas em países não endêmicos <sup>8</sup> .....	4
FIGURA 1.2 – Representação das estruturas moleculares do a) Nifurtimox e b) Benzonidazol.....	5
FIGURA 1.3 - Reação de clivagem de ligação peptídica por proteases.....	7
FIGURA 1.4 – Estrutura cristalográfica da cruzaina (PDB: 2AIM). Em destaque os resíduos que participam do mecanismo catalítico .....	7
FIGURA 1.5 - Mecanismo de catálise de enzimas cisteíno proteases .....	8
FIGURA 1.6 – Sítio da cruzaina representado por superfície de van der Waals – em amarelo está representada a posição da cisteína catalítica (PDB: 1ME4) 8	8
FIGURA 1.7 – Estrutura do K11777 .....	9
FIGURA 2.1 - Etapas do processo de descoberta de fármacos.....	13
FIGURA 2.2 - Demonstração do espaço químico e do espaço químico-biológico .....	14
FIGURA 2.3 - Estágios do processo de desenvolvimento de um fármaco onde os métodos em quiminformática são aplicados.....	15
FIGURA 2.4 - Otimização do fragmento. a) Crescimento do fragmento, pedaço a pedaço. b) Combinação de diferentes fragmentos.....	18
FIGURA 2.5 – Estratégia adotada pra planejamento baseado em fragmentos <sup>54</sup> .....	19
FIGURA 2.6 – Planejamento do LY-517717 <sup>45</sup> .....	20
FIGURA 4.1 – Procedimento usado para construir a matriz X. Iniciando pelos cálculos do GRID para uma sonda, é construído um vetor contendo todas as energias de interação nos k pontos do grid; depois, os vetores para n sondas são compilados em um único vetor n x k. ‘Empilhando’ os vetores, cada alvo corresponderá a uma linha na matriz X final. ....	26
FIGURA 4.2 - Sítio Ativo da Cruzaina .....	29
FIGURA 4.3 - Sub-sítio S2 da cruzaina em diferentes estruturas cristalográficas depositadas no PDB.....	30
FIGURA 4.4 – Sobreposição dos ligantes co-cristalizados com a cruzaina. Em amarelo, destaque para o sub-sítioS2.....	30
FIGURA 4.5 - CPCA score: 1)Catepsina B; 2)Catepsina L; 3)Cruzaina .....	31
FIGURA 4.6 - Contribuição de cada sonda para o modelo gerado: A) C3; B) NM3 e C)OH.....	31
FIGURA 4.7 - Diferença nos pseudo-campos gerados, mostrando as regiões de favorecimento para cada alvo. Sondas: A) C3, B) NM3 e C) OH. (Ciano=Cruzaina e Amarelo=Catepsina L).....	32

FIGURA 4.8 - Diferença nos pseudo-campos gerados, mostrando as regiões de favorecimento para cada alvo. Sondas: A)C3, B) NM3 e C) OH. (Ciano=Cruzaína e Amarelo=Catepsina B).....	32
FIGURA 4.9 - Sobreposição das cisteíno-proteases na extensão do sub-sítio S2: <b>Cruzaína (azul/negrito)</b> , <i>Catepsina L (branco/italico)</i> e Catepsina B (ciano/sublinhado).....	34
FIGURA 5.1 - Ensaio virtual para identificação de novos ligantes .....	38
FIGURA 5.2 - Graus de liberdade conformacionais (setas) de um ligante.....	41
FIGURA 5.3 - Estratégia utilizada para gerar o sinal de fluorescência em reações catalisadas por cisteíno-proteases .....	46
FIGURA 5.4 - Estrutura do ZFRMCA (substrato sintético para a cruzaína) .....	46
FIGURA 5.5 – a) Variação da temperatura durante a DM. b) Etapas de aquecimento e resfriamento durante o anelamento. ....	50
FIGURA 5.6 – Representação do ligante P10, extraído do PDB: 1ME3. Em destaque os grupos reconhecidos pela enzima cruzaína.....	53
FIGURA 5.7 – Exemplo de padrões SMARTS .....	53
FIGURA 5.8 – Sub-sítioS2 da cruzaína: a) sobreposto com o ligante BD3 da catepsina L (PDB 3HWN); b) após a minimização com o mesmo ligante .....	54
FIGURA 5.9 – Posições ocupadas pelos fragmentos selecionados obtidas pela docagem.....	55
FIGURA 5.10 – Modo de interação observado para o fragmento 10 .....	56
FIGURA 5.11 – RMSD dos fragmentos analisados.....	57
FIGURA 5.12 – Modo de interação predito na posição de menor energia durante a dinâmica molecular a) Nequimed147, b) Nequimed182.....	57
FIGURA 5.13 – Padrões SMARTS utilizados para a busca .....	59
FIGURA 5.14 – Estruturas das moléculas análogas ao Nequimed 147 testadas contra a enzima cruzaína. LE = eficiência do ligante kcal mol <sup>-1</sup> átomo pesado <sup>-1</sup> .....	60
FIGURA 5.15 – Estruturas dos novos esqueletos moleculares testados contra a enzima cruzaína. LE = eficiência do ligante kcal mol <sup>-1</sup> átomo pesado <sup>-1</sup> .....	61



## RESUMO

PLANEJAMENTO DE INIBIDORES DA CRUZAÍNA BASEADO EM FRAGMENTOS - A Doença de Chagas, descrita em 1909 pelo médico sanitário brasileiro Dr. Carlos Chagas, é causada pelo parasito tripanossomatídeo *Trypanosoma cruzi*. Os tratamentos atuais consistem no uso dos fármacos benzonidazol e nifurtimox que são eficazes apenas no estágio inicial da doença (fase aguda), mas possuem efeitos colaterais severos. A enzima cruzaina constitui um alvo validado contra a doença de Chagas e sua inibição vem sendo estudada para o planejamento de fármacos. Neste trabalho a estratégia adotada foi o planejamento baseado em fragmentos moleculares, através de uma abordagem computacional para a seleção dos mesmos. Utilizando a técnica GRID/CPCA, foram identificados resíduos importantes, como Ser207 e Asp69, em uma região adjacente ao sub-sítio S2 da enzima. Estes resíduos são capazes de realizar interações do tipo ligação de hidrogênio e ainda não foram explorados na busca de inibidores para esse alvo. Além disso, foram identificados padrões de reconhecimento molecular da enzima, baseado nas estruturas depositadas no *Protein Data Bank* com inibidores co-cristalizados no sítio da enzima, que foram utilizados para fazer uma busca guiada com a utilização de padrão SMARTS para construção de um banco de fragmentos. As técnicas de docagem e dinâmica molecular foram usadas para selecionar os fragmentos, os quais foram adquiridos para os ensaios *in vitro*. Vários fragmentos foram testados contra a enzima cruzaina utilizando a técnica de fluorimetria, sendo que um deles (Nequimed147) mostrou-se capaz de inibir a enzima com IC<sub>50</sub> igual a 2,39 mM, resultando em uma eficiência do ligante de 0,23 kcal mol<sup>-1</sup> átomo pesado<sup>-1</sup>. Baseado nesse fragmento, novos inibidores foram identificados com potência entre 1,25 e 4,29 mM e eficiência do ligante entre 0,21 e 0,26 kcal mol<sup>-1</sup> átomo pesado<sup>-1</sup>, com características apropriadas para o reconhecimento molecular pela enzima cruzaina na região do sub-sítio S2.

## ABSTRACT

FRAGMENT BASED DRUG DESIGN OF CRUZAIN INHIBITORS - The Chagas disease, described in 1909 by the Brazilian sanitary physician, scientist and bacteriologist Dr. Carlos Chagas, is caused by the *Trypanosomatid* parasite *Trypanosoma cruzi*. Current treatments involve the use of drugs benznidazole and nifurtimox, which are effective only in the early stage of the disease (acute phase), but have severe side effects. The enzyme cruzain is a validated target for searching small molecules to act against Chagas disease. Cruzain inhibition has been studied for the discovery and development of new drugs. In this work, the planning strategy was based on molecular fragments, using a computational approach for selecting them. Using GRID/CPCA, important residues such as Ser207 and Asp69 were identified in an adjacent cavity to the S2 sub-site of the enzyme. These residues can take part on hydrogen-bond interactions and have not been explored in the search for inhibitors for this target. In addition, we identified patterns of molecular recognition of the enzyme, based on the structures deposited in the Protein Data Bank, which are co-crystallized with inhibitors at the site of the enzyme. These inhibitors were used to guide the search via SMARTS pattern to build a fragment database. Docking and molecular dynamics were used to select the fragments for an *in vitro* assay. Many fragments were acquired and tested against cruzain using the technique of fluorescence spectroscopy. One fragment (Nequimed147) showed activity, inhibiting the enzyme with  $IC_{50}$  equal to 2.39 mM, resulting in a ligand efficiency of  $0.23 \text{ kcal mol}^{-1} \text{ weight atom}^{-1}$ . Based on this fragment, new inhibitors were identified with  $IC_{50}$  between 1.25 and 4.29 mM. Their ligand efficiencies were in the range of 0.21 and  $0.26 \text{ kcal mol}^{-1} \text{ weight atom}^{-1}$  whose molecular recognition features are suitable for the search of new enzyme inhibitors likely to bind the cavity near the S2 sub-site that has not been explored yet.

## Sumário

CAPÍTULO 1 .....	1
1 -Introdução .....	2
1.1 -Doença de Chagas .....	3
1.2 -Alvo Enzimático: Cruzaína.....	6
CAPÍTULO 2 .....	11
2 -Química Medicinal e o Planejamento de Fármacos .....	12
2.1 -Planejamento de fármacos .....	12
2.2 -Planejamento de fármacos baseado em fragmentos.....	16
CAPÍTULO 3 .....	22
3 -Objetivos .....	23
CAPÍTULO 4 .....	24
4 -Inspeção do sítio usando campos de interação molecular.....	25
4.1 -Materiais e Métodos.....	25
4.2 -Resultados e Discussão.....	27
CAPÍTULO 5 .....	35
5 -Identificação e Seleção dos Fragmentos .....	36
5.1 -Métodos Computacionais.....	36
5.1.1 -Docagem Molecular .....	36
5.1.1.1 -Funções de pontuação.....	39
5.1.1.1.1 -As funções baseadas em campo de força .....	39
5.1.1.1.2 -As funções empíricas.....	40
5.1.1.1.3 -As funções baseadas no conhecimento.....	40
5.1.1.2 -Algoritmos de busca e flexibilidade molecular .....	41
5.1.1.2.1 -Algoritmo de busca sistemática ou construção incremental.....	41
5.1.1.2.2 -Algoritmos estocásticos .....	42

5.1.2 -Simulação de Dinâmica Molecular .....	43
5.2 -Ensaio de inibição .....	45
5.3 -Materiais e métodos.....	47
5.4 -Resultados e Discussão.....	52
5.5 -Otimização do fragmento .....	58
CAPÍTULO 6 .....	63
6 -Conclusão e Perspectiva.....	64
6.1 -Conclusões .....	64
6.2 -Perspectivas .....	64
CAPÍTULO 7 .....	66
7 - Referência Bibliográfica.....	67
Apêndice .....	71

# **CAPÍTULO 1**

## **Introdução**

Neste capítulo é apresentada uma breve introdução sobre a doença de Chagas, seu panorama atual, ocorrência geográfica, manifestações clínicas e o tratamento usado. Também são abordados aspectos importantes da enzima cruzaina, que é o alvo enzimático utilizado nesse trabalho.

## 1 - Introdução

No início de 1909, um jovem e desconhecido médico brasileiro, Carlos Ribeiro Justiniano das Chagas, foi enviado para investigar um surto de malária que interrompeu a construção de uma ferrovia importante no estado de Minas Gerais. Lá ele se deparou com uma doença nova e estranha, conhecida pelos habitantes locais como *'baticum'*, uma palavra onomatopeica para as palpitações decorrentes da condição cardíaca característica da doença. Um dos engenheiros da ferrovia chamou a atenção de Chagas para o grande número de triatomíneos na área. Estes eram conhecidos localmente como barbeiro devido ao seu hábito de morder os seres humanos na face. Quando Chagas examinou o intestino dos insetos, ele encontrou as formas tripomastigotas de *Trypanosoma*, agente causador da tripanossomíase americana. Poucas semanas após a descoberta do *Trypanosoma*, ele examinou uma menina que tinha apresentado com um estranho inchaço de uma pálpebra, febre e mal estar. Uma gota de seu sangue revelou os mesmos tripanossomos encontrados nos insetos. Com este primeiro caso humano identificado, Chagas completou um ciclo de trabalho extraordinário, sem precedentes na história da medicina até agora: ele descobriu uma nova doença, identificada pelo seu agente e seu mecanismo de transmissão dentro do espaço de poucas semanas.<sup>1</sup> Pouco mais de cem anos depois da descoberta, a doença de Chagas continua essencialmente incurável.<sup>2</sup> Ainda assim, as doenças tropicais como a malária, a doença de Chagas, a doença do sono, entre outras, continuam sendo uma das principais causas de morbidade e mortalidade em todo mundo, sendo um problema global de saúde pública. Essas doenças, juntamente com a tuberculose, incapacitam ou matam milhões de pessoas e, apesar disso, do total de 1.556 medicamentos registrados entre 1975 e 2004, apenas 21 foram desenvolvidos para elas.<sup>3-4</sup>

Estudo recente sobre o financiamento mundial de inovação para doenças negligenciadas mostrou que menos de 5% foram investidos nas doenças extremamente negligenciadas, mesmo que 500 milhões de pessoas estejam ameaçadas por elas em todo o mundo.<sup>3-4</sup> A indústria farmacêutica teve pouco interesse no desenvolvimento de novos fármacos antichagásicos, devido principalmente à falta de incentivos comerciais.<sup>2</sup> O desinteresse na produção desse tipo de medicamento, considerados como “medicamentos sociais”, está ligado à sua

baixa demanda, restrita principalmente à população de baixa renda de alguns países na América Latina, portanto, com pequena margem de lucro para a indústria.<sup>5</sup> Como resposta a essa necessidade, metodologias baseadas em modelagem molecular, capazes de manipular grandes quantidades de informações utilizando ferramentas em bio- e quiminformática, têm sido desenvolvidas para auxiliar na busca de potenciais novos fármacos.

## 1.1 - Doença de Chagas

A doença de Chagas é um dos principais problemas socioeconômicos enfrentados na América Latina. Estima-se que 16 a 18 milhões de pessoas do sul dos Estados Unidos até a Patagônia estejam infectadas, causando aproximadamente 50.000 mortes por ano, sendo que outros 100 milhões de indivíduos vivem em áreas de risco de contaminação.<sup>2, 6</sup> (Dados da Organização Mundial da Saúde de 2007 citado por Man et al., 2008<sup>2, 6</sup>) O protozoário flagelado *T. cruzi*, da ordem cinetoplastida, família *Trypanosomatidae* e gênero *Trypanosoma* é o agente causador desta doença. O parasita possui um complexo ciclo biológico, o qual envolve um hospedeiro invertebrado e outro vertebrado, e apresenta três formas distintas: (i) epimastigota, forma presente no vetor e em cultura axênica; (ii) tripomastigota, forma sanguínea circulante e infectante; e (iii) amastigota, forma de replicação intracelular.<sup>6</sup> A contaminação se dá principalmente pelo contato com as fezes do inseto vetor (barbeiro), podendo também ocorrer por transfusão de sangue, transplante de órgão, ingestão de alimentos contaminados e transmissão congênita. O vetor pertence a uma família com aproximadamente 130 espécies, sendo que cerca de doze podem transmitir o *Trypanosoma*.<sup>1, 7</sup> Endêmica na América Latina, há a ocorrência da doença, devido aos processos migratórios, em locais como: Estados Unidos (300.000 pessoas infectadas), Austrália (1.500 pessoas infectadas), Japão (3.000 pessoas infectadas), Europa (80.000 pessoas infectadas), Canadá (5.500 pessoas infectadas) – FIGURA 1.1.<sup>8</sup>

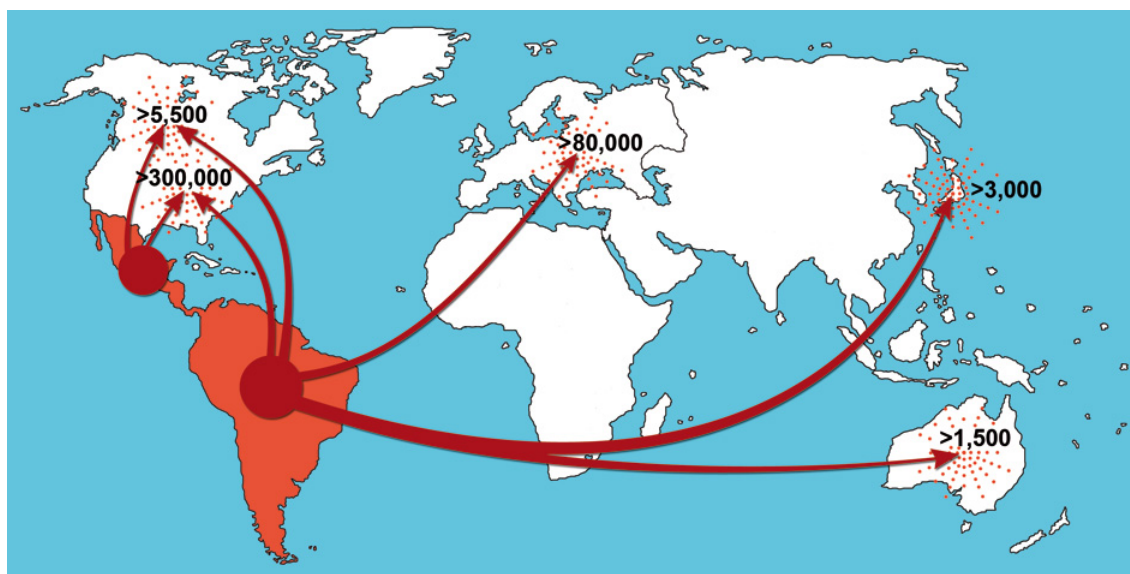


FIGURA 1.1 - Rotas de migração da América Latina e estimativa do número total de pessoas infectadas em países não endêmicos<sup>8</sup>

A doença de Chagas apresenta dois estágios clínicos: agudo (que dura em média dois meses) e crônico (que se prolonga por toda a vida do hospedeiro)<sup>1, 5-6</sup>. A fase aguda é geralmente uma doença das crianças, podendo também ocorrer em qualquer idade, e levar à morte devido a complicações decorrentes de insuficiência cardíaca e processos inflamatórios que envolvem o cérebro, como a meningoencefalite.<sup>3-6</sup> Esse estágio se inicia após um inseto infectado picar uma pessoa e defecar no local. Algumas horas depois, o local da picada produz um inchaço característico chamado de "chagoma". Ocorre também febre e taquicardia. A fase de infecção aguda normalmente não é fatal e, muitas vezes, passa despercebida.<sup>1, 5</sup>

O estágio crônico se divide em dois: sintomático e assintomático. Cerca de 60 % dos pacientes chagásicos crônicos permanecem assintomáticos por longos períodos, sendo que 30% destes desenvolvem complicações cardíacas, que podem levar à cardiopatia chagásica crônica. A doença de Chagas crônica causa deficiências significativas, tendo alto impacto social e econômico, incluindo desemprego e diminuição da capacidade produtiva. No Brasil, mais de US\$ 1,3 bilhões em salários e produtividade industrial foram perdidos devido a trabalhadores contaminados.<sup>9</sup>

A prevenção da doença tem sido feita combatendo seu vetor com inseticidas e através de medidas preventivas pessoais, como o uso de mosquiteiros, boas práticas de higiene na preparação, estocagem e consumo de



alimentos, controle do sangue e de órgãos usados para transfusão e transplantes, respectivamente.<sup>5</sup> Os tratamentos atuais são eficientes apenas no início da infecção e não apresentam doses pediátricas.<sup>10-11</sup> Surgidos na década de 1970, os dois fármacos disponíveis são: nifurtimox (Lampit®, da Bayer) e benzonidazol (Rochagan®, da Roche) – FIGURA 1.2. Estes medicamentos são eficazes somente na fase aguda, devem ser administrados por um longo período e têm sua ação afetada diretamente por condições como a idade do paciente. Além disso, apresentam significativos efeitos colaterais.<sup>2, 5-6, 10-12</sup> Não há tratamento para o estágio crônico.<sup>10-11</sup>

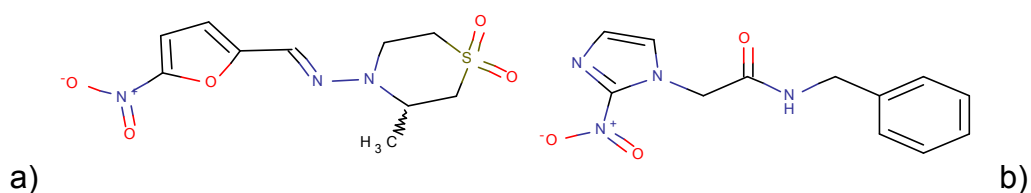


FIGURA 1.2 – Representação das estruturas moleculares do a) Nifurtimox e b) Benzonidazol

O Benzonidazol pode causar edema, febre, erupção cutânea, dores articulares e musculares, neuropatia periférica e linfadenopatia. Da mesma forma, o Nifurtimox pode provocar a perda de peso, erupção cutânea, psicose, náuseas e vômitos, leucopenia, neurotoxicidade, neuropatia periférica e anormalidades do tecido (das glândulas supra-renais e mamárias, do cólon, esôfago, testículos, ovários). Estudos retrospectivos na América Latina revelaram que, embora ambos fármacos apresentem boa atividade, durante o tratamento de casos recentes de infecção aguda e crônica, atinge um índice médio de cura parasitológica em torno de 60%. Na fase crônica (pacientes com mais de 10 anos de infecção), a cura parasitológica é obtida somente em 10-20% dos pacientes.<sup>12</sup> Assim, se fazem necessários estudos para obtenção de uma fórmula acessível financeiramente, eficiente para tratar pacientes no início da infecção; e medicamentos novos para tratamento seguro da fase crônica.<sup>9</sup>

Dentre as estratégias adotadas para o tratamento de várias doenças está a inibição de alvos enzimáticos. Além dos avanços significativos nos aspectos biológicos, genéticos e evolucionários do parasito, o sequenciamento de seu genoma contribuiu com a identificação de diferentes alvos biológicos promissores, sendo a maior parte de enzimas. Na TABELA 1.1 são apresentados alguns desses alvos e suas funções biológicas.<sup>6</sup>

TABELA 1.1 - Funções biológicas de alguns alvos estudados na busca de novos fármacos para a doença de Chagas

<b>Enzima</b>	<b>Via ou classe metabólica</b>
<b>Cisteíno proteases, Serino proteases, metaloproteínas, Treonino proteases</b>	Proteases – Múltiplas funções que envolvem desde invasão celular até o escape do parasita do sistema imune.
<b>Esterol 14-demetilase, Lanosterol sintase, Esqualeno epoxidase, Esqualeno sintase</b>	Biossíntese de esteróis – Essencial para a composição estrutural de membranas, mitocôndrias e plasma.
<b>Gliceraldeído-3-fosfato desidrogenase, Hexoquinase, Fosfofrutoquinase</b>	Via glicolítica – Produção de energia
<b>Tripanotiona redutase, Tripanotiona sintase</b>	Metabolismo dependente de grupos tióis – Mecanismo de defesa contra estresse oxidativo
<b>6-Fosfogluconato desidrogenase</b>	Metabolismo de pentose fosfato – Mecanismo de defesa contra estresse oxidativo

## 1.2 - Alvo Enzimático: Cruzaína

Inibidores da protease provaram ser medicamentos eficazes para o diabetes, osteoporose, certos tipos de câncer e doenças infecciosas.<sup>12</sup> As cisteíno proteases são enzimas proteolíticas que dependem do grupo tiol nucleofílico de um resíduo de cisteína para a sua atividade enzimática.<sup>13</sup>

A Cruzaína, uma cisteíno protease chave, é a principal protease do *Trypanosoma cruzi*, sendo essencial para a sobrevivência do parasita. Essa enzima está envolvida em todas as fases do ciclo de vida do *T. Cruzi*, e pertence à superfamília papaína de cisteíno proteases, com estrutura e funções similares às enzimas humanas catepsina B, L, K, S, F e V.<sup>12, 14</sup> Ela está envolvida na replicação, diferenciação celular do parasita e nutrição com a hidrólise de proteínas do hospedeiro. Ela é essencial para a infecção das células hospedeiras, metabolismo e imunoevasão (evasão da resposta imune).<sup>13, 15-23</sup>

A cruzaína catalisa reações de hidrólise de ligações peptídicas (FIGURA 1.3), e seu número de classificação segundo a comissão de nomenclatura de enzimas (*Enzyme Commission*) é EC 3.4.22.51. Os números 3, 4 e 22

correspondem respectivamente à classe de hidrolases, à sua atuação em ligação peptídica e ao grupo de cisteíno endopeptidases enquanto que 51 é o número específico da cruzaina.



FIGURA 1.3 - Reação de clivagem de ligação peptídica por proteases

A estrutura cristalográfica da cruzaina revela que ela é composta por uma cadeia polipeptídica de 215 resíduos de aminoácidos, dobrado em dois domínios (FIGURA 1.4). Um domínio é principalmente helicoidal (domínio L), e o outro é composto por folhas  $\beta$  antiparalelas (domínio R). Como esperado para as enzimas da família da papaína, entre os dois domínios existe uma extensa fenda, onde está localizado o sítio ativo, sendo a tríade catalítica composta pelos resíduos Cys25, His159 e Asn175.

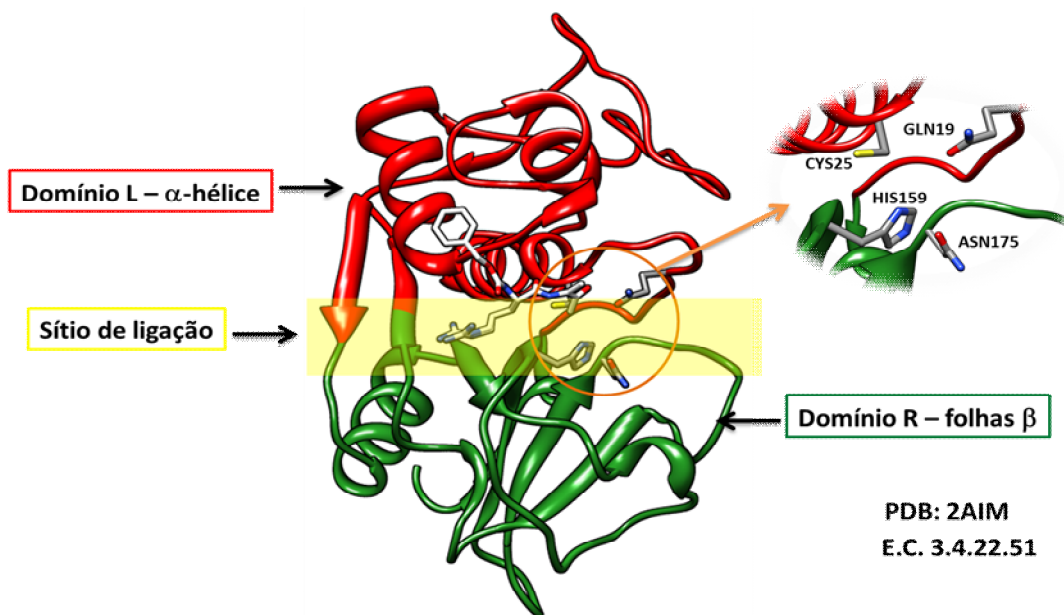


FIGURA 1.4 – Estrutura cristalográfica da cruzaina (PDB: 2AIM). Em destaque os resíduos que participam do mecanismo catalítico

O mecanismo de hidrólise de peptídeos desempenhado pelas cisteíno proteases envolve quatro etapas (FIGURA 1.5). A primeira é a desprotonação do grupo tiol da cisteína catalítica pela histidina. No próximo passo o enxofre aniônico nucleofílico se liga ao carbono carbonílico do substrato formando um intermediário tetraédrico, o qual é estabilizado pela cavidade do oxiânion. Tal cavidade é caracterizada pela presença do resíduo Gln19, que ajuda na catálise estabilizando o

intermediário oxianion. Em seguida o fragmento N-terminal do substrato é liberado e a histidina é restituída na sua forma desprotonada. Na sequência o intermediário acil-enzima é hidrolisado e um segundo intermediário tetraédrico é formado, novamente sendo estabilizado pela cavidade do oxianion. Em seguida, ocorre a liberação do substituinte N-terminal, restituindo a enzima no estado inicial.<sup>24-26</sup>

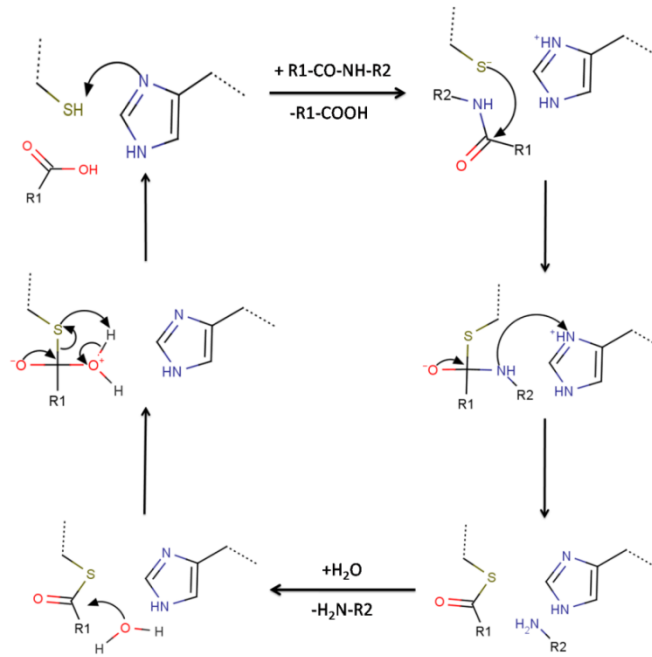


FIGURA 1.5 - Mecanismo de catálise de enzimas cisteíno proteases

A terminologia descrita para proteases foi adotada para a cruzaina, denominando sub-regiões do sítio ativo de S1 a S4<sup>27</sup>, os sub-sítios estão localizados conforme mostradas na FIGURA 1.6.

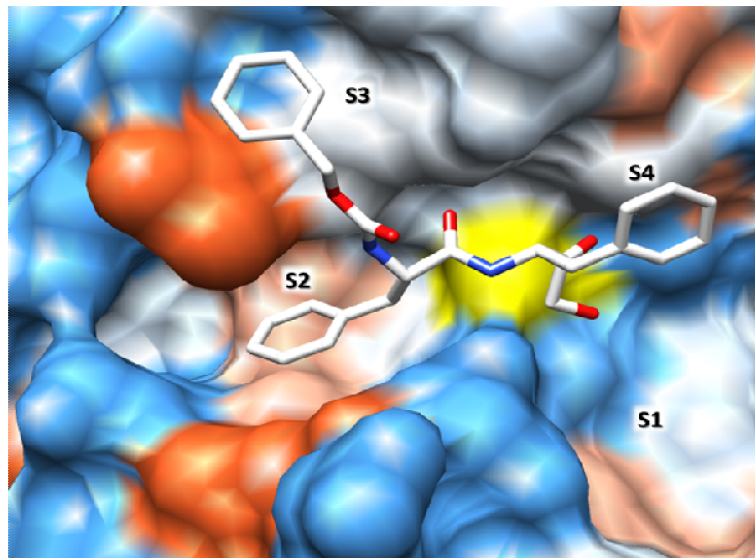


FIGURA 1.6 – Sítio da cruzaina representado por superfície de van der Waals – em amarelo está representada a posição da cisteína catalítica (PDB: 1ME4)

O sub-sítio S1 é onde se localiza a tríade catalítica da enzima, o sub-sítio S2 possui uma característica hidrofóbica localizado em uma região mais profunda, os sub-sítios S3 e S4 são mais superficiais e possuem também característica hidrofóbica. Apenas o centro do sítio possui característica hidrofílica, onde há o reconhecimento da ligação amídica.

A cruzaina possui excelentes provas de validação pré-clínica como um alvo de fármacos viável. Vários inibidores dessa protease, com diferentes esqueletos moleculares e mecanismos catalíticos, mostraram ser capazes de curar a infecção chagásica em modelos celulares, caninos e em ratos.<sup>12, 17, 28-32</sup> Inibidores seletivos desta enzima bloqueiam a proliferação do parasito epimastigota extracelular e amastigota intracelular, e suspendem a metaciclogênese (transformação de epimastigotas para tripomastigotas metacíclicos).<sup>13, 17, 33</sup>

Portanto, a cruzaina representa um alvo interessante e potencial para busca de novos fármacos antichagásicos. As estruturas de inibidores existentes para a cruzaina são diversas, incluindo uma série de vinil-sulfonas,  $\alpha$ -hidroxicetonas, tiossemicarbazona e derivados de 2-hidroxi-aril-N-acilhidrazona, entre outros.<sup>13</sup> Dentre esses destaca-se o K11777 (FIGURA 1.7), uma vinil-sulfona.

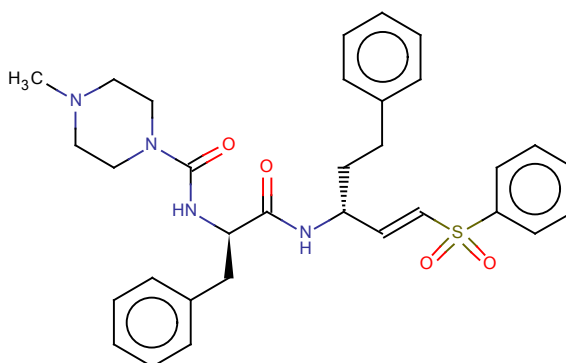


FIGURA 1.7 – Estrutura do K11777

O K11777, (2R)-N-[(1E,3R)-1-(benzenosulfonil)-5-fenilpent-1-en-3-il]-2-[(4-metilpiperazin-1-il)carbonilamino]-3-fenilpropanamida, foi eficaz na cura ou alívio da infecção por *T. Cruzi* em estudos pré-clínicos de prova de conceito e entrou nas investigações de desenvolvimento pré-clínico de fármaco. Em estudos, o K11777 demonstrou ser eficaz o suficiente para curar a doença em modelos agudos da infecção em camundongos e também reverter o dano cardíaco em cães. Estes

estudos demonstraram pela primeira vez que a inibição da enzima cruzaina efetivamente interfere na infecção pelo parasito *T. cruzi* e os inibidores peptídicos vinil sulfônicos não são tóxicos em doses terapêuticas, colocando sob questionamento a falta de seletividade deste inibidor em relação às catepsinas humanas.<sup>12, 34</sup>

Como dito anteriormente, a cruzaina pertence à família de enzimas proteases e possui um alto grau de identidade principalmente com a catepsina L (EC 3.4.22.15) e a catepsina B (EC 3.4.22.1).<sup>35</sup> As catepsinas estão envolvidas em muitos processos fisiológicos como na reabsorção óssea, liberação de hormônios da tireoide e possuem um papel fundamental no sistema imunológico. Elas também estão envolvidas em um número de processos degenerativos que incluem a osteoporose, artrite reumatoide, enfisema e distrofia muscular. Há também forte evidência da participação das catepsinas no câncer por atuação sobre os processos de invasão de tumores e metástases. As catepsinas, em especial a catepsina L, são alvos atrativos para o desenvolvimento de agentes anticâncer.<sup>36</sup>

# **CAPÍTULO 2**

## **A Química Medicinal e o Planejamento de Fármacos**

Neste capítulo é apresentado o papel da Química Medicinal e as ferramentas da quiminformática utilizadas no planejamento de fármacos.

## 2 - Química Medicinal e o Planejamento de Fármacos

### 2.1 - Planejamento de fármacos

A química medicinal é uma disciplina baseada na química, envolvendo também aspectos das ciências biológicas, médicas e farmacêuticas. Ela está preocupada, dentre outros fatores, com o planejamento, identificação e preparação de compostos biologicamente ativos, além da interpretação do seu modo de ação a nível molecular.<sup>37</sup> Os químicos medicinais que estão engajados na descoberta de fármacos fazem parte de equipes interdisciplinares, e devem, portanto, compreender não só o campo da química orgânica, mas também uma série de outras disciplinas para ajudar a levar o planejamento adiante.

No início da descoberta de medicamentos (de 1950 a cerca de 1980), os químicos medicinais se baseavam principalmente em dados obtidos em ensaios *in vivo*. No período mais recente (cerca de 1980 até o presente), o desenvolvimento de novas tecnologias, tais como ensaios em massa *in vitro*, bibliotecas de compostos de grande porte, alvos moleculares definidos e planejamento de fármacos baseado na estrutura do alvo, mudou esse panorama anterior relativamente simples. Embora estas novas tecnologias apresentem muitas oportunidades para o químico medicinal, a grande quantidade de novos requisitos de segurança que surgiu dificultam a tarefa de levar a atividade *in vitro* para a atividade *in vivo*.<sup>38</sup>

Na FIGURA 2.1 é demonstrado o processo de descoberta de fármacos, que começa com a identificação de uma necessidade médica, incluindo um juízo sobre a adequação das terapias existentes (se houver). A partir dessa análise, juntamente com uma avaliação do conhecimento atual sobre a doença-alvo, serão construídas hipóteses sobre como, possivelmente, melhorar a terapia. Etapas subsequentes do processo incluem a detecção de relevante atividade biológica *in vitro* para um composto estruturalmente novo, e em seguida, a verificação da atividade *in vivo* em um modelo animal adequado do composto da fase anterior, seguido da maximização desta atividade através da preparação de análogos estruturais e, finalmente, seleção de um composto como o candidato para desenvolvimento do fármaco. O candidato a fármaco é então submetido a testes



toxicológicos em animais, como exigidos por lei. Se a substância passa em todos os testes, todos os dados da pesquisa acumulados são reunidos e apresentados aos órgãos reguladores do governo (por exemplo, a ANVISA), antes que os ensaios clínicos sejam iniciados. Na fase clínica, ocorre a avaliação sequencial de tolerância em voluntários humanos saudáveis (Fase I), eficácia e o intervalo da dose em pacientes (Fase II), seguido por ensaios difundidos em milhares de pacientes apropriados para desenvolver um amplo banco de dados de eficácia e segurança (Fase III). Para os poucos candidatos a fármacos que sobreviverem até a fase III, um novo relatório que contém todos os dados da pesquisa acumulada é apresentado para a revisão aprofundada por parte dos especialistas do órgão regulador. Somente com a aprovação, o novo fármaco pode ser oferecido aos médicos e seus pacientes para tratar a doença para a qual foi planejado. Na fase clínica IV, são realizados os monitoramentos pós-comercialização dos novos fármacos.<sup>38</sup>

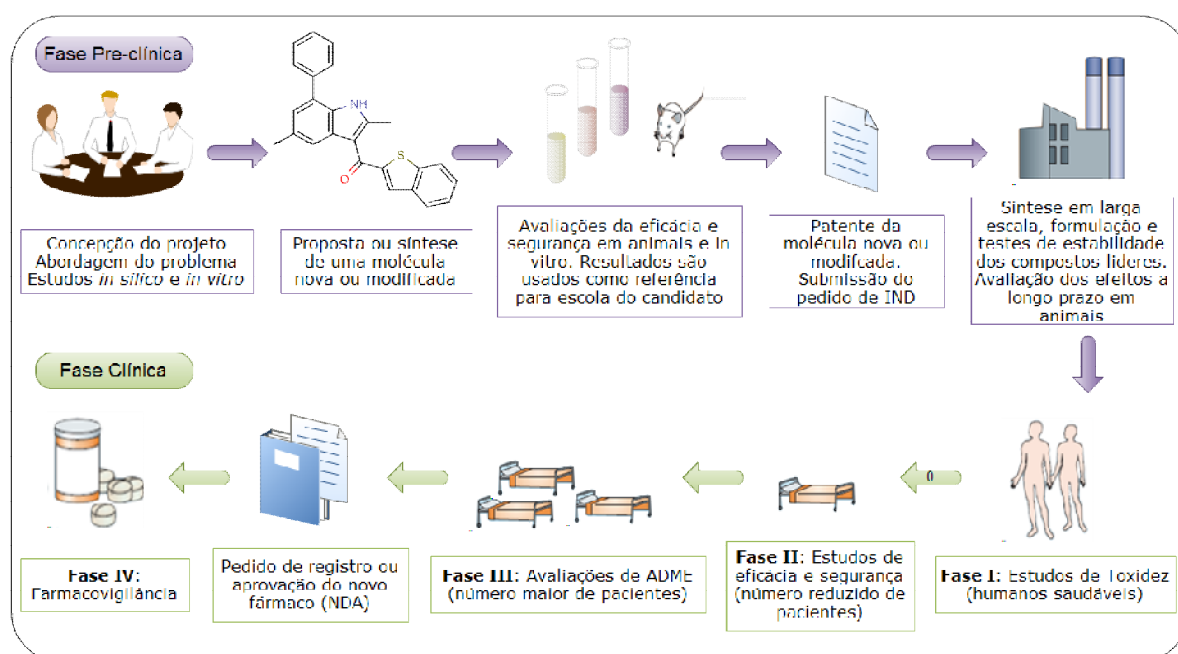


FIGURA 2.1 - Etapas do processo de descoberta de fármacos

Planejar e desenvolver um novo medicamento são processos longos, complexos, caros e muito arriscados. Na fase de pesquisa, somente uma pequena fração dos compostos são selecionados como candidatos a fármacos para o desenvolvimento. Experiências demonstraram que aproximadamente um de 15-25 candidatos a fármacos sobrevive aos rigorosos testes de segurança e eficácia.

Apesar dos elevados riscos, os potenciais benefícios para os milhões de pacientes com doenças graves representam uma motivação crescente.<sup>38</sup>

A taxa de sucesso no planejamento de fármacos é de apenas 10% apesar de todos os avanços obtidos nos últimos anos, sendo que cerca de 30% dos compostos que chegam aos estágios finais de desenvolvimento clínico é abandonado em decorrência da ausência de eficácia e insuficiente segurança clínica e toxicológica.<sup>38-40</sup>

Um dos obstáculos encontrados com certeza é o amplo espaço químico utilizado para a realização da busca. Atualmente o número de compostos existentes no *Chemical Abstracts* já ultrapassa 60 milhões.<sup>41</sup> Com sua magnitude, o espaço químico pode ser visto como sendo análogo ao universo, com os compostos químicos ocupando o lugar das estrelas.<sup>42</sup> Dada a vastidão do espaço químico, o desafio para os químicos medicinais é identificar as regiões que podem conter compostos biologicamente ativos, isto é, o espaço químico de interesse biológico. Os limites desse espaço são definidos pela interação específica de ligações entre as pequenas moléculas e os alvos macromoleculares, tais como proteínas, DNA e RNA. Medido em termos de propriedades físico-químicas e descritores topológicos, compostos terapêuticos parecem se agrupar em “galáxias”. Um século de química medicinal e o avanço dos programas HTS sugerem que compostos que se ligam às certas "classes de alvos" (proteínas da mesma superfamília, como o receptor acoplado à proteína G; GPCRs) são agrupados em regiões distintas do espaço químico, como demonstrado na FIGURA 2.2.<sup>42</sup>

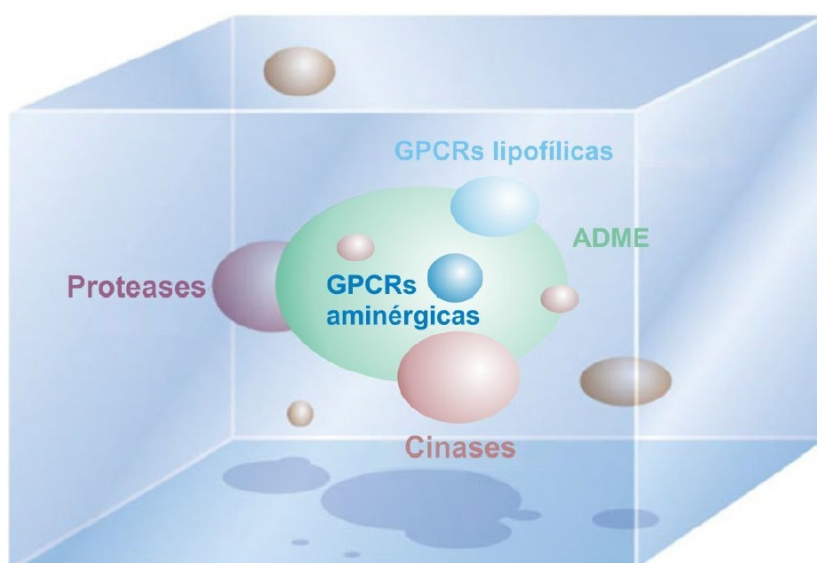


FIGURA 2.2 - Demonstração do espaço químico e do espaço químico-biológico

Para navegar pelo vasto espaço químico-biológico, os químicos medicinais utilizam de métodos em quiminformática. A realidade é que atualmente o uso de computadores e métodos computacionais permeia todos os aspectos da descoberta de fármacos. Aqueles que são mais hábeis com as ferramentas computacionais têm a vantagem de oferecer novos candidatos a fármacos mais rapidamente e com custos mais baixos. Baseando-se na premissa de que o conhecimento da estrutura e das propriedades dos ligantes ativos contra o alvo biológico de interesse e daqueles sem atividade biológica é valioso para o enriquecimento, identificação e otimização dos candidatos, a química medicinal emprega os métodos em quiminformática desde a etapa de identificação do ligante até a fase pré-clínica (FIGURA 2.3).<sup>43</sup>



FIGURA 2.3 - Estágios do processo de desenvolvimento de um fármaco onde os métodos em quiminformática são aplicados

Métodos estão sendo desenvolvidos atualmente para uma varredura mais eficaz do espaço químico pela seleção de um número relativamente pequeno de "fragmentos" de baixo peso molecular. Embora, em teoria, um número incrivelmente grande de compostos fármacos-similares possa ser sintetizado, o espaço químico cresce em função do número de átomos do composto (isto é, seu peso molecular). Logo, como aumenta o número de átomos, também aumenta o número de permutações possíveis. Assim, a restrição do peso molecular pela triagem apenas de fragmentos permite explorar uma área muito maior do espaço químico possível utilizando um número bem menor de compostos (para um determinado número de átomos). Com baixo peso molecular, fragmentos possuem em geral baixa atividade biológica *in vitro*, devido ao pouco número de interações com o alvo. Estes então são crescidos para formar ligantes com alta afinidade e maior peso molecular. Diversas abordagens baseadas nesta ideia têm sido desenvolvidas, sendo este o tema discutido a seguir.<sup>42</sup>

## 2.2 - Planejamento de fármacos baseado em fragmentos

O planejamento de fármacos baseado em fragmentos (FBDD, do inglês *fragment-based drug discovery*) emergiu nas últimas décadas e provou ser um novo paradigma para o planejamento de fármacos. Esforços significativos das indústrias farmacêuticas e companhias biotecnológicas aplicando esta abordagem nos anos 1990 e início dos anos 2000 resultaram no desenvolvimento de mais de dez candidatos clínicos. Como consequência, FBDD tem sido usado como principal alternativa para métodos tradicionais usados no planejamento de fármacos, como HTS e ensaios virtuais.

Existem duas diferenças básicas entre FBDD e HTS. A primeira é o tamanho do espaço químico. O número de composto fármacos-similares está estimado em  $10^{60}$ , indicando que o espaço químico de compostos para HTS é muito grande para ser amostrado com uma biblioteca de tamanho realístico. O número de compostos em bibliotecas de fragmentos é tipicamente relatado como alguns milhares de compostos, os quais cobrem a maior parte do espaço químico de compostos fragmentos-similares comercialmente disponíveis.<sup>44</sup>

Os princípios fundamentais de reconhecimento molecular, que desempenharam um papel fundamental no desenvolvimento do planejamento de fármacos baseado em fragmentos (FBDD), começaram a mais de duas décadas. Em 1981, Jencks introduziu a ideia de que moléculas grandes podem ser consideradas como a combinação de dois ou mais "fragmentos" que contêm todas as características necessárias para a ligação com a proteína-alvo. Em 1985, Goodford desenvolveu um programa de computador chamado GRID, onde pequenos fragmentos ou 'sondas' foram usados para inspecionar um suposto site de ligação, a fim de destacar os pontos importantes de interação. Análise do GRID, aplicada a alguns sítios de interação de proteínas, indicou que algumas das posições favoráveis encontradas para fragmentos moleculares individuais estavam presentes em inibidores conhecidos. Fragmentos e moléculas grandes têm que obedecer ao mesmo conjunto de regras de reconhecimento molecular quando se ligam a uma proteína-alvo. Durante um evento de interação, os fragmentos e as macromoléculas têm de superar uma barreira entrópica associada à perda de entropia do corpo rígido sobre a ligação para o alvo. Esta barreira tem apenas uma pequena dependência do tamanho molecular. Fragmentos podem arcar com o custo entrópico de ligação

somente usando uma alta proporção de seus átomos para formar interações e, portanto, fragmentos são assim considerados ligantes eficientes (altas energias de interação por unidade de massa molecular).<sup>45-46</sup>

O final dos anos 1990 e início de 2000 foi marcado pela publicação de uma série de artigos importantes que se mostraram influentes na formação do caminho que o FBDD desenvolveu. Em 1996, cientistas da Abbott relataram o uso de "SAR por RMN". Esta técnica faz uso de alterações nos espectros de RMN de proteína devido à interação do ligante e foi utilizada para identificar fragmentos de baixa afinidade de interação em sítios adjacentes da proteína FK506. Foi possível utilizar informações derivadas do RMN para vincular esses fragmentos com um 'linker' curto e flexível, resultando em um inibidor nanomolar.<sup>45-46</sup>

Inúmeros critérios para seleção de fragmentos para ensaios têm sido sugeridos. Tipicamente, eles envolvem restrições de peso molecular, número de átomos diferentes de hidrogênio, lipofilicidade, doadores e aceitadores de ligação de hidrogênio, área superficial polar, dentre outros. Pesquisadores da Astex propuseram a regra dos três (Ro3):  $MW \leq 300$ ,  $clogP \leq 3$ , doadores de ligações de hidrogênio  $\leq 3$ , aceitadores de ligação de hidrogênio  $\leq 3$ .<sup>47</sup> Contudo, eles não apresentaram análises que dessem suporte a esses parâmetros e outros grupos acabam por não utilizá-los, adotando parâmetros diferentes.<sup>48-49</sup>

Fornecedores de compostos foram rápidos em responder ao aumento da demanda por bibliotecas de fragmentos. As primeiras bibliotecas de fragmentos disponíveis comercialmente foram para o mercado em 2003, com Asinex e Maybridge fornecendo as primeiras coleções que seguiam à 'regra dos três' (Ro3). Atualmente, muitos fornecedores vendem bibliotecas de fragmentos comerciais.<sup>45-46</sup>

Apesar de fragmentos apresentarem baixa afinidade, devido à presença de poucos sítios de interação, eles podem interagir preferencialmente com 'hot spots' da superfície da proteína alvo com modo de interação otimizado devido sua estrutura simples. Compostos orgânicos pequenos interagem quase exclusivamente com regiões bem-definidas e localizadas na proteína, sendo ligantes altamente eficientes.<sup>44, 50</sup> Uma vez que estes pontos chave de interação são identificados, interações com regiões adjacentes da superfície da proteína podem ser posteriormente exploradas para aumentar a seletividade e melhorar a atividade.<sup>48</sup>

Um conceito valioso e bastante usado para comparar a atividade biológica de fragmentos moleculares é a eficiência do ligante (LE). Este conceito é útil para comparar fragmentos em séries diferentes e para estabelecer fragmentos prioritários. Este termo é definido como sendo a energia livre de interação de um composto para uma proteína específica dividida por cada átomo diferente de hidrogênio (HAC) presente nesse composto (equação 1). Em geral, um fragmento é considerado promissor se sua eficiência (LE) é maior ou igual a  $0,30 \text{ kcal mol}^{-1}$  átomo pesado<sup>-1</sup> e este mesmo número é usado para comparar ligantes em uma série e para guiar a otimização de um fragmento em ligantes maiores.<sup>45-46, 51-52</sup>

$$LE = -\Delta G/HAC = \frac{-RT \ln(IC_{50})}{HAC} \quad \text{equação 1}$$

Uma vez identificados fragmentos ativos, estes são otimizados utilizando técnicas como: crescimento do fragmento ou a combinação de diferentes fragmentos (FIGURA 2.4). O primeiro caso consiste em adicionar funcionalidade a um fragmento inicial que explora uma região específica do sítio de interação para encontrar interações favoráveis que aumentem a afinidade de ligação. Isso tem sido usado com sucesso para desenvolver inibidores para uma série de alvos, incluindo inibidores da p38 MAP quinase (Astex), DNA girase (Hoffman-La Roche), ERM metil transferase (Abbott), e uroquinase (Abbott).

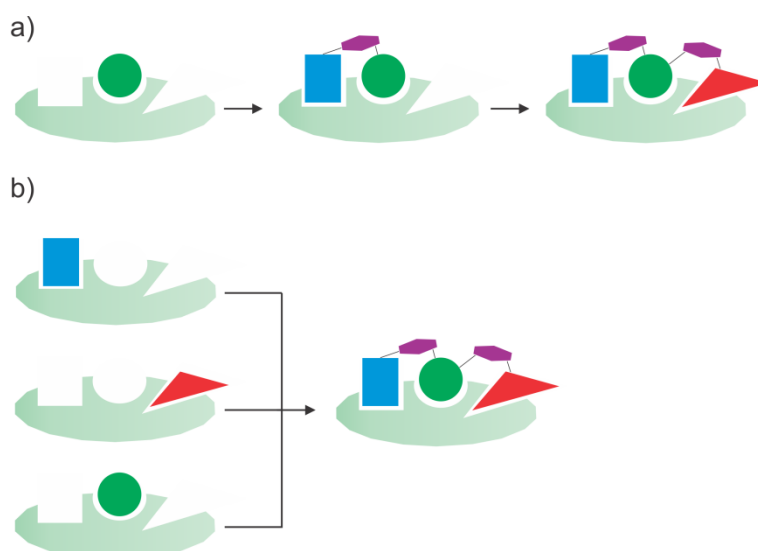


FIGURA 2.4 - Otimização do fragmento. a) Crescimento do fragmento, pedaço a pedaço. b) Combinação de diferentes fragmentos

Compostos de alta afinidade também podem ser criados através da ligação de fragmentos de baixa afinidade. Isso requer vários fragmentos que se

liguem a sítios de interação adjacentes, o conhecimento de seus modos de ligação e orientações e, finalmente, um *'linker'* apropriado que possa unir os fragmentos, mantendo as interações de ligação dos fragmentos individuais.<sup>53</sup>

O conhecimento sobre onde fragmentos químicos específicos se ligam em uma proteína pode ser utilizado em uma grande variedade de técnicas computacionais, tais como buscas farmacofóricas de bibliotecas químicas disponíveis. Estas pesquisas permitem aos pesquisadores encontrar ou sintetizar compostos que validem a hipótese de ligação e estendam o fragmento em outras partes do sítio.<sup>53</sup>

Fragmentos apresentam baixa afinidade, o que implica em dizer que é necessário um sistema robusto, capaz de detectar e quantificar interações fracas da pequena molécula com a macromolécula. A disponibilidade de um ensaio de alta qualidade pode permitir a exploração de um composto com baixo peso molecular e baixa complexidade. Dentre as técnicas pode-se destacar o uso de calorimetria, ressonância magnética nuclear, cristalografia de raios-X e espectrometria de massas.<sup>48, 50</sup> Vários trabalhos na literatura relatam metodologias de ensaios virtuais e modelagem molecular para a identificação e seleção de fragmentos com potencial para inibir alvos enzimáticos. Na FIGURA 2.5 está representado um esquema da estratégia em geral abordada.<sup>54</sup>

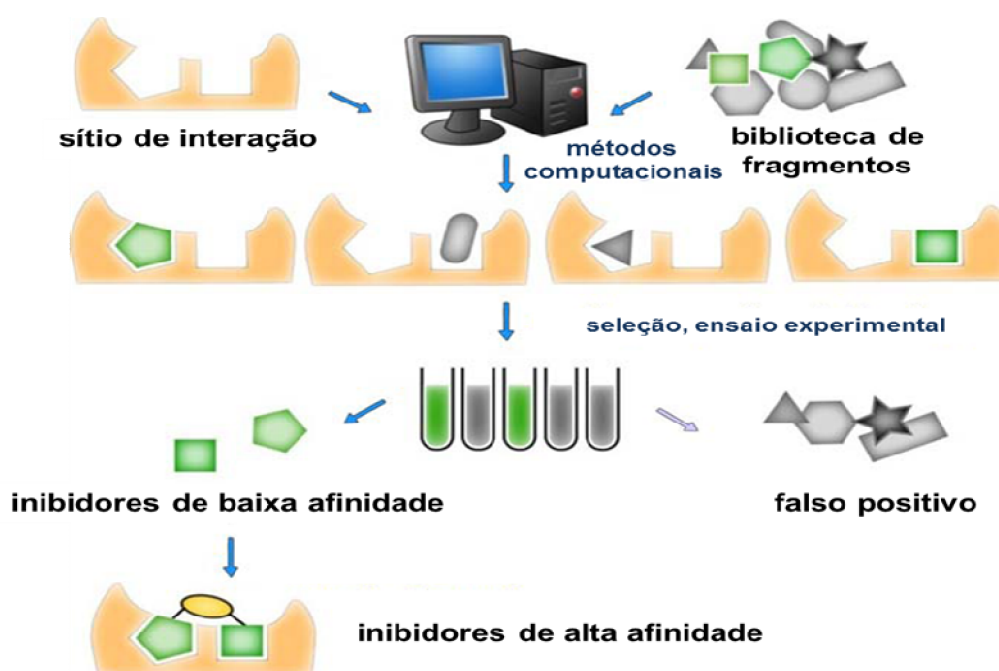


FIGURA 2.5 – Estratégia adotada pra planejamento baseado em fragmentos<sup>54</sup>

Coleções virtuais de fragmentos são submetidas a ensaios virtuais, e.g. docagem molecular e simulações de dinâmica molecular, onde são pontuados, e então selecionados pelo químico medicinal para os ensaios experimentais. Inibidores de baixa afinidade, porém com elevada eficiência de ligante, são então identificados e posteriormente otimizados para se obter inibidores de elevada afinidade.<sup>54</sup>

O primeiro composto derivado do planejamento baseado em fragmentos a entrar nos testes clínicos foi o LY-517717, um inibidor altamente seletivo do fator Xa descoberto em 2002. Esta molécula é administrada via oral e possui MW de 459 Da. Em contraste com inibidores indiretos de FXa (por exemplo, Fondaparinux), LY-517717 inibe diretamente FXa, não necessitando de antitrombina como cofator. Este inibidor mostrou propriedades antitrombóticas em animais experimentais e um perfil farmacocinético favorável em estudos de fase I. Em voluntários saudáveis, LY-517717 mostrou uma atividade anticoagulante que atingiu seu pico dentro de 0,5-4 horas após a administração oral. Estas características farmacológicas conferem ao fármaco um potencial para segurança e eficácia na dose diária. Atualmente, esse composto se encontra em fase clínica II.<sup>55</sup>

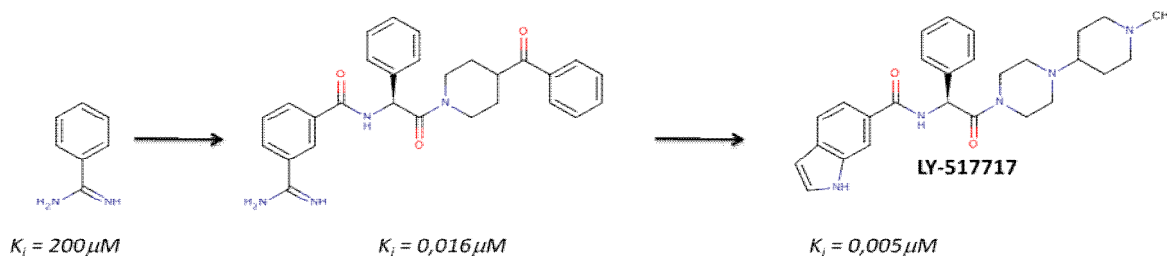


FIGURA 2.6 – Planejamento do LY-517717<sup>45</sup>

Um conhecido aglutinante, a benzamidina, foi usado como um ‘fragmento-modelo’ para o sub-sítio S1 da enzima Fator Xa por apresentar o modo de interação bem definido por estrutura cristalográfica. Planejamento baseado na estrutura do alvo foi usado para atingir sub-sítios adjacentes da enzima e dirigir as etapas de síntese química. Foram identificados compostos com potência na ordem de nM, tal como mostrado na FIGURA 2.6. Esse composto, apesar de potente, apresentou baixa biodisponibilidade oral, que foi atribuída ao grupo benzamidina altamente básico, o qual é relatado por impedir a absorção. Otimizações subsequentes usando uma combinação entre química medicinal e planejamento baseado na estrutura do alvo levou à substituição de grupos como a benzamidina



por grupos que mantivessem as interações favoráveis antes observadas para o grupo benzamidina. As etapas de otimização levaram à descoberta do candidato biodisponível por via oral, LY517717.<sup>45</sup>

# **CAPÍTULO 3**

## **Objetivos**

Os objetivos desta dissertação de mestrado são apresentados neste capítulo.

### 3 - Objetivos

Identificar novos fragmentos moleculares com interação no sub-sítio S2 da cruzaina.

Estratégias específicas usadas para atingir o objetivo geral:

- ✓ Selecionar fragmentos moleculares por métodos computacionais;
  - Construir uma coleção de fragmentos;
  - Identificar pontos de seletividade (cruzaina-catepsina L) no sítio da enzima-alvo (cruzaina);
  - Identificar a conformação e a energia de interação de fragmentos utilizando métodos de docagem e dinâmica molecular;
- ✓ Realizar ensaios bioquímicos para verificar a inibição da atividade da enzima;
  - Realizar estudos de relação estrutura-atividade para fragmentos que apresentarem atividade inibitória.

# **CAPÍTULO 4**

## **Inspeção do Sítio: Identificação de Pontos de Seletividade**

Neste capítulo é feita uma breve introdução sobre os métodos, as estratégias e critérios usados na inspeção do sítio da cruzaina. Também são apresentados os resultados obtidos e a discussão dos mesmos.

## 4 - Inspeção do sítio usando campos de interação molecular

No programa GRID<sup>56</sup>, uma grade tridimensional é construída sobre uma região molecular de interesse ou dentro de alguma cavidade da proteína. Sondas (grupos funcionais que representam diferentes propriedades físico-químicas) são movidas ponto-a-ponto, e para cada ponto, a energia de interação entre a sonda e o sistema é computada usando um campo de força. No GRID, o campo de força é baseado no potencial de Lennard-Jones, uma função eletrostática dependente da distância, e um termo para ligação de hidrogênio. Para cada ponto na grade, a energia de interação é obtida, o que em conjunto forma um campo de interação molecular (MIF – *Molecular Interaction Field*), que representa o potencial de interação do composto/proteína com certo grupo químico (sonda).<sup>57</sup>

MIFs podem ser calculados em biomoléculas tais como receptores ou enzimas, para sugerir pontos de interação importantes entre a proteína e o ligante, ou diretamente dos ligantes para sugerir posições onde um possível receptor poderia interagir com o ligante.<sup>57</sup>

### 4.1 - Materiais e Métodos

Os estudos de inspeção do sítio da cruzafina foram realizados no Laboratório de Quiminformática e Modelagem Molecular do Grupo de Estudos em Química Medicinal de Produtos Naturais – NEQUIMED-PN, do Instituto de Química de São Carlos (IQSC/USP). Os trabalhos foram desenvolvidos em uma estação de trabalho operando Linux Fedora 10.

A abordagem original (GRID/PCA) é baseada na análise dos campos de interação molecular gerados pelo GRID usando análise de componentes principais (PCA) para destacar as diferenças mais relevantes entre os alvos enzimáticos dados. Mas este método permite análise de pares de alvos e a análise de mais de dois alvos é complexa.<sup>57</sup>

GRID/CPCA é baseada no GRID/PCA, mas apresenta soluções a esse problema descrito. Iniciando de uma ou mais estruturas 3D de muitos alvos, uma descrição multivariada do sítio de interação é feita usando o GRID. Então campos de interação molecular (MIF) são analisados por PCA consensual (CPCA). Os diferentes MIFs obtidos para as diferentes sondas são adicionados, lado-a-lado,

adicionando novas variáveis para o mesmo alvo. Como resultado se obtém gráficos de contorno destacando as regiões e o tipo de interações nessas regiões que podem ser usados para introduzir seletividade em um ligante potencial dessas proteínas-alvo.

A matriz descrevendo o sistema tem uma linha para cada estrutura 3D estudada e  $n \times k$  colunas, correspondendo a  $n$  sondas multiplicado pelas  $k$  variáveis presentes em um único MIF (FIGURA 4.1). O PCA de cada matriz produz uma parcela de pontuação onde cada estrutura 3D é representada por um único ponto. Quando diversas estruturas diferentes são usadas para representar cada alvo, as parcelas de pontuação devem mostrá-las agrupadas, o que significa que diferenças entre as estruturas 3D do mesmo alvo são menos importantes do que diferenças entre proteínas distintas.

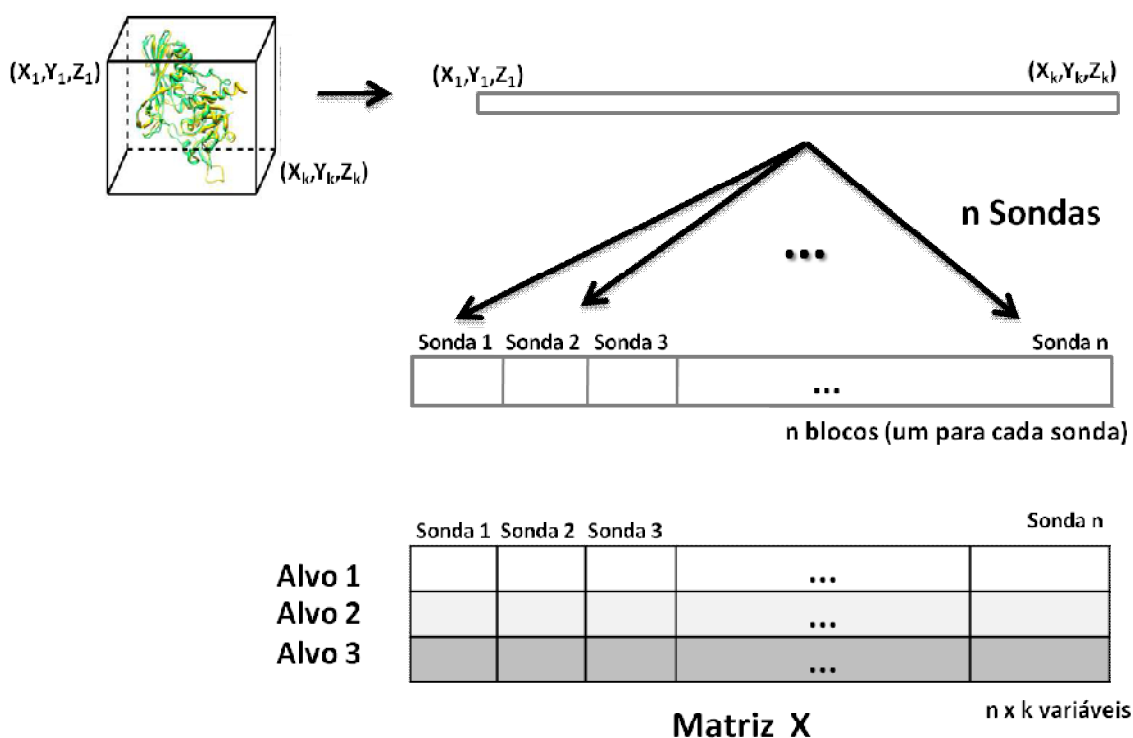


FIGURA 4.1 – Procedimento usado para construir a matriz X. Iniciando pelos cálculos do GRID para uma sonda, é construído um vetor contendo todas as energias de interação nos  $k$  pontos do grid; depois, os vetores para  $n$  sondas são compilados em um único vetor  $n \times k$ . 'Empilhando' os vetores, cada alvo corresponderá a uma linha na matriz X final.

Além disso, nesta metodologia as sondas diferentes são organizadas em diferentes blocos de variáveis. Portanto, é possível aplicar um procedimento de escalonamento, que normaliza a sua importância no modelo. O processo de ponderação é chamado *Block Unscaled Weights* (BUW) e serve para cumprir o objetivo delineado. A operação escalona cada bloco variável separadamente,

enquanto que as escalas relativas de variáveis dentro de cada bloco permanecem inalteradas.

Os cálculos foram feitos no programa GRID v22c<sup>56</sup> e os campos de interação moleculares gerados analisados no programa GOLPE v4.6<sup>58</sup>. O seguinte pré-tratamento dos dados foi feito: (a) o valor de máximo *cutoff* foi considerado igual à zero para considerar somente interações positivas; (b) BUW foi utilizado para normalizar os dados; (c) variáveis com valores menores que 0,01 kcal mol<sup>-1</sup> e com desvio padrão abaixo de 0,02 kcal mol<sup>-1</sup> foram removidos para eliminar ruídos no sistema.

Visando construir um modelo para busca de ligantes seletivos à cruzaina, realizou-se uma busca por estruturas com elevada identidade com este alvo utilizando o programa BLAST. As estruturas selecionadas foram extraídas do *Protein Data Bank*<sup>59</sup> e alinhadas utilizando o programa Sybyl v8.0<sup>60</sup>, onde também foram separados os ligantes cristalizados com as enzimas. Para o alinhamento foi utilizada a estrutura da cruzaina de código PDB 1ME4 como referência.

Para os cálculos de campo de interação molecular (MIF) foi utilizada uma grade tridimensional de 10Å x 17Å x 11Å, com espaçamento entre os planos de grade igual a 1Å. As sondas aplicadas foram: C3, NM3 e OH.

## 4.2 - Resultados e Discussão

As estruturas selecionadas para o estudo GRID/CPCA, bem como os dados referentes aos seus parâmetros estruturais, são apresentadas na TABELA 4.2. Na busca por estrutura com elevada identidade com a cruzaina, foram destacadas duas estruturas para a catepsina L (EC 3.4.22.15) e duas para a catepsina B (EC 3.4.22.1), já que estas apresentam elevada identidade de sequência de aminoácidos com relação à cruzaina, sendo aumentada quando se compara apenas o sítio de interação. Um importante parâmetro nos estudos cristalográficos é a resolução dos dados, que é expressa em Å, onde números menores indicam resoluções mais altas. Em altas resoluções (< 1,5 Å) o modelo é provavelmente mais que 95% uma consequência dos dados observados. Contudo, em baixas resoluções (> 2,5 Å), a modelagem dos detalhes nas estruturas das proteínas é muito mais subjetiva do que baseada em informações experimentais.<sup>61</sup>

TABELA 4.2 - Código PDB das estruturas estudadas, parâmetros estruturais e grau de identidade

Código	Resolução (Å)	R <sub>factor</sub>	R <sub>free</sub>	(R <sub>free</sub> -R <sub>factor</sub> )	Identidade (%) <sup>a</sup>	Identidade sítio (%) <sup>b</sup>
<i>Cruzaína</i>						
1EWL	2,00	0,171	0,233	0,062	--	--
1EWM	2,00	0,173	0,206	0,033	--	--
1EWO	2,10	0,221	0,262	0,041	--	--
1EWP	1,75	0,183	0,206	0,023	--	--
1F29	2,15	0,185	0,208	0,023	--	--
1F2A	1,60	0,184	0,219	0,035	--	--
1F2B	1,80	0,173	0,203	0,030	--	--
1ME3	1,20	0,103	0,135	0,032	--	--
1ME4	1,20	0,091	0,124	0,033	--	--
1U9Q	2,30	0,136	0,218	0,082	--	--
<i>Catepsina B</i>						
1GMY	1,90	0,161	0,199	0,038	31,2	54,1
1CSB	2,00	0,194	n/a	n/a	30,7	52,5
<i>Catepsina L</i>						
3BC3	2,20	0,183	0,234	0,051	45,7	65,6
3HWN	2,33	0,252	0,347	0,095	46,1	57,4

\*Identidade calculada usando como estrutura de referência no alinhamento: a) estrutura PDB: 1ME4; b) sítio de interação da cruzaína, tomando uma região de 10 Å em torno do inibidor T10 (obtido no co-cristal PDB: 1ME4)

A habilidade do modelo em explicar os dados experimentais é usualmente identificada por meio do parâmetro chamado R<sub>factor</sub>, porém, esse parâmetro está sujeito ao ajuste forçado (over-fitting) dos dados. Assim, foi introduzido o conceito de validação cruzada no refinamento cristalográfico, com a introdução do parâmetro R<sub>free</sub>, cuja definição é idêntica à do R<sub>factor</sub>, exceto que o R<sub>free</sub> é calculado para um pequeno conjunto de reflexões que nunca são usadas no refinamento do modelo. O parâmetro R<sub>free</sub> mede a capacidade do modelo em prever observações experimentais que não são usadas para ajustar o modelo. Os valores desejáveis são < 0,25 para o R<sub>factor</sub> e < 0,40 para R<sub>free</sub>. A diferença entre os parâmetros R<sub>free</sub> e R<sub>factor</sub> é uma medida parcial da extensão na qual o modelo superestima os dados. Esta diferença (R<sub>free</sub> - R<sub>factor</sub>) deve ser baixa para o modelo final, idealmente <0,05.<sup>62</sup> Baseado nesses fatores, várias estruturas têm



características apropriadas para os ensaios virtuais. As estruturas 1ME4 e 1ME3 são as que apresentam melhor resolução, sendo a primeira escolhida como estrutura referência nos alinhamentos com as estruturas cristalográficas das catepsinas.

A caixa de grade, criada para os cálculos de campo de interação molecular, foi feita visando delinear a área adjacente ao sub-sítio S2 do sítio ativo, conforme destacado na FIGURA 4.2. As sondas aplicadas foram escolhidas para tentar descrever características relevantes: C3 – estérea e hidrofóbica, NM3 – estérea e carregada positivamente e OH – doadora e aceitadora de ligação de hidrogênio.

Por inspeção visual do sítio de interação da cruzaina a partir das estruturas cristalográficas depositadas no PDB, foi possível observar a existência de resíduos flexíveis no sub-sítio S2 (FIGURA 4.3), como Glu205, Leu67 e Asn69. Isso é corroborado pelos resultados obtidos por Durrant e colaboradores<sup>33</sup>, onde estudos de dinâmica molecular na presença de ligante próximo ao S2 foram feitos e se observou que a área adjacente ao sub-sítio S2 ficou mais exposta, sendo um ponto interessante para o planejamento de inibidores dessa enzima. Além disso, dos ligantes co-cristalizados na cruzaina, disponíveis no PDB, nenhum explora essa região, como pode ser visto na FIGURA 4.4.

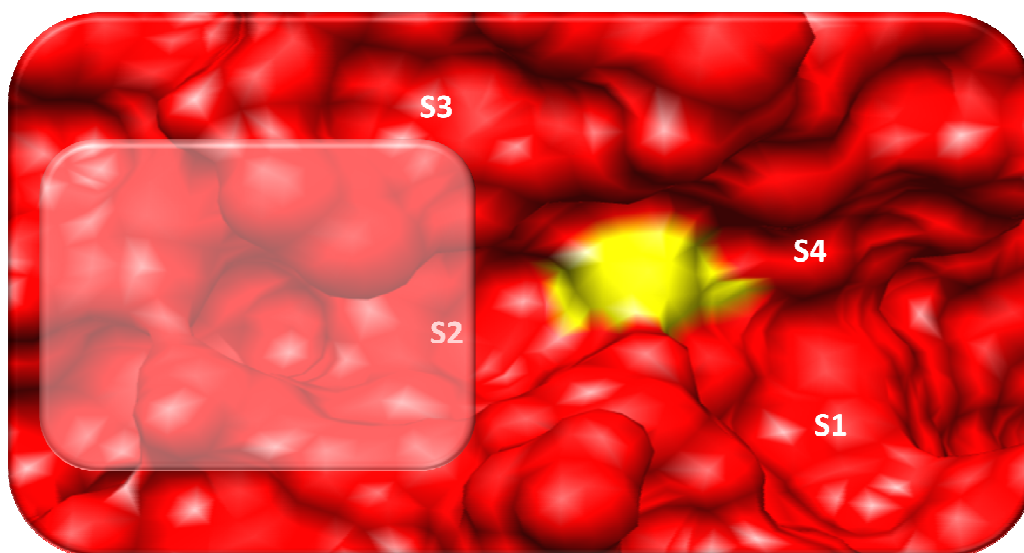


FIGURA 4.2 - Sítio Ativo da Cruzaina

De modo paralelo, na análise do alinhamento da estrutura da cruzaina com as estruturas das catepsinas L e B foi interessante perceber a estrutura cristalográfica PDB 3HWN, da catepsina L, onde o sub-sítio S2 é explorado para obtenção de um ligante seletivo em relação a outras catepsinas<sup>63</sup>. Assim esse sub-

sítio foi utilizado neste estudo na tentativa de buscar pontos de seletividade em relação à cruzaina.

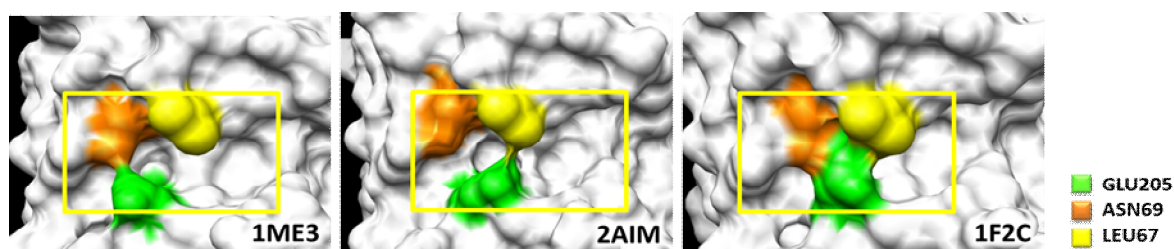


FIGURA 4.3 - Sub-sítio S2 da cruzaina em diferentes estruturas cristalográficas depositadas no PDB

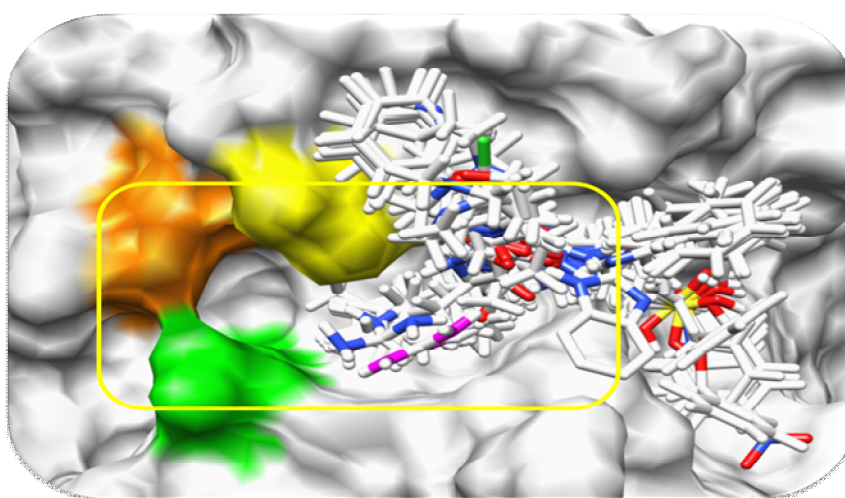


FIGURA 4.4 – Sobreposição dos ligantes co-cristalizados com a cruzaina. Em amarelo, destaque para o sub-sítioS2

Foram geradas quatro componentes principais. Usualmente, a primeira representa a discriminação entre os grupos de enzimas e a segunda pode ser usada para classificar as sondas de acordo com a força de interação. A análise do gráfico de *scores* de PCA para o modelo gerado indica uma boa distribuição dos grupos de enzimas em relação à PC1 e PC2, como verificado na FIGURA 4.5. A primeira componente discrimina entre o grupo 3 (representando as estruturas da Cruzaina) e os grupos 1 (Catepsina B) e 2 (Catepsina L).

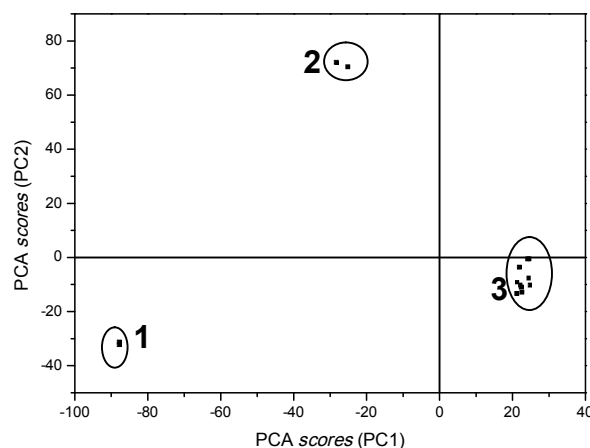


FIGURA 4.5 - CPCA score: 1)Catepsina B; 2)Catepsina L; 3)Cruzaína

A FIGURA 4.6 mostra a contribuição de cada sonda para a pontuação total. Como se pode perceber, todas as três sondas empregadas neste estudo apresentaram contribuição similar para a primeira e segunda componentes.

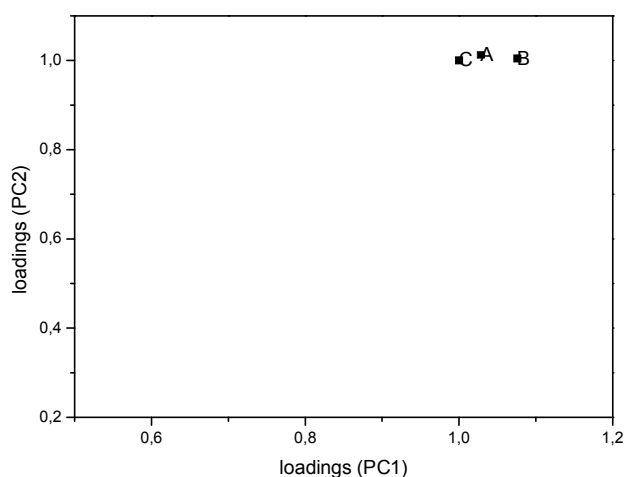


FIGURA 4.6 - Contribuição de cada sonda para o modelo gerado: A) C3; B) NM3 e C)OH

As figuras 4.7 e 4.8 mostram a distribuição dos pseudo-campos gerados pelo GRID/CPCA para a cruzaína em relação à catepsina L e catepsina B, respectivamente. Como observado, a seletividade da cruzaína em relação aos alvos é estudada através da seleção de pontos relevantes no gráfico da diferença dos pseudo-campos de interação gerados para a cruzaína em relação a cada uma das catepsinas.

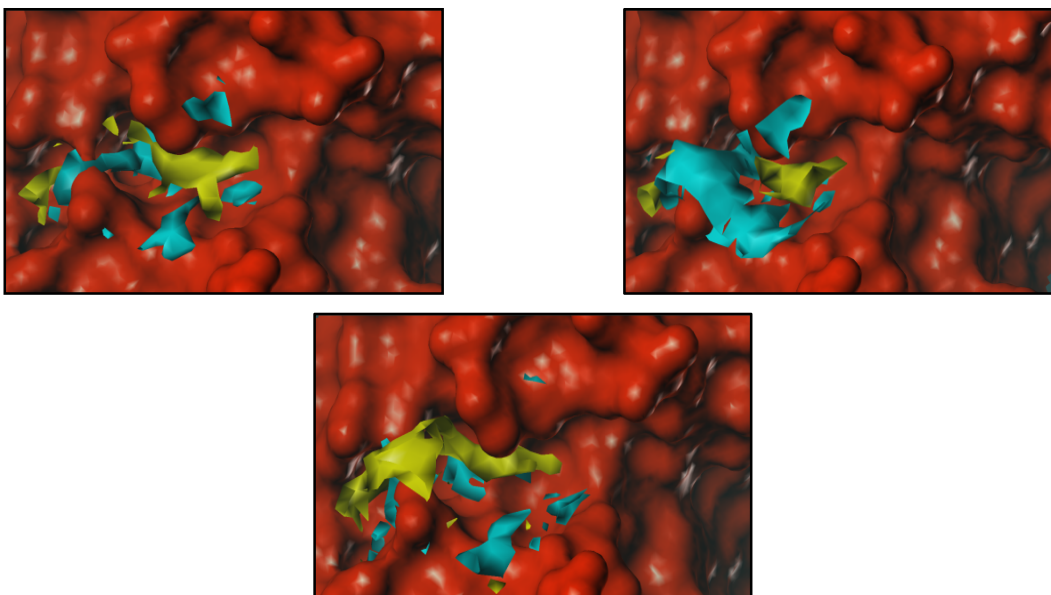


FIGURA 4.7 - Diferença nos pseudo-campos gerados, mostrando as regiões de favorecimento para cada alvo. Sondas: A) C3, B) NM3 e C) OH. (Ciano=Cruzaína e Amarelo=Catepsina L)

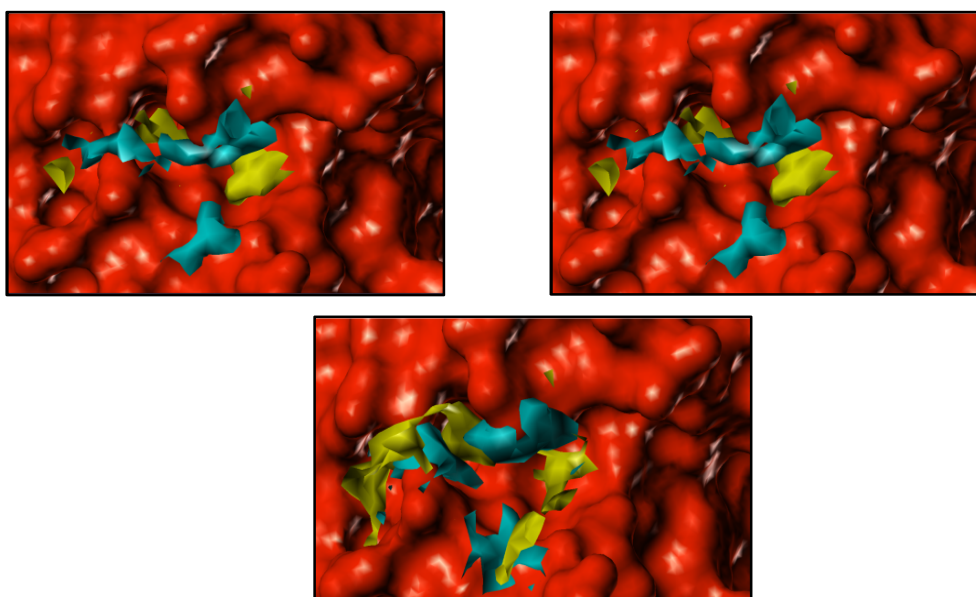


FIGURA 4.8 - Diferença nos pseudo-campos gerados, mostrando as regiões de favorecimento para cada alvo. Sondas: A)C3, B) NM3 e C) OH. (Ciano=Cruzaína e Amarelo=Catepsina B)

A partir da análise das estruturas PDB dos alvos e os campos gerados, foi possível determinar diferenças químicas importantes entre as estruturas das catepsinas e a cruzaina. Além disso, foi feita a identificação dos resíduos de aminoácidos correspondentes a essas diferenças, que estão listados na TABELA 4.3.

TABELA 4.3 - Relação dos aminoácidos correspondentes aos pontos relevantes para a seletividade na cruzaina e os aminoácidos correspondentes nas Catepsinas L e B. (Em negrito são apontados os resíduos com propriedades diferentes)

Sonda	Cruzaína	Catepsina L	Catepsina B
C3	<b>Met68</b>	Met	<b>Tyr</b>
	<b>Leu67</b>	Leu	<b>Gly</b>
	<b>Leu157</b>	Met	----
	<b>Ala131</b>	<b>Ser</b>	<b>Glu</b>
NM3	<b>Leu67</b>	Leu	<b>Gly</b>
	<b>Glu205</b>	<b>Ala</b>	Glu
	<b>Glu112</b>	Asp	<b>Tyr</b>
OH	<b>Ser207</b>	Ser	<b>Val</b>
	<b>Gln156</b>	Asp	<b>Pro/Ala</b>
	<b>Asn69</b>	Asp	<b>Pro/Ala</b>
	<b>Glu205</b>	<b>Ala</b>	Glu

Na FIGURA 4.9 é apresentada a sobreposição das três enzimas e destacados os resíduos de aminoácido considerados importantes para a interação. Para um melhor entendimento, os resultados serão discutidos separados por tipo de sonda.

**C3** (FIGURA 4.7A e 4.8A): Esta sonda foi escolhida por apresentar caráter hidrofóbico e também pelo fator estéreo. Como podemos observar, o modelo apresentou um ponto de seletividade em relação a ambas as catepsinas, na região do sítio da cruzaina próximo aos aminoácidos Leu157, Leu67, Met68 e Ala131, que apresentam cadeia lateral apolar e pertencem ao sub-sítio S2. Nessas mesmas posições, ocorre a presença de aminoácidos como tirosina e glutamato na catepsina B; e serina na catepsina L – que são caracterizados por apresentarem cadeia lateral polar não carregada ou carregada negativamente (Glu). Isso torna então o sub-sítio S2 (e a cavidade próxima a ele), relevante para a seletividade em relação a interações hidrofóbicas do tipo van der Waals.

**NM3** (FIGURA 4.7B e 4.8B): Esta sonda foi escolhida por apresentar grupos hidrofóbicos e ter uma carga positiva. Foi observada uma pequena variação entre os campos gerados por esta sonda e pela sonda C3. Podemos observar um campo de interação gerado próximo ao Glu205 e Glu112 que são aminoácidos que apresentam cadeia lateral polar, porém carregada negativamente. Assim, é possível

determinar um possível ponto de interação do tipo ligação iônica próximo ao sítio ativo.

**OH** (FIGURA 4.7C e 4.8C): Esta sonda se caracteriza por realizar ligação de hidrogênio tanto aceitadora quanto doadora. É possível determinar pontos importantes de interação próximos aos aminoácidos Ser207, Gln156, Asn69 e Glu205. Estes aminoácidos apresentam cadeia lateral polar não carregada ou carregada negativamente com a presença de grupo carboxílico, amida e hidroxila, que são capazes de realizar interações do tipo ligação de hidrogênio. Nas mesmas posições, ocorre a presença de aminoácidos como valina, prolina e alanina, que possuem cadeia lateral apolar, sendo então estes pontos importantes para a seletividade.

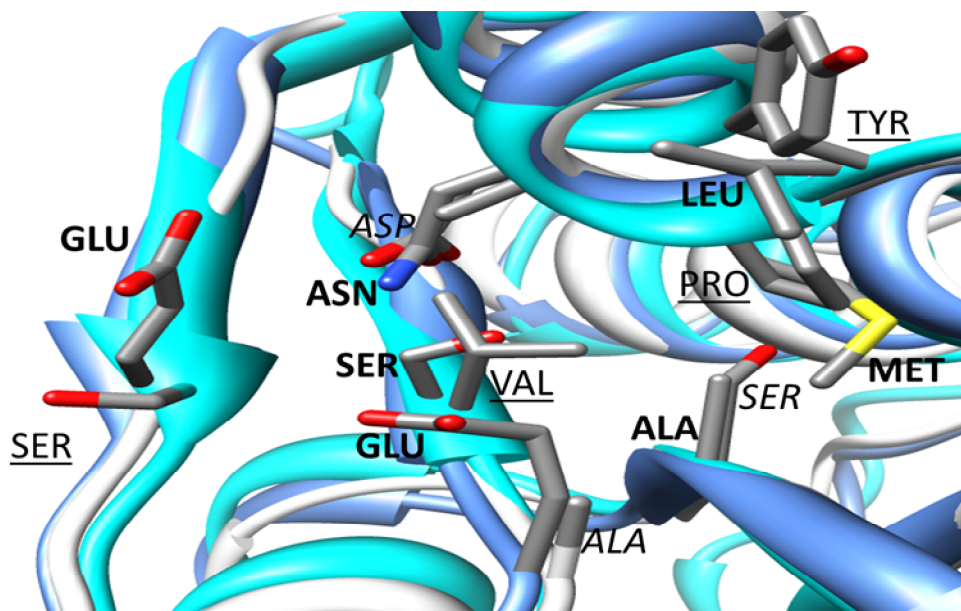


FIGURA 4.9 - Sobreposição das cisteíno-proteases na extensão do sub-sítio S2: **Cruzaína** (azul/negrito), *Catepsina L* (branco/itálico) e Catepsina B (ciano/sublinhado).

# **CAPÍTULO 5**

## **Identificação e Seleção dos Fragmentos**

Neste capítulo é feita uma breve introdução sobre os métodos em quiminformática, as estratégias e critérios usados na seleção dos fragmentos. Além disso, também são apresentados os resultados obtidos e a discussão dos mesmos.

## 5 - Identificação e Seleção dos Fragmentos

Métodos e ferramentas computacionais têm-se tornado importantes para a exploração sistemática do espaço químico-biológico durante as etapas de identificação e otimização de ligantes. A quiminformática possibilitou a introdução de tecnologias avançadas na busca por novos fármacos, evitando assim a síntese de compostos triviais que dificilmente se tornariam novos medicamentos.

### 5.1 - Métodos Computacionais

#### 5.1.1 - Docagem Molecular

A compreensão dos mecanismos de reconhecimento molecular enzima-ligante é um dos principais desafios da biologia molecular e um dos aspectos centrais para o sucesso na descoberta e planejamento de novos fármacos dentro da área de química medicinal. O uso de metodologias computacionais para obter uma descrição acurada e automatizada do processo de reconhecimento molecular enzima-ligante é motivado principalmente pela redução no tempo.

A docagem molecular (do inglês *molecular docking*) é um método baseado na estrutura do receptor (TBVS) e pode ser definida como uma tentativa de prever a presença, a conformação e a orientação da estrutura do complexo formado entre um ligante (uma pequena molécula ou até mesmo uma proteína) e um receptor (enzima, DNA, canais iônicos, receptores, dentre outros).<sup>64</sup> Desta forma, um pré-requisito é a informação tridimensional detalhada da biomacromolécula em estudo, a qual é geralmente conseguida através da análise de estruturas obtidas pela resolução de estruturas de proteínas por raios-X dentre outros como os estudos de espectroscopia de ressonância magnética ou através de modelagem por homologia.<sup>65-66</sup>

Métodos de docagem são amplamente utilizados no planejamento de fármacos baseado na estrutura do alvo, tanto para a descoberta de novas substâncias bioativas, através de técnicas conhecidas como triagem virtual, quanto para o refinamento e otimização de compostos protótipos previamente identificados. O objetivo desse método é encontrar e quantificar o modo de interação/ligação



correto de uma molécula ligante no sítio de interação de uma macromolécula, de maneira que a função da mesma possa ser regulada (inibida ou ativada). Além disso, é possível também identificar novos ligantes usando o ensaio virtual e prever a afinidade de ligação de compostos relacionados a partir de uma série de ligantes ativos.<sup>64,67</sup>

O reconhecimento molecular enzima-ligante depende da complementaridade de características estruturais e físico-químicas da macromolécula com o ligante, visando a identificação do arranjo conformacional preferencial, correspondente à melhor energia livre de interação. As características estruturais determinam os arranjos espaciais moleculares, dado por variações na orientação, posição espacial e rotações de ligações covalentes de ambas as moléculas. As características físico-químicas definem o grau de afinidade e especificidade entre o ligante e o sistema molecular em estudo, dados por interações intermoleculares, que compreendem forças eletrostáticas, de dispersão, interações hidrofóbicas, ligações de hidrogênio e ligações covalentes.<sup>64</sup>

Há três características principais na docagem molecular: i) representação do sistema; ii) busca conformacional no espaço; iii) pontuação de potenciais candidatos. A docagem simula a interação do ligante na superfície da proteína. Portanto, a primeira questão é como definir a superfície da proteína. Elas podem ser descritas por modelos matemáticos como, por exemplo, descritores de formas geométricas ou grades. Alternativamente, a descrição pode envolver um tratamento estático ou dinâmico das cadeias proteicas.

A docagem envolve a análise de duas moléculas separadas, começa pelo enovelamento da proteína e as conformações dos ligantes. Esse processo exige um procedimento eficiente de busca e uma função de pontuação. Os dois elementos críticos para este fim são a velocidade e a eficiência em cobrir um espaço conformacional relevante; é essencial que a função de pontuação seja rápida o suficiente para permitir a obtenção de um grande número de potenciais soluções, e em princípio, discriminar entre as conformações docadas de forma efetiva e não efetiva. A função de pontuação deve incluir um peso apropriado para todos os componentes energéticos da interação. Desta forma, uma solução apresentada por um programa de docagem em particular, não deve ser vista como uma representação completa, mas sim uma peça dentro de um vasto espaço conformacional.

A predição correta da estrutura do complexo (posição) não requer informações a respeito da constante de afinidade. Entretanto, a predição da constante de afinidade (pontuação) requer a informação sobre a posição relativa do ligante na macromolécula. A função de pontuação deve considerar os seguintes fatores: estereo, eletrostático, ligação de hidrogênio, restrição conformacional do ligante (se flexível) e enzima. Considerando o equilíbrio de formação do complexo, há também os fatores de dessolvatação, entropia rotacional e entropia translacional para se levar em conta.

Coleções extensas de compostos podem ser “docadas” computacionalmente na estrutura da macromolécula. Cada composto é testado em milhares de possíveis poses. Como resultado, as moléculas são pontuadas com base na sua complementaridade com a macromolécula e as mais bem pontuadas podem ser subsequentemente avaliadas em um ensaio experimental, FIGURA 5.1.<sup>67</sup>

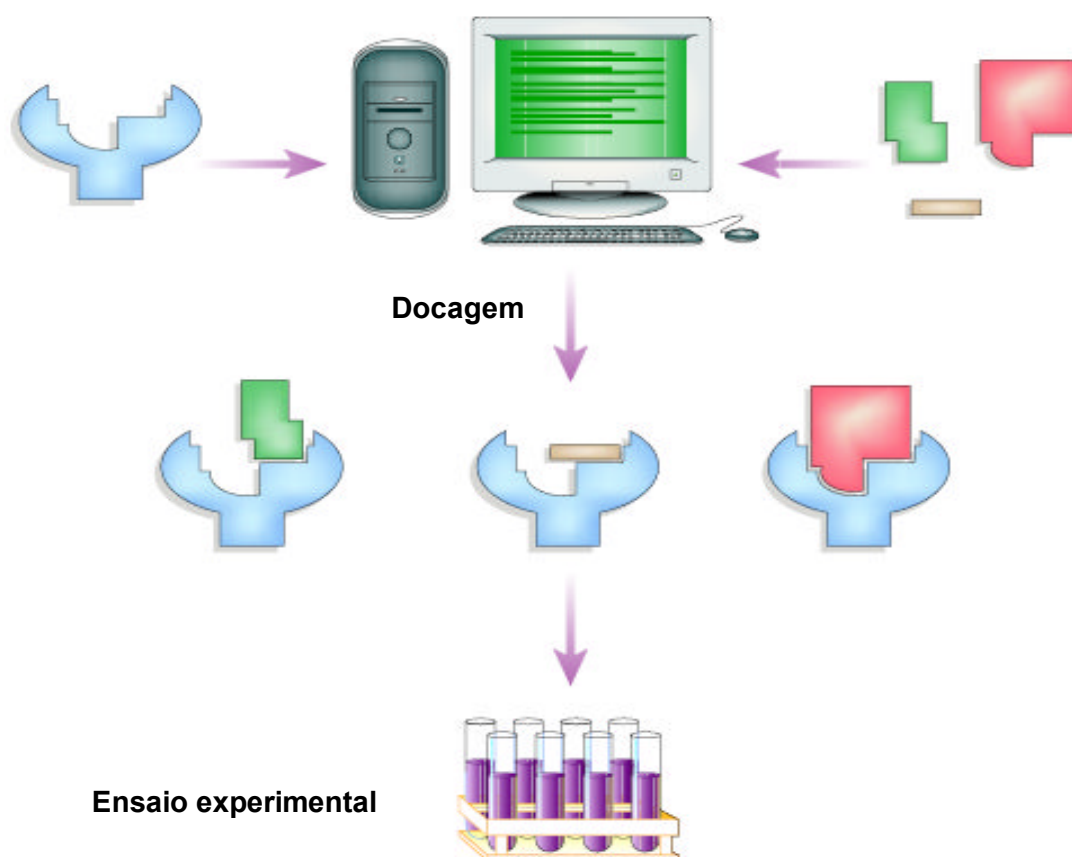


FIGURA 5.1 - Ensaio virtual para identificação de novos ligantes

No processo de reconhecimento molecular, o receptor e o ligante passam por mudanças conformacionais e a consideração da flexibilidade molecular destes implica no tratamento de centenas de milhares de graus de liberdade, por parte dos algoritmos de docagem. Além disso, o reconhecimento molecular é um

processo dinâmico e altamente complexo, envolvendo um grande número de interações intermoleculares entre o ligante, a molécula receptora e o solvente.<sup>64</sup>

Devido a essa complexidade, os protocolos de docagem geralmente são divididos em dois componentes: um algoritmo de busca dos graus de liberdade configuracional e conformacional e uma função de pontuação ou avaliação. O algoritmo de busca deve elucidar todos os possíveis modos de ligação entre o ligante e o receptor, explorando os graus de liberdade translacional e rotacional dos ligantes e do receptor. Usualmente, as funções de pontuação indicam a complementaridade estérea e química entre o ligante e o receptor.<sup>64, 68-69</sup>

As funções de pontuação possuem duas funções, direcionar a posição durante o procedimento de docagem e prever a afinidade de interação da posição final. A maioria dos programas de docagem utiliza modelos simples de funções de energia potencial, geralmente baseados em campos de força da mecânica molecular clássica. Alguns programas utilizam essas funções mais simples durante a fase de execução e posteriormente avaliam as conformações obtidas com funções de pontuação mais sofisticadas, com a inclusão de termos entrópicos. Entretanto, apesar do grande custo computacional, alguns programas também utilizam essas funções de pontuação na fase de busca conformacional, como tentativa de obter predições mais acuradas.

De uma forma geral a predição da afinidade é utilizada para listar de maneira ordenada as moléculas com maiores chances de apresentar atividade biológica. As funções de pontuação são classificadas em funções baseadas no campo de força, no conhecimento e empírica e conforme descritas na próxima seção.

### **5.1.1.1 - Funções de pontuação**

#### **5.1.1.1.1 - As funções baseadas em campo de força**

Um campo de força é uma representação física de um sistema em simulações computacionais que consiste de uma forma funcional e de um conjunto de parâmetros de ajuste às suas respectivas funções. Os campos de força moleculares atualmente mais utilizados são GROMOS<sup>70</sup>, AMBER<sup>71</sup> e CHARMM<sup>72</sup>. A maioria dos campos de força divide a função de energia potencial em termos de

contribuições entre átomos ligados e entre átomos não-ligados. Nos métodos de simulação envolvendo macromoléculas biológicas, a função avaliação é geralmente um campo de força molecular clássico. Entretanto, os programas de docagem, na maioria dos casos, fazem uso apenas de um termo descrevendo a rotação de uma ligação química no ligante e de termos descrevendo as interações não-ligadas para cálculo da interação entre receptor e a molécula ligante.

#### **5.1.1.1.2 - As funções empíricas**

Os métodos semi-empíricos e empíricos para cálculo de energia livre de ligação utilizam uma função cujos coeficientes foram pré-otimizados tomando-se como base resultados experimentais de estruturas receptor-ligante e suas respectivas constantes de inibição. Estas funções reproduzem dados experimentais associados à energia livre de interação, assumindo que esta pode ser decomposta como uma soma de diversas funções básicas que dependem das coordenadas do ligante e do receptor. Dentre as funções de pontuação empíricas, a mais comumente utilizada é ChemScore, presente em vários programas de docagem, como por exemplo GOLD e FRED. As funções de pontuação empíricas incluem contribuições não entálpicas assim como termos de rotação, as quais incluem penalidades entrópicas para a interação através da atribuição de pesos na soma do número de ligações rotacionáveis nos ligantes.<sup>64, 73</sup>

#### **5.1.1.1.3 - As funções baseadas no conhecimento**

As funções de pontuação baseadas no conhecimento procuram utilizar as informações derivadas do número cada vez maior de estruturas de complexos determinados experimentalmente. Nessa análise, um conjunto de ligantes e fragmentos de proteínas identificados é utilizado para listar as distâncias entre os fragmentos de todas as estruturas e compilá-los. Assim como nos métodos empíricos, essas funções tentam obter implicitamente efeitos da ligação que são difíceis de modelar explicitamente. DrugScore<sup>74</sup> e SMOG<sup>75</sup> são exemplos de implementações de funções baseadas em conhecimento. A maior vantagem dessas funções de pontuação é sua simplicidade computacional, o que permite testar grandes coleções de compostos de maneira eficiente. Por outro lado, possui a

desvantagem de sua derivação ser essencialmente baseada em informações implícitas, extraídas de conjuntos limitados de estruturas experimentais de complexos proteína-ligante.

### 5.1.1.2 - Algoritmos de busca e flexibilidade molecular

Durante o reconhecimento molecular receptor-ligante as moléculas podem sofrer mudanças conformacionais, ou seja, podem alterar a forma para melhor se adequarem umas às outras. Embora a consideração da flexibilidade molecular seja uma característica importante para a representação da realidade em um algoritmo de docagem, muitas vezes isto se torna inviável devido principalmente ao grande número de graus de liberdade da molécula receptora. Os primeiros algoritmos de docagem tratavam tanto receptor quanto ligante como moléculas rígidas, considerando apenas os graus de liberdade translacionais e rotacionais da molécula ligante. Atualmente, a maioria dos programas de docagem inclui a flexibilidade do ligante considerando, além dos graus de liberdade translacionais e rotacionais, os graus de liberdade conformacionais (ângulos diedros associados às ligações covalentes simples conforme demonstrado na FIGURA 5.2). Os algoritmos mais amplamente usados são os de busca sistemática e os de busca estocásticos.<sup>64</sup>

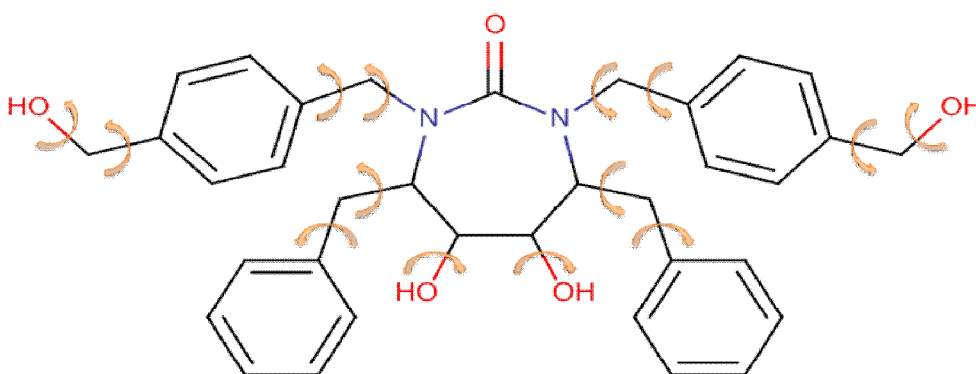


FIGURA 5.2 - Graus de liberdade conformacionais (setas) de um ligante

#### 5.1.1.2.1 - Algoritmo de busca sistemática ou construção incremental

Estes algoritmos exploram todos os graus de liberdade de uma molécula, mas enfrentam o problema da “explosão combinatória”. Para evitar uma

busca exaustiva, muitos algoritmos de busca usam um método de construção incremental, crescendo o ligante no sítio ativo, que consiste basicamente de três passos: seleção do fragmento principal, ajuste posicional do fragmento principal e crescimento incremental do ligante. No primeiro passo, o ligante é dividido em um fragmento rígido com cadeias laterais flexíveis; subsequentemente, essas cadeias laterais são divididas em cada nova ligação rotacional, no segundo e terceiro passos, o fragmento é colocado e as cadeias laterais são ligadas incrementalmente, considerando-se os graus de liberdade rotacionais possíveis. O crescimento incremental pode ser realizado de diferentes modos. Um método é docar vários fragmentos moleculares na região do sítio de interação da enzima e em seguida conectá-los covalentemente (método *de novo*) ou, alternativamente, dividir os ligantes docados em partes rígidas e flexíveis. Neste caso, uma vez que o fragmento rígido é definido, ele é docado do sítio ativo, em seguida as regiões flexíveis são adicionadas incrementalmente. Exemplos de programas que utilizam este algoritmo são DOCK, FlexX, FLOG, Hammerhead e Surflex.<sup>76</sup>

#### 5.1.1.2.2 - Algoritmos estocásticos

O algoritmo estocástico faz uma busca aleatória para um ligante ou mesmo para uma população de ligantes. Cada novo complexo de interação entre proteína e ligante obtido é avaliado baseado numa função de probabilidade pré-definida. Os dois métodos mais amplamente utilizados são Monte Carlo e algoritmo genético.<sup>77</sup>

O algoritmo de Monte Carlo possui basicamente a seguinte forma: geração da conformação inicial do ligante no sítio ativo consistindo de uma forma aleatória de conformação, e, em seguida, tomando essa conformação como referência, geração de uma nova conformação. Se a conformação gerada possuir energia menor que a conformação de referência, a nova conformação é imediatamente aceita e tomada como referência para a próxima iteração. Se a conformação gerada possuir energia maior que a conformação de referência, o critério Metrópolis é usado para definir se a nova conformação é aceita ou não. Esse processo é repetido até que o número desejado de conformações seja obtido. O critério de Metrópolis consiste em gerar um número randômico entre 0 e 1 e compará-lo com o fator de Boltzmann,  $e^{-\Delta V/k_B T}$ . Se o fator de Boltzmann for maior

que o número randômico gerado, então, a nova conformação é aceita, caso contrário é rejeitada. Exemplos de programas que utilizam este algoritmo são DockVision, ICM, MCDOCK, ProDOCK e SLIDE.<sup>78-79</sup>

Os algoritmos genéticos são baseados na teoria da evolução de Darwin. A pose do ligante é representada por um cromossomo constituído de genes. Os genes codificam cada ângulo torsional, rotacional e translacional do ligante. As poses envolvem transmissão da informação dos genes (reprodução), alterada por operações genéticas como cruzamento e mutação. Existem vários métodos de seleção da próxima geração, mas, a mais utilizada é a sobrevivência da pose de melhor complementaridade, onde duas conformações com melhor energia passam para a próxima geração. O algoritmo possui basicamente os seguintes passos: faz pequenas mudanças aleatórias na conformação inicial, pontua cada mudança de acordo com o valor da função de pontuação, determina quais mudanças são favoráveis à interação. Se a melhor modificação tem um valor menor de energia é aceita, e cada nova conformação é reportada com sua pontuação. Exemplos de programas são AutoDock, DARWIN, DIVALI, GOLD, EADock, FITTED e PSI-DOCK.<sup>79</sup>

### **5.1.2 - Simulação de Dinâmica Molecular**

Dinâmica Molecular (DM) é um método de simulação muito utilizado para o entendimento dos mais diversos fenômenos físicos. É uma técnica utilizada para computar propriedades de equilíbrio e transporte de um sistema clássico de muitos corpos, simulando os movimentos de partículas de um sistema atômico-molecular ao longo do tempo. Na DM, o comportamento dinâmico de um sistema biomolecular é obtido através das equações de movimento de Newton para cada átomo do sistema. Dada a energia como função das coordenadas atômicas, as forças que atuam sobre cada átomo são calculadas através do gradiente da energia potencial em relação à posição dos átomos. Em cada passo de tempo da dinâmica, as equações de movimento são integradas numericamente, num processo iterativo, obtendo-se as energias e trajetórias para todas as partículas (átomos) e para o sistema como um todo. O resultado é a trajetória que especifica como as posições e velocidades das partículas no sistema variam com o tempo.<sup>80</sup>

Uma característica marcante do método DM é o estabelecimento de forte sinergia com estudos experimentais. Os resultados das investigações experimentais, geralmente motivam e guiam os estudos de DM. Os resultados obtidos pela dinâmica, por sua vez, fornecem explicações detalhadas a nível microscópico dos fenômenos observados experimentalmente.

Um dos maiores desafios para a realização de simulações de dinâmica molecular é a parametrização dos campos de força. Nem sempre o ajuste de curvas de energia potencial obtidas por cálculos de estrutura eletrônica de pequenos grupos de moléculas é satisfatório. No caso de moléculas como as proteínas, a parametrização é ainda mais complexa.

A qualidade do campo de força depende, evidentemente, da complexidade do sistema estudado. Desta forma, a precisão dos resultados obtidos usando um destes campos de força também depende desta complexidade. Porém, uma propriedade destes sistemas de grande complexidade é que grande parte da dinâmica é dominada por fatores coletivos, como o empacotamento das moléculas, e desta forma os erros associados aos pares de átomos acabam sendo minimizados.

A escolha dos potenciais de interação intra- e intermolecular constitui uma etapa essencial para a descrição correta do sistema em estudo, já que são estes potenciais que determinam as forças atuantes em cada partícula e, conseqüentemente, determinarão como o sistema irá evoluir no tempo, para gerar as trajetórias para análise. De um modo geral, assume-se que o potencial entre duas moléculas pode ser descrito como uma soma de interações entre cada par de átomos ou sítios de interação. No caso de movimentos internos na molécula serem importantes, deve-se somar também os potenciais de interações intramoleculares, relacionadas à deformação da geometria molecular.

Assim, o campo de força é definido como o conjunto de parâmetros necessários para descrever todas essas interações. Existem muitos tipos de campos de força desenvolvidos por diferentes grupos como já comentado anteriormente. Entre os mais conhecidos destacam-se o OPLS (*Optimized Potentials for Liquid Simulations*), com parâmetros apropriados para proteínas, lipídeos, sacarídeos, entre outros. Outros campos de força bem estabelecidos e usados para biomoléculas em geral são AMBER, o GROMOS e o CHARMM.



Para simular um sistema de partículas através de DM são necessárias algumas técnicas específicas, para cada etapa do processo de descrição das posições e velocidades das partículas a cada instante. Dentre estas etapas estão: geração da configuração inicial das moléculas; cálculo das forças exercidas sobre cada partícula; movimento das partículas e análise das trajetórias.

Embora os métodos de dinâmica molecular permitam que todos os graus de liberdade do complexo receptor-ligante sejam explorados, esses métodos frequentemente ficam presos em mínimos locais de energia, devido à impossibilidade de atravessarem altas barreiras de energia em um tempo de simulação viável computacionalmente na temperatura de simulação. Para lidar com essa limitação, algumas estratégias têm sido desenvolvidas, como, por exemplo, aumentar a temperatura da simulação<sup>65</sup>, suavizar a superfície de energia potencial<sup>81</sup>, simular diferentes partes do sistema receptor-ligante com diferentes temperaturas<sup>82</sup>, além de iniciar os cálculos de dinâmica com diferentes posições iniciais da molécula ligante.

## 5.2 - Ensaio de inibição

O ensaio cinético padrão utilizado para a determinação da atividade das enzimas cruzaína e catepsina L é a fluorescência. O mecanismo fundamental de fluorescência é a emissão de fótons quando a molécula excitada no estado singleto retorna ao estado fundamental. O processo de emissão e excitação no fenômeno de fluorescência é muito rápido com tempo de vida na ordem de nanossegundos. Desta forma, é possível observar facilmente fluorescência à temperatura ambiente e diretamente em solução, o que torna este procedimento experimental bastante simples.

O método de fluorescência tem sido amplamente utilizado para a determinação da cinética das enzimas do tipo cisteíno-proteases. Zimmerman e colaboradores sintetizaram a primeira classe de substratos fluorogênicos para as proteases tripsina, elastase e quimotripsina. Esses substratos eram amidas de acil aminoácidos ou peptídeos com o substituinte 7-amino-4-metilcumarina (MCA). A FIGURA 5.3 ilustra a estratégia utilizada para a determinação da formação de produto catalisada pelas enzimas.<sup>83</sup>

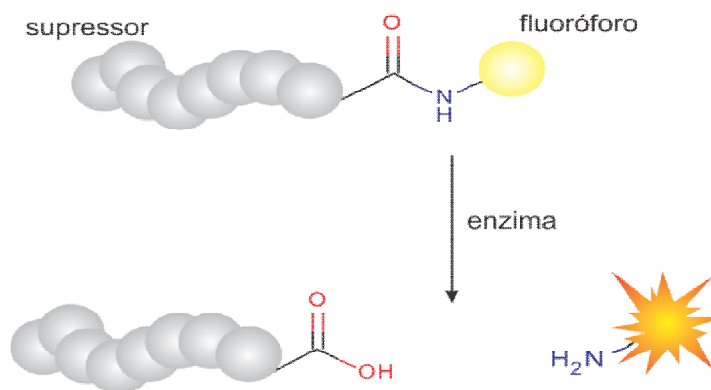


FIGURA 5.3 - Estratégia utilizada para gerar o sinal de fluorescência em reações catalisadas por cisteíno-proteases

Aos grupos terminais carbóxi e amino de um dipeptídeo são ligados a um grupo fluorescente e supressor, respectivamente. Após a quebra da ligação peptídica pela enzima, o grupo fluorescente é liberado em solução permitindo a detecção da formação do produto e determinação da cinética de reação catalisada pela enzima.<sup>17</sup>

Diversos substratos fluorogênicos foram utilizados como substratos para a enzima cruzaina. O substrato estabelecido como padrão para as análises cinéticas de novos inibidores da enzima foi o N-benzóxicarbonil-L-fenilalanina-L-arginina-7-amido-metil-cumarina (Z-Phe-Arg-MCA ou ZFRMCA), FIGURA 5.4.

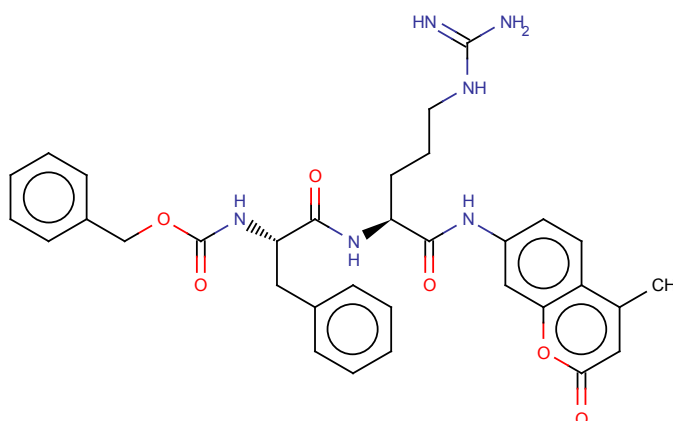


FIGURA 5.4 - Estrutura do ZFRMCA (substrato sintético para a cruzaina)

Os valores das constantes cinéticas da enzima cruzaina reportadas na literatura para este substrato são:  $K_M$  entre 0,96 e 1,4  $\mu\text{M}$ ,  $k_{cat}$  entre 3,9 e 11,6  $\text{s}^{-1}$  e  $k_{cat}/K_M$  entre 4677 e 9667  $\text{mM}^{-1} \text{s}^{-1}$ . Os valores das constantes cinéticas para a

enzima catepsina L são:  $K_M$  entre 1,2 e 2,4  $\mu\text{M}$ ,  $k_{cat}$  entre 10 e 25,8  $\text{s}^{-1}$  e  $k_{cat}/K_M$  entre 4166 e 11700  $\text{mM}^{-1} \text{s}^{-1}$ .<sup>84-90</sup>

### 5.3 - Materiais e métodos

A estratégia adotada para a seleção dos fragmentos consistiu, inicialmente, na aplicação de filtros de seleção molecular. Um filtro para selecionar compostos fragmentos-similares (que seguem a “regra dos três”) e outro contendo grupos reconhecidos pela cruzaina, utilizando padrões SMARTS<sup>91</sup>.

Com o banco montado, os fragmentos foram então submetidos à docagem. Os programas de docagem molecular diferem na maneira de lidar com a flexibilidade da enzima e do ligante, seu algoritmo e função de pontuação.<sup>77</sup> Neste trabalho, o programa de docagem escolhido foi o Glide v5.5.<sup>92-93</sup> O Glide busca por interações favoráveis entre uma ou mais moléculas ligantes tipicamente pequenas e uma molécula alvo tipicamente maior, geralmente uma proteína. Cada ligante deve ser uma única molécula, enquanto o receptor pode incluir mais de uma molécula, por exemplo, uma proteína e um cofator.

O programa pode ser executado nos modos de docagem rígido ou flexível, este último gera automaticamente conformações para cada ligante de entrada. A combinação de posição e orientação de um ligante em relação ao receptor, juntamente com sua conformação na docagem flexível, é referido como uma posição do ligante. Nele, uma série hierárquica de filtros para a busca de possíveis posições do ligante na região do sítio de interação da macromolécula é usada. A forma e as propriedades do receptor são representadas em uma grade por diferentes conjuntos de campos que fornecem progressivamente cada vez mais precisa pontuação da posição do ligante.<sup>92-93</sup>

Os filtros iniciais testam o ajuste espacial do ligante no sítio de interação definido, e examina a complementaridade das interações enzima-ligante através de um método padronizado baseado em grid após a função empírica ChemScore.<sup>94</sup> Posições que passam por esses testes iniciais entram na fase final do algoritmo, que envolve avaliação e minimização do ligante usando uma função de padrão de energia mecânica molecular, neste caso, o campo de força OPLS-AA (AA = *all atom*, ou seja, inclui todos os átomos explicitamente) em conjunto com um modelo dependente da distância. Um pequeno número dos ligantes melhores

pontuados é então submetido a um procedimento de Monte Carlo, no qual mínimos locais nas proximidades conformacionais são examinados. Finalmente, as posições minimizadas são repontuadas usando a função de pontuação GlideScore, que é baseada na ChemScore. Essa função possui um módulo de precisão extra, o qual possui penalidades para poses que violam princípios físico-químicos bem estabelecidos como grupos polares ou com carga serem expostos ao solvente. Os objetivos principais do método XP são eliminar falsos positivos e proporcionar uma melhor correlação entre as posições e as pontuações.<sup>95</sup>

Uma importante facilidade implementada no Glide é a possibilidade da seleção de restrições para aplicar na macromolécula. O Glide incorpora a satisfação dessas limitações em vários dos seus filtros hierárquicos, permitindo uma rápida rejeição das estruturas docadas que não cumpram os requisitos.<sup>92-93</sup>

A docagem foi feita seguindo as seguintes etapas: 1<sup>a</sup>) preparação da estrutura da enzima; 2<sup>a</sup>) geração dos multiconfômeros, utilizando o programa Omega v2.4.2<sup>96</sup>; 3<sup>a</sup>) docagem utilizando o módulo XP do programa Glide v5.5<sup>92</sup>. Para obtenção de uma estrutura relaxada no sub-sítio S2, foi extraída do PDB a estrutura cristalográfica da cruzaina sob código 1ME4 e alinhada no programa Sybyl8.0<sup>97</sup> com a estrutura da catepsina L sob código 3HWN. A catepsina L e o ligante T10 da cruzaina foram excluídos, permanecendo então apenas a estrutura da cruzaina com o ligante BD3 - (2S)-2-[(3-tert-butil-1-metil-1H-pirazol-5-il)formamido]-3-[3-(5-carboximidoil-1,3,4-oxadiazol-2-il)phenil]-N-(2-iminoetil)propanamida. O BD3 foi então minimizado no sítio da cruzaina. Esse ligante foi escolhido, pois, como dito no capítulo anterior, na catepsina L ele ocupa a região adjacente ao sub-sítio S2, onde estão localizados os resíduos flexíveis da cruzaina.

A estrutura 1ME4 minimizada foi então preparada no Receptor Grid Generation do programa Glide: foram removidas as moléculas de água, ajustada a ordem de ligação e a carga formal do ligante. O sítio foi definido tendo como referência a estrutura do ligante BD3, no sub-sítio S2 da enzima. Foram configuradas também restrições de ligação de hidrogênio e de interações hidrofóbicas. Paralelamente, o banco refinado de fragmentos foi submetido ao programa Omega v2.4.2<sup>96</sup>. Ele foi utilizado para a construção da coleção multiconformacional, partindo de uma coleção de fragmentos padrão incorporada ao programa.

Os parâmetros utilizados para gerar a coleção multiconformacional foram os padrões do OMEGA2, onde os parâmetros de construção 3D foram: o campo de força utilizado foi o mmff94s\_NoEstat (Merck Molecular Force Field - inclui todos os termos do mmff94s, exceto interações coulômbicas); parâmetros de “torsão” das ligações: janela de energia utilizada  $10 \text{ kcal mol}^{-1}$ , desvio quadrático médio 0,5, número máximo de conformêmeros 200. Os multiconfômeros foram então docados no módulo XP do programa Glide v5.5<sup>95</sup>.

As moléculas com maior pontuação foram analisadas pela inspeção visual da docagem, sendo que as selecionadas foram submetidas a estudos de dinâmica molecular, utilizando o programa Amberv11<sup>98</sup>, campo de força mmff99sb. Como estrutura de início da dinâmica foi utilizado o complexo enzima-fragmento extraído da docagem.

Como o complexo possui moléculas que são diferentes de resíduos de aminoácidos, o primeiro passo foi parametrizar o ligante. Para isso, foi calculada computacionalmente a carga parcial dos átomos do fragmento já na conformação em que se encontra no complexo com a cruzaína. O programa utilizado para essa etapa foi o Gaussian<sup>99</sup>, com o método/função de base HF/6-31G\* para o cálculo de cargas.

Após isso, toda uma etapa foi feita de modo a colocar os valores de carga, posição e tipo dos átomos dessa molécula de um modo que o programa AMBER fosse capaz de ler. Para isso, foram utilizados os programas Antechamber e tLeap. Os arquivos de parâmetros para os fragmentos foram salvos para o programa reconhecer tal molécula quando computar o complexo enzima-fragmento.

Com o fragmento devidamente parametrizado, de modo que o tornasse possível de ser reconhecido, foi feita a preparação da estrutura do complexo. Para isso, a estrutura saída da docagem teve os hidrogênios retirados (para evitar conflito na nomenclatura de átomos de hidrogênio) e o estado de protonação das histidinas corrigido (manualmente). Com o tLeap, o complexo foi carregado, e antes de colocar água e íons, teve seus parâmetros salvos. Após isso, o sistema foi neutralizado adicionando íons de sódio, e então foi adicionada uma caixa octaédrica truncada de água, do tipo TIP3, de 10 Å em volta do sistema.

O processo de dinâmica foi dividido em basicamente duas etapas: a pré-dinâmica, consistindo de uma minimização, aquecimento do sistema, equilibração pós-aquecimento; e a dinâmica propriamente dita.

Na primeira etapa, inicialmente foi feita uma minimização para reacomodar os átomos do sistema de forma que os mesmos não assumissem alta energia cinética durante o aquecimento, devido a repulsões eletrostáticas e contatos estéreos. O segundo passo foi aquecer o sistema de 0 K (os átomos inicialmente estão com velocidade nula) a 310 K durante 50 ps, temperatura em que a dinâmica ocorreu. Então, o sistema foi deixado a volume constante por 300 ps, para que a densidade do complexo e das águas pudessem se estabilizar. Após isso, foi feita a equilíbrio, durante 300 ps, na verdade uma pré-dinâmica, para verificar se o sistema estava estável para se iniciar a dinâmica propriamente dita.

Na segunda etapa, a dinâmica, o sistema foi submetido ao processo NpT, à pressão constante, 310 K de temperatura, durante um período de 8 ns. Esse tempo é possivelmente suficiente para que o sistema se estabilize ou que o fragmento saia do sítio ativo. Após isso, o sistema foi submetido a um estudo de anelamento por 170 ps (FIGURA 5.5).

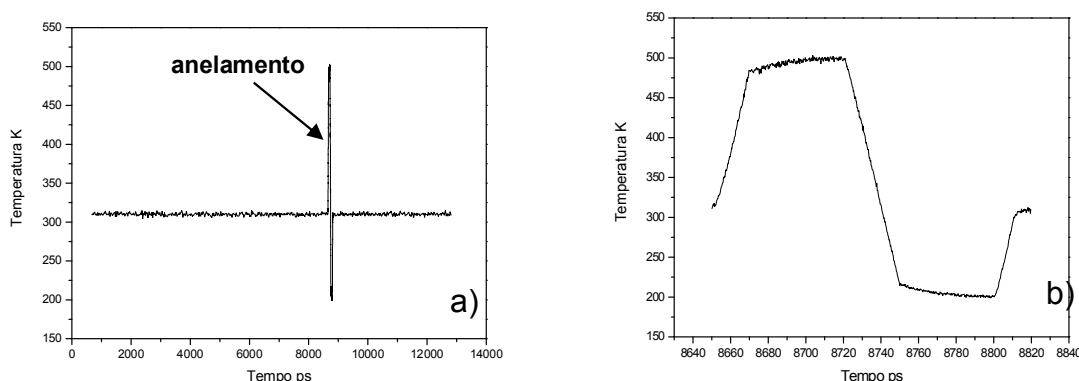


FIGURA 5.5 – a) Variação da temperatura durante a DM. b) Etapas de aquecimento e resfriamento durante o anelamento.

O anelamento seguiu o seguinte procedimento: aquecimento até 500K em 20 ps, equilíbrio a temperatura constante por 50 ps, resfriamento a 200K em 30ps, equilíbrio por 50 ps, e aquecimento até a temperatura inicial, em 20 ps. Em seguida, o sistema foi novamente submetido ao sistema NpT, à pressão constante, 310 K de temperatura, durante um período de 4 ns. Caso ao fim desses 4 ns o sistema estivesse passando por alguma transformação, ou tivesse acabado de passar por uma, esse tempo era estendido, de modo a acompanhar essa mudança.

A DM foi utilizada para obter informações sobre a estabilidade do fragmento no sítio e a manutenção das ligações de hidrogênio definidas como obrigatórias no processo de docagem. Os fragmentos selecionados pela DM foram

então submetidos a ensaios de inibição. Os ensaios foram realizados no Laboratório de Biotermodinâmica de Fármacos do Grupo de Estudos em Química Medicinal de Produtos Naturais – NEQUIMED-PN, do Instituto de Química de São Carlos (IQSC/USP). O equipamento utilizado é da empresa Biotek, modelo Synergy HT Microplate Reader. As condições experimentais para as enzimas foram:

- Cruzaína - tampão fosfato de sódio 100 mM em pH 6,3, 100 mM de NaCl, 10 mM EDTA, 5,0 % v/v de DMSO, 5,0 mM de DTT e Triton X-100 0,01 % v/v.
- Catepsina L - tampão acetato de sódio 100 mM em pH 5,5, 100 mM de NaCl, 10 mM EDTA, 5,0 % v/v de DMSO, 5,0 mM de DTT e Triton X-100 0,01 % v/v.

As concentrações das enzimas utilizadas foram 2,0 nM e do substrato ZFRMCA 2,0  $\mu$ M. Os valores de concentração para inibir 50% da atividade enzimática ( $IC_{50}$ ) foram determinados através da curva dose-resposta. As concentrações dos fragmentos nos ensaios variaram de 4,0mM a 4,0 $\mu$ M. Estas concentrações, do mais concentrado para o menos concentrado, foram obtidas segundo um esquema de diluição serial na proporção de diluição 2 vezes. O procedimento utilizado para ambas as enzimas foi incubar as mesmas na solução tampão contendo DTT por 5 minutos. Em seguida, o inibidor foi incubado também por 5 minutos antes de adicionar o substrato. A hidrólise do substrato foi acompanhada por fluorescência com a excitação em 355 nm e emissão 460 nm. A velocidade máxima foi determinada considerando a velocidade inicial da hidrólise. Todos os experimentos foram feitos em triplicata para avaliação do erro experimental. O valor do  $IC_{50}$  foi calculado através de curva dose-resposta.

Os reagentes fosfatos de sódio monobásico e dibásico, ácido etilenodiamino tetra-acético (EDTA), dimetilsulfóxido (DMSO), Triton-x100, acetato de sódio, ZFRMCA, enzima catepsina L de fígado humano foram adquiridos da empresa Sigma-Aldrich. O reagente Ditiotreitól (DDT) foi adquirido da empresa USB Corporation com grau de pureza “ultra-puro”.

As substâncias obtidas para os ensaios de inibição foram adquiridas das empresas Otava e Enamine. Todos os reagentes têm certificação do fabricante de pureza > 99 %. Nenhuma purificação/avaliação de integridade adicional foi realizada.

A enzima cruzaina utilizada neste trabalho foi expressa e purificada pelo Dr. Helton José Wiggers, no Laboratório de Biologia Molecular do NEQUIMED do Instituto de Química de São Carlos (IQSC), Universidade de São Paulo (USP).

## 5.4 - Resultados e Discussão

O banco de dados utilizado para iniciar o estudo foi obtido no banco virtual de moléculas comerciais ZINC<sup>100</sup>, contendo 2.240.335 estruturas químicas. Esse valor representa cerca de 4 % de todas as moléculas registradas no *Chemical Abstracts*<sup>41</sup>, que corresponde a aproximadamente 60 milhões de compostos. Em seguida foram aplicados filtros de seleção no banco utilizando o programa FILTER 2.0.2<sup>101</sup>. Esse programa é utilizado para remover compostos indesejáveis utilizando uma combinação de cálculos de propriedades físicas e conhecimento de grupos funcionais.

Um filtro para selecionar compostos fragmentos-similares foi utilizado. Esse filtro foi criado baseado na regra dos três ( $MW \leq 300$ ,  $clogP \leq 3$ , doadores de ligações de hidrogênio  $\leq 3$ , aceptores de ligação de hidrogênio  $\leq 3$ , área superficial polar  $\leq 60 \text{ \AA}^2$ <sup>47</sup>). Contudo, foi permitida uma flexibilização na regra, podendo ser violados até três parâmetros, no intuito de permitir que esqueletos privilegiados presentes em vários fármacos da terapêutica pudessem constar do banco de fragmentos final. Isso se fez necessário, pois vários desses esqueletos moleculares apresentam um ou mais átomos capazes de realizar interação como ligação de hidrogênio.<sup>48-49</sup> Assim, esses heteroátomos presentes no heterociclos diminuiriam a possibilidade de substituintes com característica hidrofílica estarem incluídos no banco final. Após essa etapa, o banco de fragmentos-similares ficou com 76.800 moléculas. Buscando refinar o banco gerado, uma inspeção nos inibidores co-cristalizados no sítio de interação da cruzaina foi realizada para se retirar informações acerca de grupos de reconhecimento pela enzima nesses ligantes.

Na FIGURA 5.6 estão representados os grupos que se repetiram nas estruturas da cruzaina analisadas. Das 19 estruturas, 11 apresentaram o grupo fenila no sub-sítio S2, 16 apresentam a amida realizando interações do tipo ligação de hidrogênio com os resíduos Gly66 ou Asp158. Esses grupos foram então adotados como âncoras, ou seja, grupos que são reconhecidos pela enzima para a montagem de filtros moleculares utilizando padrões SMARTS.



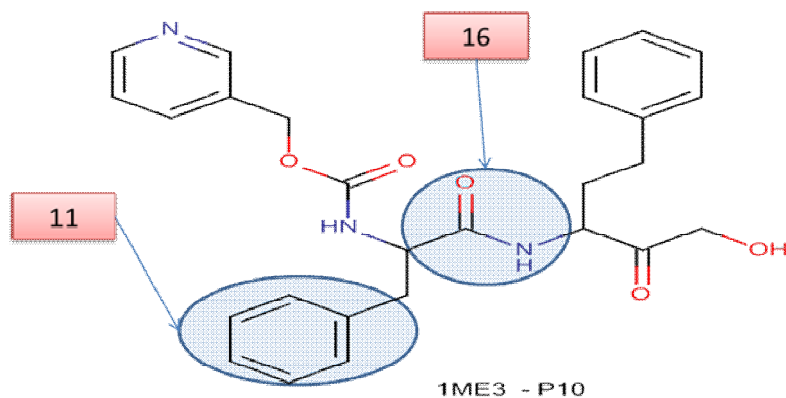


FIGURA 5.6 – Representação do ligante P10, extraído do PDB: 1ME3. Em destaque os grupos reconhecidos pela enzima cruzaina

Em trabalhos anteriores, já havia sido utilizada a mesma hipótese de que ciclos aromáticos ou alifáticos poderiam atuar como grupo de reconhecimento no sub-sítio S2, além de se perceber a importância da amida próximo à Gly66.<sup>18-19, 23, 102</sup>

A SMARTS (SMiles ARbitrary Target Specification) é uma linguagem que permite especificar subestruturas usando regras que são simples extensões do SMILES. Na linguagem SMARTS são incluídos operadores lógicos e descritores moleculares adicionais.<sup>91</sup>

O programa FILTER 2.0.2<sup>101</sup> determina se uma estrutura satisfaz uma série de requisitos subestruturais definidos por notação SMARTS. Cada requisito deve ser respeitado e os ganhos de desempenho podem ser alcançados garantindo-se que os mais facilmente testados e mais restritivos requerimentos sejam testados primeiro. A especificação de requisitos subestruturais é mais bem ilustrada com exemplo (FIGURA 5.7). O filtro pode ser usado para pesquisar a ocorrência de um determinado padrão subestrutural em um banco de dados de estruturas.<sup>48</sup>

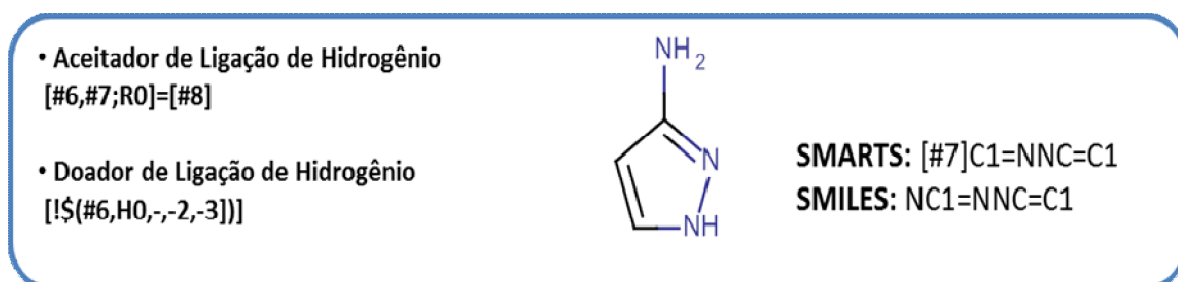


FIGURA 5.7 – Exemplo de padrões SMARTS

Os filtros com padrões SMARTS foram utilizados para fazer a busca mais focada no banco fragmento-similar, resultando em 227 (duzentos e vinte e sete) fragmentos que foram então docados. A representação em SMILES do banco está no Apêndice.

A docagem foi feita seguindo as seguintes etapas: 1<sup>a</sup>) preparação da estrutura da enzima; 2<sup>a</sup>) geração dos multiconfôrmeros, utilizando o programa Omegav2.4.2<sup>96</sup>; 3<sup>a</sup>) docagem utilizando o módulo XP do programa Glide v5.5<sup>92</sup>. Inicialmente o ligante BD3 foi minimizado no sítio da cruzaina, o que resultou na abertura do sub-sítio S2 conforme a FIGURA 5.8.

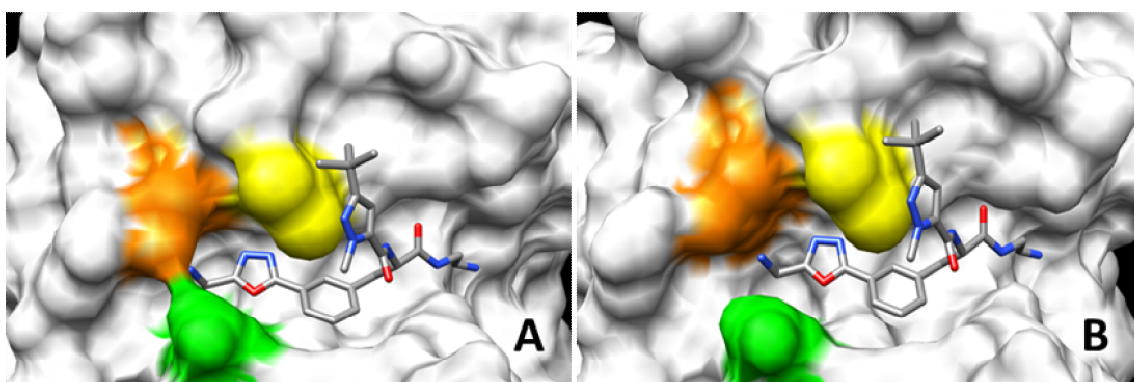


FIGURA 5.8 – Sub-sítioS2 da cruzaina: a) sobreposto com o ligante BD3 da catepsina L (PDB 3HWN); b) após a minimização com o mesmo ligante

A estrutura 1ME4 minimizada foi então preparada no *Receptor Grid Generation* do programa Glide. O banco contendo 227 moléculas foi submetido ao programa Omega v2.4.2<sup>96</sup> para geração de multiconfôrmeros e então docado no módulo XP do programa Glide v5.5<sup>95</sup>, usando as restrições de ligação de hidrogênio nos resíduos Glu205 e Ser207; e restrições de interações hidrofóbicas próximas aos resíduos Met68, Leu67, Leu157 e Ala131. Estes resíduos foram identificados nos estudos de mapeamento do receptor, descrito no capítulo anterior. Para serem aprovados, os fragmentos deveriam cumprir ao menos um requisito na restrição relacionada à ligação de hidrogênio, e cumprir o requisito da interação hidrofóbica.

Após a docagem, os fragmentos mais bem pontuados foram submetidos à análise visual. Esta é uma das etapas mais importantes durante o processo, já que é nela que se faz a seleção dos compostos que prosseguirão no planejamento. Nesta etapa exige-se do químico medicinal o uso da intuição química e conhecimento do sistema para decidir quais fragmentos apresentam as características mais adequadas e que justifiquem a sua seleção. São avaliados,

entre outros fatores, a complementaridade química pelo sítio, interações e energia de interação. Após esta etapa, foram selecionados 12 fragmentos para estudos de dinâmica molecular. Na FIGURA 5.9 são apresentadas as posições de menor energia obtidas para os fragmentos selecionados, e na TABELA 5.1 são apresentados os valores de pontuação obtidos (nesse caso, quanto menor o valor, melhor é a energia de interação).

Para exemplificar o modo de interação obtido, na FIGURA 5.10 é apresentado um dos fragmentos que seguiu os critérios adotados na inspeção visual. Como se pode observar, o fragmento manteve as interações hidrofóbicas, previstas pela restrição e, além disso, predisse interações do tipo ligação de hidrogênio dentro do sub-sítio S2, com os resíduos de aminoácido Met68 e Asn69 (cadeia principal e lateral destes resíduos, respectivamente). Apesar de alguns destes fragmentos apresentarem baixa pontuação ( $\sim -3 \text{ kcal mol}^{-1}$ ), a boa complementaridade destes com a macromolécula na posição predita pela docagem levou à sua seleção para a etapa de dinâmica molecular. Nessa etapa, foi verificado o comportamento do fragmento no sítio, quando a proteína é tratada de forma flexível.

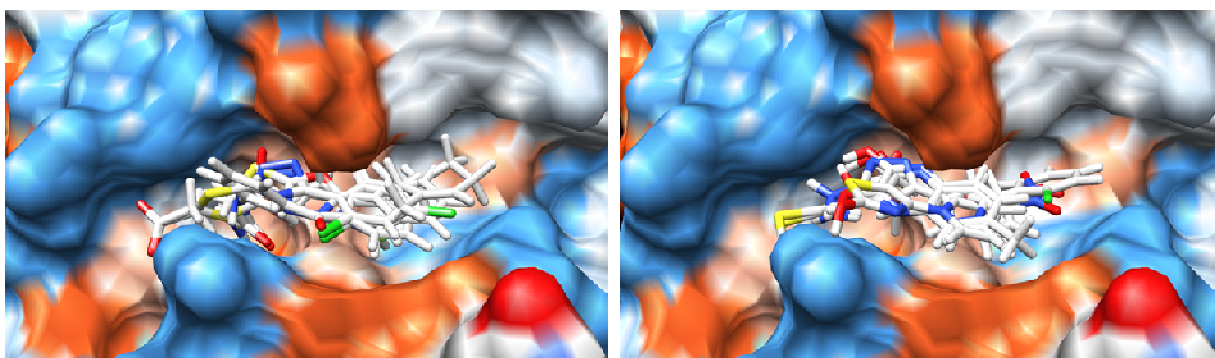
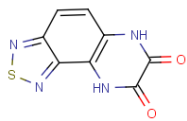
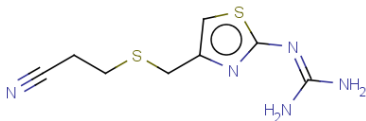
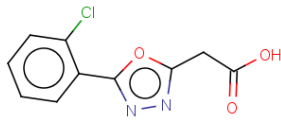
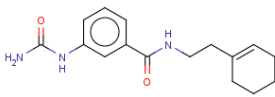
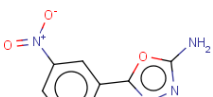
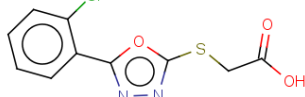
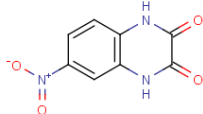
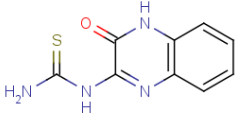
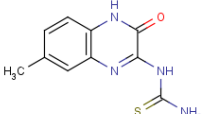
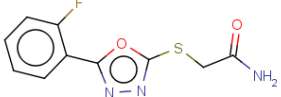
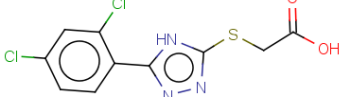
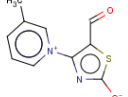


FIGURA 5.9 – Posições ocupadas pelos fragmentos selecionados obtidas pela docagem

TABELA 5.1 - Pontuação obtida na docagem molecular ( $\text{kcal mol}^{-1}$ )

1  -8,73	2  -6,45	3  -3,38
4  -6,09	5  -5,25	6  -3,95
7  -6,41	8  -4,41	9  -4,46
10  -6,07	11  -4,84	12  -7,36

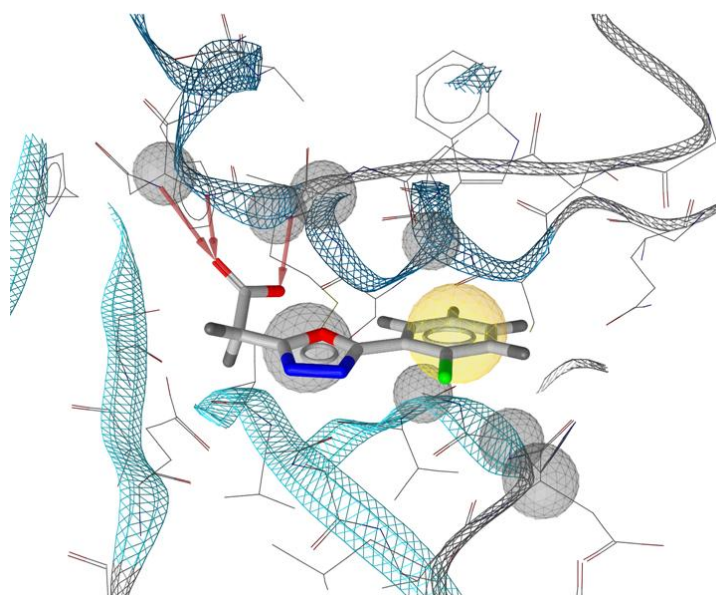


FIGURA 5.10 – Modo de interação observado para o fragmento 10

Os 12 fragmentos foram então submetidos a simulações de DM. Uma das etapas seguidas na DM é o *simulated annealing* (SA), um artifício usado para fornecer energia ao sistema para que o mesmo tenha chance de ultrapassar barreiras energéticas que o limitam a um mínimo local. Isso possibilita uma maior varredura no espaço conformacional do sistema, em busca de uma posição energeticamente mais favorável.

Em altas temperaturas, o sistema é capaz de ocupar regiões de alta energia do espaço conformacional e passar por grandes barreiras de energia. Com a

diminuição de temperatura, os estados de menor energia tornam-se mais prováveis de acordo com a distribuição de Boltzmann. No zero absoluto, o sistema deve ocupar o estado de menor energia (por exemplo, conformação de energia mínima global).<sup>103</sup>

Nos estudos de DM, os resultados sugerem uma posição para os fragmentos 3, 5 e 10 (Nequimed147, 182 e 183, respectivamente) no sítio de interação da enzima, consistente com os resultados obtidos na docagem molecular, já que esses três fragmentos permaneceram no sub-sítio S2, apresentando RMSD médio por volta de 1,5 Å (FIGURA 5.11). Os demais fragmentos não mantiveram o modo de interação durante a simulação, deslocando-se para outros sub-sítios dentro do sítio ativo da enzima. Apesar de a docagem prever uma boa afinidade dos compostos pelo alvo, quando se considera a flexibilidade da proteína essas interações não são suficientes para manter os compostos na mesma posição. Diante deste quadro, somente aqueles que mantiveram as posições foram selecionados para os testes de inibição.

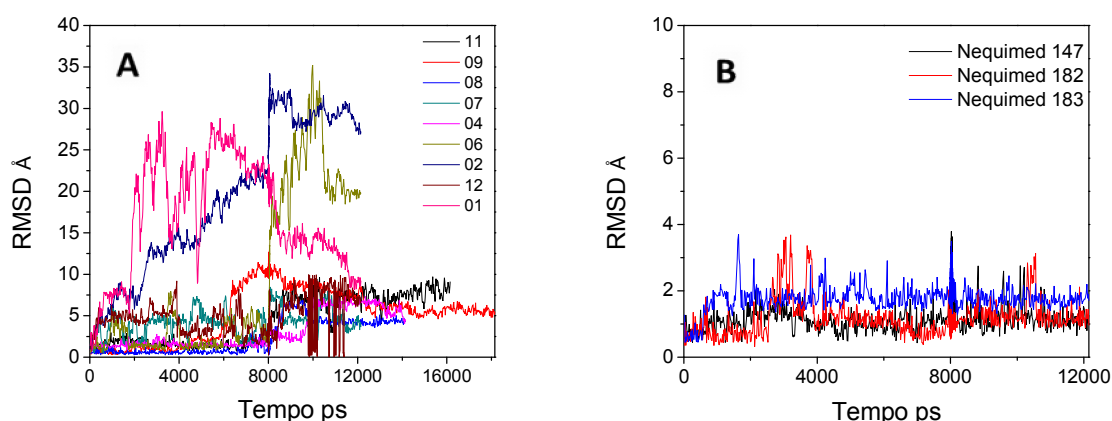


FIGURA 5.11 – RMSD dos fragmentos analisados

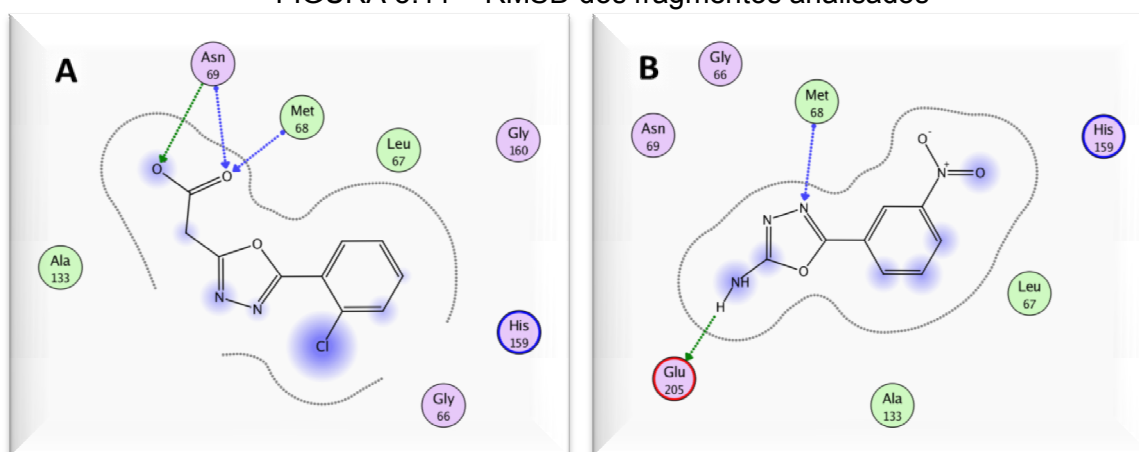


FIGURA 5.12 – Modo de interação predito na posição de menor energia durante a dinâmica molecular a) Nequimed147, b) Nequimed182

Na FIGURA 5.12 está representado o modo de interação de dois dos compostos que foram adquiridos para os testes de inibição. Todos eles mantiveram as mesmas características consideradas durante a inspeção visual da docagem, como por exemplo: ligações de hidrogênio com os resíduos Glu205, Asn69 e interação hidrofóbica com os resíduos Met68, Leu67.

Os fragmentos Nequimed147, 182 e 183 foram adquiridos e ensaiados contra a cruzaina. Para isso, um sistema robusto é requerido, pois, quando se fala em fragmentos implica em dizer que serão quantificadas interações fracas. Se essas interações podem ser detectadas, compostos com baixa massa molecular e baixa complexidade podem ser ensaiados. A disponibilidade de um ensaio de alta qualidade pode permitir a exploração de um composto de interações fracas que teoricamente não seriam um ponto de partida viável.

A solubilidade é uma questão que necessita ser previamente tratada com cuidado na seleção dos compostos para ensaios em alta concentração, para permitir que essas interações sejam detectadas. Quando exibe alta solubilidade e baixo ClogP, fragmentos tendem a ser menos propensos a agregarem. Porém não se pode esquecer que se ele for excessivamente hidrofílico, corre-se o risco de não interagir no sítio do alvo em estudo, então certo grau de caráter hidrofóbico é necessário para que interaja no alvo.

Dos três compostos testados, apenas o Nequimed147 inibiu a cruzaina apresentando um  $IC_{50}$  de  $2,39 \pm 0,21$  mM e a catepsina L com  $IC_{50}$   $2,74 \pm 0,04$  mM. Essa pequena diferença na interação não resultou em uma diferença significativa em termos de eficiência do ligante, que foi de  $0,23 \text{ kcal mol}^{-1} \text{ átomo pesado}^{-1}$  para ambas as enzimas. Assim, este composto foi utilizado como guia para a etapa seguinte deste trabalho, na tentativa de otimizar a eficiência do ligante, como descrito na próxima seção.

## 5.5 - Otimização do fragmento

Uma vez que o fragmento Nequimed147 foi identificado como inibidor da cruzaina, novos fragmentos foram selecionados para otimização da eficiência do ligante. Para isso, uma nova estratégia foi adotada, seguindo quatro etapas: 1) obtenção de um banco de moléculas disponíveis comercialmente no ZINC<sup>100</sup>, com

aproximadamente 18 milhões de compostos; 2) refino do banco realizando busca utilizando SMARTS baseado no esqueleto do Nequimed147; 3) Docagem e seleção dos compostos por inspeção visual; e 4) testes de inibição.

A etapa de refino do banco foi realizada no programa FILTER 2.0.2<sup>101</sup> utilizando o padrão SMARTS conforme descrição da FIGURA 5.13. A estrutura proposta deveria ser formada por um anel ligado a um heterociclo de 5 membros. Esse heterociclo de 5 membros deveria ter um substituinte que contemplasse grupos capazes de realizar ligação de hidrogênio.

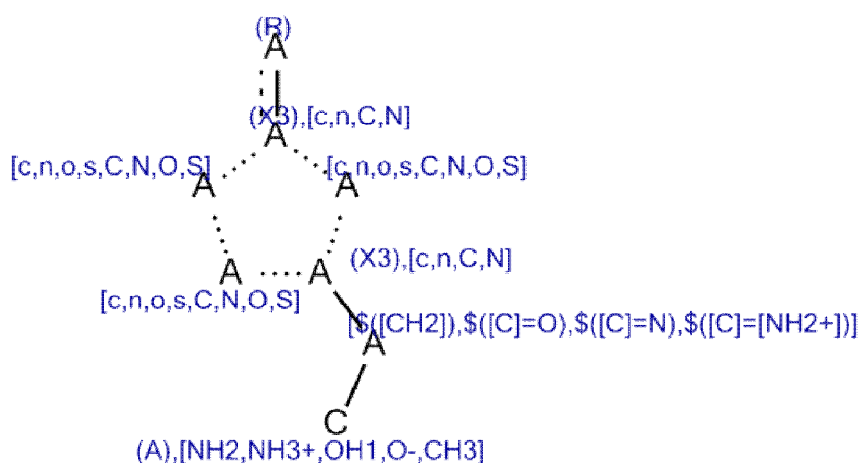


FIGURA 5.13 – Padrões SMARTS utilizados para a busca

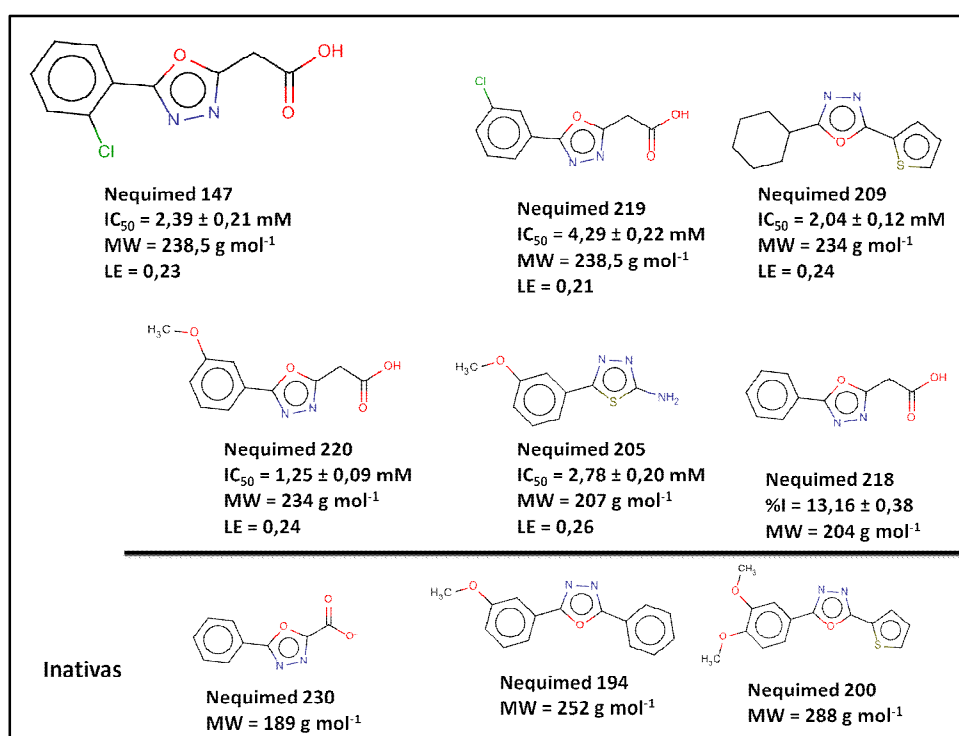
A docagem molecular seguiu o mesmo protocolo utilizado na etapa de seleção e identificação anterior, utilizando o módulo XP do programa Glide v5.5<sup>95</sup>. Os resíduos Glu205 e Ser207 foram utilizados como restrição, bem como uma região hidrofóbica próxima aos resíduos Leu67 e Ala131. Na TABELA 5.2 são apresentadas as pontuações obtidas para as moléculas selecionadas após a inspeção visual.

Os fragmentos foram então ensaiados utilizando o protocolo descrito na seção 5.3, sendo que oito destes apresentaram atividade inibitória frente à cruzaina. A curva dose-resposta contra essa enzima e o ensaio contra a catepsina L foram feitos somente para aqueles que inibiram a cruzaina em percentual maior ou igual a 50% a uma concentração de 3,76mM do fragmento (FIGURA 5.14 e 5.15). Alguns compostos não apresentaram solubilidade em água suficiente para a realização dos ensaios ou para realizar o experimento da curva dose-resposta.



TABELA 5.2 - Pontuação obtida por docagem molecular das moléculas selecionadas

Molécula	Pontuação (kcal mol <sup>-1</sup> )	Molécula	Pontuação(kcal mol <sup>-1</sup> )
Nequimed 194	-5,63	Nequimed 223	-5,45
Nequimed 200	-6,03	Nequimed 224	-5,13
Nequimed 205	-6,28	Nequimed 225	-4,77
Nequimed 209	-5,09	Nequimed 226	-6,17
Nequimed 218	-6,06	Nequimed227	-6,28
Nequimed 219	-6,07	Nequimed228	-5,55
Nequimed 220	-6,08	Nequimed229	-5,32
Nequimed 221	-7,08	Nequimed230	-5,32
Nequimed 222	-5,53	Nequimed231	-5,88

FIGURA 5.14 – Estruturas das moléculas análogas ao Nequimed 147 testadas contra a enzima cruzainá. LE = eficiência do ligante kcal mol<sup>-1</sup> átomo pesado<sup>-1</sup>



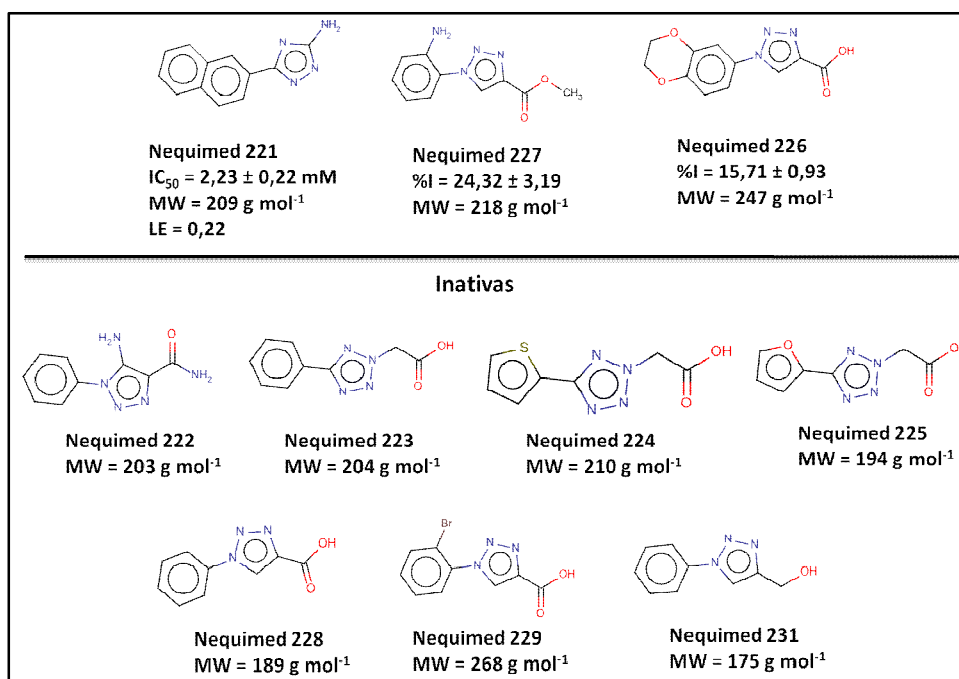


FIGURA 5.15 – Estruturas dos novos esqueletos moleculares testados contra a enzima cruzaina. LE = eficiência do ligante  $\text{kcal mol}^{-1} \text{ átomo pesado}^{-1}$

Dos cinco fragmentos ensaiados contra a catepsina L, três inibiram a enzima em uma concentração de 3,76 mM do inibidor, sendo os percentuais de inibição: Nequimed205 -  $34 \pm 0,86\%$ , Nequimed219 -  $15,49 \pm 1,52\%$  e Nequimed221 -  $53,61 \pm 1,89\%$ . O  $IC_{50}$  para o Nequimed221 contra a catepsina L foi  $3,34 \pm 0,22 \text{ mM}$  com LE de  $0,22 \text{ kcal mol}^{-1} \text{ átomo pesado}^{-1}$ .

Na seleção da série de compostos análogos ao Nequimed147 (FIGURA 5.14) foram variados os substituintes no anel fenila, a distância ácido carboxílico-oxadiazol e a natureza do substituinte nesse heterociclo (doador/aceitador de ligação de hidrogênio e hidrofóbico). A alteração na posição do halogênio ou a substituição dele por hidrogênio no anel levou a uma diminuição na potência (Nequimed218 e 219). Quando este átomo foi substituído pelo grupo metóxi, a potência aumentou significativamente, levando a um discreto aumento na eficiência do ligante (Nequimed220), de 0,23 para 0,24.

Com a manutenção na estrutura do grupo 2-(3-metoxifenil)-1,3,4-oxadiazol foi avaliado o ácido carboxílico. Quando este foi substituído por um grupo hidrofóbico como a fenila, o fragmento não foi capaz de inibir a enzima (Nequimed194). Por outro lado, quando o ácido foi substituído por um grupo doador de ligação de hidrogênio, a potência diminui assim como a massa molar do fragmento, porém a eficiência do ligante aumenta para 0,26 (Nequimed205).

Além dos análogos do Nequimed147, fragmentos com forma próxima a ele, mas grupos constituintes diferentes também foram selecionados, conforme a série de tetrazóis e triazóis apresentada na FIGURA 5.15. O composto Nequimed223 difere do Nequimed218 apenas pela substituição do oxadiazol pelo tetrazol; porém não apresentou percentual de inibição para a enzima cruzafina. Isso demonstra que o oxadiazol tem contribuição efetiva para a atividade, que é perdida quando da substituição pelo tetrazol. A presença dos grupos 2,3-diidro-1,4-benzodioxano e naftaleno no lugar da fenila parece ser um fator positivo para a interação, o que é suportado pelo fato de o sub-sítio S2 ser bastante hidrofóbico. Assim esses grupos apresentariam boa complementaridade pelo sítio.

# **CAPÍTULO 6**

## **Conclusões e Perspectivas**

## 6 - Conclusões e Perspectivas

### 6.1 - Conclusões

A busca por fragmentos capazes de inibir a enzima cruzaina foi realizada utilizando métodos virtuais baseados na estrutura do receptor, com o mapeamento do sítio de interação utilizando campos de interação molecular, docagem e dinâmica molecular. A inspeção do sítio possibilitou a determinação de resíduos importantes para a interação e seletividade, em relação à catepsina L, em uma região ainda não explorada no sítio da cruzaina. Esses resíduos são Glu205, Ser207 e Asn69.

Partindo de um banco de moléculas comerciais (ZINC) contendo cerca de 2 milhões de estruturas, o uso de filtros de seleção molecular, baseados em propriedades fragmentos-similares descritos na literatura, levou a obtenção de um banco focado de fragmentos, com 227 substâncias. Associando estudos de docagem, inspeção visual e dinâmica molecular, três fragmentos foram selecionados. Inicialmente esses fragmentos foram ensaiados para a determinação do IC<sub>50</sub> utilizando fluorimetria e foi identificado um fragmento (Nequimed147) com potência de 2,39 mM para a cruzaina e 2,74 mM para a catepsina L.

Baseado nesse fragmento, novos inibidores foram identificados com potência entre 1,25 e 4,29 mM e eficiência do ligante entre 0,21 e 0,26 kcal mol<sup>-1</sup> átomo pesado<sup>-1</sup>, com características apropriadas para o reconhecimento molecular pela enzima cruzaina em uma região do sub-sítio S2 ainda não explorada na literatura. Os resultados dos ensaios *in vitro* serviram de validação da estratégia adotada pelos métodos computacionais.

### 6.2 - Perspectivas

Neste trabalho, o uso de métodos computacionais possibilitou a seleção de fragmentos capazes de inibir a enzima cruzaina. Além disso, uma região ainda não explorada para planejamentos de inibidores desta enzima foi caracterizada por campos moleculares de interação. A validação ortogonal será feita pela realização de estudos de cristalização com a enzima cruzaina. De acordo com o modo de interação predito pela dinâmica molecular, os fragmentos identificados podem crescer na direção do sub-sítio S1, para região da cisteína catalítica.

A confirmação do modo de interação através da obtenção da estrutura tridimensional fornecerá informações importantes que possibilitarão o crescimento do fragmento visando a ocupação de outros sub-sítios no alvo. Além disso, substituintes poderão ser introduzidos para melhorar as interações com o alvo, resultando no aumento da potência.

Desse modo, novos esqueletos moleculares baseados nesses fragmentos poderão ser modelados pelos métodos computacionais e sintetizados para serem avaliados quanto à atividade tripanosomicida.

# **CAPÍTULO 7**

## **Referências Bibliográficas**

## 7 - Referência Bibliográfica

1. MASSAD, E. "The elimination of Chagas' disease from Brazil". *Epidemiology and Infection*, 2008. **136**(9): p.1153-1164
2. MAN, C.C.; FERREIRA, E.I.; et al. "Prodrugs for the treatment of neglected diseases". *Molecules*, 2008. **13**(3): p.616-677
3. CHIRAC, P. e TORREELE, E. "Global framework on essential health R&D". *Lancet*, 2006. **367**(9522): p.1560-1561
4. MORAN, M.; GUZMAN, J.; et al. "Neglected disease research and development: How much are we really spending?". *PLoS Medicine*, 2009. **6**(2): p.0137-0146
5. COURA, J.R. e CANÇADO, J.R. *A história da terapêutica da doença de Chagas*. [acesso em 24 jan. 2011]; Disponível em: <<http://www.fiocruz.br/chagas/cgi/cgilua.exe/sys/start.htm?sid=126>>
6. DIAS, L.C.; DESSOY, M.A.; et al. "Quimioterapia da doença de Chagas: Estado da arte e perspectivas no desenvolvimento de novos fármacos". *Quim Nova*, 2009. **32**(9): p.2444-2457
7. KEALEY, A. e SMITH, R. "Neglected tropical diseases: Infection, modeling, and control". *Journal of Health Care for the Poor and Underserved*, 2010. **21**(1): p.53-69
8. COURA, J.R. e VÍAS, P.A. "Chagas disease: A new worldwide challenge". *Nature*, 2010. **465**(7301 SUPPL.):
9. DNDI. [acesso em 24 jan. 2011]; Disponível em: <<http://www.dndi.org.br/pt/doencas-negligenciadas/doenca-de-chagas.html>>
10. DNDI. [acesso em 20 jan. 2011]; Disponível em: <<http://www.dndi.org.br/pt/doencas-negligenciadas/doenca-de-chagas/tratamentos-atuais.html>>
11. MONCAYO, A. e ORTIZ YANINE, M.I. "An update on Chagas disease (human American trypanosomiasis)". *Annals of Tropical Medicine and Parasitology*, 2006. **100**(8): p.663-677
12. MCKERROW, J.H.; DOYLE, P.S.; et al. "Two approaches to discovering and developing new drugs for Chagas disease". *Memorias do Instituto Oswaldo Cruz*, 2009. **104**(SUPPL. 1): p.263-269
13. ROMEIRO, N.C.; AGUIRRE, G.; et al. "Synthesis, trypanocidal activity and docking studies of novel quinoxaline-N-acylhydrazones, designed as cruzain inhibitors candidates". *Bioorg Med Chem*, 2009. **17**(2): p.641-652
14. MOTT, B.T.; FERREIRA, R.S.; et al. "Identification and optimization of inhibitors of trypanosomal cysteine proteases: Cruzain, rhodesain, and TbCatB". *J Med Chem*, 2010. **53**(1): p.52-60
15. DOS SANTOS FILHO, J.M.; LEITE, A.C.L.; et al. "Design, synthesis and cruzain docking of 3-(4-substituted-aryl)-1,2,4-oxadiazole-N-acylhydrazones as anti-Trypanosoma cruzi agents". *Bioorg Med Chem*, 2009. **17**(18): p.6682-6691
16. FREITAS, R.F.; OPREA, T.I. e MONTANARI, C.A. "2D QSAR and similarity studies on cruzain inhibitors aimed at improving selectivity over cathepsin L". *Bioorg Med Chem*, 2008. **16**(2): p.838-853
17. BRAK, K.; DOYLE, P.S.; et al. "Identification of a new class of nonpeptidic inhibitors of cruzain". *J Am Chem Soc*, 2008. **130**(20): p.6404-6410
18. CHEN, Y.T.; LIRA, R.; et al. "Synthesis of macrocyclic trypanosomal cysteine protease inhibitors". *Bioorg Med Chem Lett*, 2008. **18**(22): p.5860-5863
19. JAISHANKAR, P.; HANSELL, E.; et al. "Potency and selectivity of P2/P3-modified inhibitors of cysteine proteases from trypanosomes". *Bioorg Med Chem Lett*, 2008. **18**(2): p.624-628
20. BRAK, K.; KERR, I.D.; et al. "Nonpeptidic tetrafluorophenoxymethyl ketone cruzain inhibitors as promising new leads for chagas disease chemotherapy". *J Med Chem*, 2010. **53**(4): p.1763-1773
21. HERNANDES, M.Z.; RABELLO, M.M.; et al. "Studies toward the structural optimization of novel thiazolylhydrazone- based potent antitrypanosomal agents". *Bioorg Med Chem*, 2010. **18**(22): p.7826-7835
22. BUCKNER, F.S. e NAVABI, N. "Advances in Chagas disease drug development: 2009-2010". *Current Opinion in Infectious Diseases*, 2010. **23**(6): p.609-616

23. CHEN, Y.T.; BRINEN, L.S.; et al. "In vitro and in vivo studies of the trypanocidal properties of WRR-483 against *Trypanosoma cruzi*". *PLoS Neglected Tropical Diseases*, 2010. **4**(9):
24. MCGRATH, M.E.; EAKIN, A.E.; et al. "The crystal structure of Cruzain: A therapeutic target for Chagas' disease". *J Mol Biol*, 1995. **247**(2): p.251-259
25. LECAILLE, F.; KALETA, J. e BRÖMME, D. "Human and parasitic Papain-like cysteine proteases: Their role in physiology and pathology and recent developments in inhibitor design". *Chem Rev*, 2002. **102**(12): p.4459-4488
26. POWERS, J.C.; ASGIAN, J.L.; et al. "Irreversible inhibitors of serine, cysteine, and threonine proteases". *Chem Rev*, 2002. **102**(12): p.4639-4750
27. BRINEN, L.S.; HANSELL, E.; et al. "A target within the target: Probing cruzain's P1' site to define structural determinants for the Chagas' disease protease". *Structure*, 2000. **8**(8): p.831-840
28. ENGEL, J.C.; DOYLE, P.S.; et al. "Cysteine protease inhibitors cure an experimental *Trypanosoma cruzi* infection". *Journal of Experimental Medicine*, 1998. **188**(4): p.725-734
29. DU, X.; GUO, C.; et al. "Synthesis and structure-activity relationship study of potent trypanocidal thio semicarbazone inhibitors of the trypanosomal cysteine protease cruzain". *J Med Chem*, 2002. **45**(13): p.2695-2707
30. GREENBAUM, D.C.; MACKEY, Z.; et al. "Synthesis and structure-activity relationships of parasiticidal thiosemicarbazone cysteine protease inhibitors against *Plasmodium falciparum*, *Trypanosoma brucei*, and *Trypanosoma cruzi*". *J Med Chem*, 2004. **47**(12): p.3212-3219
31. FUJII, N.; MALLARI, J.P.; et al. "Discovery of potent thiosemicarbazone inhibitors of rhodesain and cruzain". *Bioorg Med Chem Lett*, 2005. **15**(1): p.121-123
32. FRICKER, S.P.; MOSI, R.M.; et al. "Metal compounds for the treatment of parasitic diseases". *J Inorg Biochem*, 2008. **102**(10): p.1839-1845
33. DURRANT, J.D.; KERÄNEN, H.; et al. "Computational identification of uncharacterized cruzain binding sites". *PLoS Neglected Tropical Diseases*, 2010. **4**(5):
34. KERR, I.D.; LEE, J.H.; et al. "Vinyl sulfones as antiparasitic agents and a structural basis for drug design". *J Biol Chem*, 2009. **284**(38): p.25697-25703
35. DARDONVILLE, C. "Recent advances in antitrypanosomal chemotherapy: Patent literature 2002-2004". *Expert Opinion on Therapeutic Patents*, 2005. **15**(9): p.1241-1257
36. CHOWDHURY, S.F.; SIVARAMAN, J.; et al. "Design of noncovalent inhibitors of human cathepsin L from the 96-residue proregion to optimized tripeptides". *J Med Chem*, 2002. **45**(24): p.5321-5329
37. WERMUTH, C.G.; GANELLIN, C.R.; et al. "Chapter 36. Glossary of Terms Used in Medicinal Chemistry (IUPAC Recommendations 1997)". *Annu Rep Med Chem*, 1998. **33**: p.385-395
38. LOMBARDINO, J.G. e LOWE III, J.A. "The role of the medicinal chemist in drug discovery - Then and now". *Nature Reviews Drug Discovery*, 2004. **3**(10): p.853-862
39. ZHAO, H. e GUO, Z. "Medicinal chemistry strategies in follow-on drug discovery". *Drug Discovery Today*, 2009. **14**(9-10): p.516-522
40. HOPKINS, A.L. "Network pharmacology: The next paradigm in drug discovery". *Nature Chemical Biology*, 2008. **4**(11): p.682-690
41. CAS. *Chemical Abstract Service - Database Counter*. [acesso em 03 fev. 2011]; Disponível em: <<http://www.cas.org/cgi-bin/cas/regreport.pl>>
42. LIPINSKI, C. e HOPKINS, A. "Navigating chemical space for biology and medicine". *Nature*, 2004. **432**(7019): p.855-861
43. JORGENSEN, W.L. "The Many Roles of Computation in Drug Discovery". *Science*, 2004. **303**(5665): p.1813-1818
44. ORITA, M.; OHNO, K. e NIIMI, T. "Two 'Golden Ratio' indices in fragment-based drug discovery". *Drug Discovery Today*, 2009. **14**(5-6): p.321-328
45. CHESSARI, G. e WOODHEAD, A.J. "From fragment to clinical candidate-a historical perspective". *Drug Discovery Today*, 2009. **14**(13-14): p.668-675
46. CONGREVE, M.; CHESSARI, G.; et al. "Recent developments in fragment-based drug discovery". *J Med Chem*, 2008. **51**(13): p.3661-3680



47. CONGREVE, M.; CARR, R.; et al. "A 'Rule of Three' for fragment-based lead discovery?". *Drug Discovery Today*, 2003. **8**(19): p.876-877
48. BLOMBERG, N.; COSGROVE, D.A.; et al. "Design of compound libraries for fragment screening". *J Comput Aided Mol Des*, 2009. **23**(8): p.513-525
49. PETROS, A.M.; DINGES, J.; et al. "Discovery of a potent inhibitor of the antiapoptotic protein Bcl-X<sub>L</sub> from NMR and parallel synthesis". *J Med Chem*, 2006. **49**(2): p.656-663
50. CIULLI, A.; WILLIAMS, G.; et al. "Probing hot spots at protein-ligand binding sites: A fragment-based approach using biophysical methods". *J Med Chem*, 2006. **49**(16): p.4992-5000
51. HOPKINS, A.L.; GROOM, C.R. e ALEX, A. "Ligand efficiency: A useful metric for lead selection". *Drug Discovery Today*, 2004. **9**(10): p.430-431
52. SCHULTES, S.; DE GRAAF, C.; et al. "Ligand efficiency as a guide in fragment hit selection and optimization". *Drug Discovery Today: Technologies*, 2010. **7**(3):
53. GODDETTE, D. "Fragment-based drug design optimizes hits". *Genetic Engineering News*, 2006. **26**(9): p.24-25
54. CHEN, Y. e POHLHAUS, D.T. "In silico docking and scoring of fragments". *Drug Discovery Today: Technologies*, 2010. **7**(3): p.e149-e156
55. AGNELLI, G.; HAAS, S.; et al. "A phase II study of the oral factor Xa inhibitor LY517717 for the prevention of venous thromboembolism after hip or knee replacement". *Journal of Thrombosis and Haemostasis*, 2007. **5**(4): p.746-753
56. GRID. v22c. Molecular Discovery Ltd. . 1998.
57. KASTENHOLZ, M.A.; PASTOR, M.; et al. "GRID/CPCA: A new computational tool to design selective ligands". *J Med Chem*, 2000. **43**(16): p.3033-3044
58. GOLPE. v4.6. Multivariate Infometric Analysis, S.r.l. 1999.
59. BERMAN, H.M.; WESTBROOK, J.; et al. *The Protein Data Bank*, in *Nucleic Acids Research*. 2000. p. 235-242.
60. . *SYBYL Version 8.0*. 1992: St. Louis Missouri, USA.
61. DAVIS, A.M.; TEAGUE, S.J. e KLEYWEGT, G.J. "Application and limitations of x-ray crystallographic data in structure-based ligand and drug design". *Angewandte Chemie - International Edition*, 2003. **42**(24): p.2718-2736
62. BRUNGER, A.T. "Free R value: A novel statistical quantity for assessing the accuracy of crystal structures". *Nature*, 1992. **355**(6359): p.472-475
63. BETHEL, P.A.; GERHARDT, S.; et al. "Design of selective Cathepsin inhibitors". *Bioorg Med Chem Lett*, 2009. **19**(16): p.4622-4625
64. MORGON, N.H. e COUTINHO, K. *Métodos de Química Teórica e Modelagem Molecular*. 2007, São Paulo: Editora Livraria da Física. 539p.
65. BROOIJMANS, N. e KUNTZ, I.D. 2003. p. 335-373.
66. ROSINI, M.; ANDRISANO, V.; et al. "Rational approach to discover multipotent anti-Alzheimer drugs". *J Med Chem*, 2005. **48**(2): p.360-363
67. SHOICHET, B.K. "Virtual screening of chemical libraries". *Nature*, 2004. **432**(7019): p.862-865
68. TAYLOR, R.D.; JEWSBURY, P.J. e ESSEX, J.W. "A review of protein-small molecule docking methods". *J Comput Aided Mol Des*, 2002. **16**(3): p.151-166
69. RESTER, U. "Dock around the clock - Current status of small molecule docking and scoring". *QSAR and Combinatorial Science*, 2006. **25**(7): p.605-615
70. VAN GUNSTEREN, W.F. e BERENDSEN, H.J.C. *Groningen Molecular Simulation (GROMOS)*. 1987.
71. WEINER, S.J.; KOLLMAN, P.A.; et al. "A new force field for molecular mechanical simulation of nucleic acids and proteins". *J Am Chem Soc*, 1984. **106**(3): p.765-784
72. BROOKS, B.R.; BRUCCOLERI, R.E.; et al. "CHARMM - A PROGRAM FOR MACROMOLECULAR ENERGY, MINIMIZATION, AND DYNAMICS CALCULATIONS". *J Comput Chem*, 1983. **4**(2): p.187-217
73. GOHLKE, H.; HENDLICH, M. e KLEBE, G. "Predicting binding modes, binding affinities and 'hot spots' for protein-ligand complexes using a knowledge-based scoring function". *Perspect Drug Discovery Des*, 2000. **20**: p.115-144

74. GOHLKE, H.; HENDLICH, M. e KLEBE, G. "Knowledge-based scoring function to predict protein-ligand interactions". *J Mol Biol*, 2000. **295**(2): p.337-356
75. DEWITTE, R.S. e SHAKHNOVICH, E.I. "SMoG: De novo design method based on simple, fast, and accurate free energy estimates. 1. Methodology and supporting evidence". *J Am Chem Soc*, 1996. **118**(47): p.11733-11744
76. EWING, T.J.A.; MAKINO, S.; et al. "DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases". *J Comput Aided Mol Des*, 2001. **15**(5): p.411-428
77. ZOETE, V.; GROSDIDIER, A. e MICHIELIN, O. "Docking, virtual high throughput screening and in silico fragment-based drug design". *Journal of Cellular and Molecular Medicine*, 2009. **13**(2): p.238-248
78. METROPOLIS, N.; ROSENBLUTH, A.W.; et al. "Equation of state calculations by fast computing machines". *The Journal of Chemical Physics*, 1953. **21**(6): p.1087-1092
79. MOITESSIER, N.; ENGLEBIENNE, P.; et al. "Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go". *British Journal of Pharmacology*, 2008. **153**(SUPPL. 1): p.S7-S26
80. KARPLUS, M. e PETSKO, G.A. "Molecular dynamics simulations in biology". *Nature*, 1990. **347**(6294): p.631-639
81. PAK, Y. e WANG, S. "Application of a Molecular Dynamics Simulation Method with a Generalized Effective Potential to the Flexible Molecular Docking Problems". *J Phys Chem B*, 2000. **104**(2): p.354-359
82. MANGONI, M.; ROCCATANO, D. e DI NOLA, A. "Docking of flexible ligands to flexible receptors in solution by molecular dynamics simulation". *Proteins: Structure, Function and Genetics*, 1999. **35**(2): p.153-162
83. ZIMMERMAN, M.; ASHE, B.; et al. "Sensitive assays for trypsin, elastase, and chymotrypsin using new fluorogenic substrates". *Anal Biochem*, 1977. **78**(1): p.47-51
84. EAKIN, A.E.; MILLS, A.A.; et al. "The sequence, organization, and expression of the major cysteine protease (cruzain) from *Trypanosoma cruzi*". *J Biol Chem*, 1992. **267**(11): p.7411-7420
85. LIMA, A.P.C.A.; DOS REIS, F.C.G.; et al. "Cysteine protease isoforms from *Trypanosoma cruzi*, cruzipain 2 and cruzain, present different substrate preference and susceptibility to inhibitors". *Molecular and Biochemical Parasitology*, 2001. **114**(1): p.41-52
86. ALVES, L.C.; JUDICE, W.A.S.; et al. "Substrate specificity of recombinant cysteine proteinase, CPB, of *Leishmania mexicana*". *Molecular and Biochemical Parasitology*, 2001. **116**(1): p.1-9
87. BARICOS, W.H.; ZHOU, Y.; et al. "Human kidney cathepsins B and L. Characterization and potential role in degradation of glomerular basement membrane". *Biochem J*, 1988. **252**(1): p.301-304
88. MASON, R.W. "Species variants of cathepsin L and their immunological identification". *Biochem J*, 1986. **240**(1): p.285-288
89. FAIRHEAD, M. e VAN DER WALLE, C.F. "The heavy-light chain loop of human cathepsin-L modulates its activity and stability". *Protein and Peptide Letters*, 2008. **15**(1): p.47-53
90. WIGGERS, H.J. *Planejamento de Inibidores da Enzima Cruzaina de Trypanosoma cruzi: dos Ligantes aos Compostos Matrizes*. São Carlos, Programa de Pós-Graduação em Química-UFSCar, 2011. Tese de doutorado, 173p.
91. SMARTS. *Daylight - Chemical Information Systems, Inc.* [acesso em; Disponível em: <[http://www.daylight.com/dayhtml\\_tutorials/languages/smarts/index.html](http://www.daylight.com/dayhtml_tutorials/languages/smarts/index.html)>
92. FRIESNER, R.A.; BANKS, J.L.; et al. "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy". *J Med Chem*, 2004. **47**(7): p.1739-1749
93. HALGREN, T.A.; MURPHY, R.B.; et al. "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening". *J Med Chem*, 2004. **47**(7): p.1750-1759
94. ELDRIDGE, M.D.; MURRAY, C.W.; et al. "Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes". *J Comput Aided Mol Des*, 1997. **11**(5): p.425-445

95. FRIESNER, R.A.; MURPHY, R.B.; et al. "Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes". *J Med Chem*, 2006. **49**(21): p.6177-6196
96. OMEGA. vv2.4.2. OpenEye Scientific Software.
97. SYBYL. v8.0. Tripos Associates. . 1992.
98. CASE, D.A.; CHEATHAM III, T.E.; et al. "The Amber biomolecular simulation programs". *Journal of Computational Chemistry*, 2005. **26**(16): p.1668-1688
99. FRISCH, M.J.T., G. W.; SCHLEGEL, H. B.; SCUSERIA, G. E.; ROBB, M. A.; CHEESEMAN, J. R.; MONTGOMERY, JR., J. A.; VREVEN, T.; KUDIN, K. N.; BURANT, J. C.; MILLAM, J. M.; IYENGAR, S. S.; TOMASI, J.; BARONE, V.; MENNUCCI, B.; COSSI, M.; SCALMANI, G.; REGA, N.; PETERSSON, G. A.; NAKATSUJI, H.; HADA, M.; EHARA, M.; TOYOTA, K.; FUKUDA, R.; HASEGAWA, J.; ISHIDA, M.; NAKAJIMA, T.; HONDA, Y.; KITAO, O.; NAKAI, H.; KLENE, M.; LI, X.; KNOX, J. E.; HRATCHIAN, H. P.; CROSS, J. B.; BAKKEN, V.; ADAMO, C.; JARAMILLO, J.; GOMPERTS, R.; STRATMANN, R. E.; YAZYEV, O.; AUSTIN, A. J.; CAMMI, R.; POMELLI, C.; OCHTERSKI, J. W.; AYALA, P. Y.; MOROKUMA, K.; VOTH, G. A.; SALVADOR, P.; DANNENBERG, J. J.; ZAKRZEWSKI, V. G.; DAPPRICH, S.; DANIELS, A. D.; STRAIN, M. C.; FARKAS, O.; MALICK, D. K.; RABUCK, A. D.; RAGHAVACHARI, K.; FORESMAN, J. B.; ORTIZ, J. V.; CUI, Q.; BABOUL, A. G.; CLIFFORD, S.; CIOŚLOWSKI, J.; STEFANOV, B. B.; LIU, G.; LIASHENKO, A.; PISKORZ, P.; KOMAROMI, I.; MARTIN, R. L.; FOX, D. J.; KEITH, T.; AL-LAHAM, M. A.; PENG, C. Y.; NANAYAKKARA, A.; CHALLACOMBE, M.; GILL, P. M. W.; JOHNSON, B.; CHEN, W.; WONG, M. W.; GONZALEZ, C.; AND POPLE, J. A. ed.^eds. Gaussian 03, ed. INC., G. 2004: Wallingford CT.
100. "ZINC - A free database dor virtual screening".
101. FILTER. OpenEye Scientific Software.
102. BRYANT, C.; KERR, I.D.; et al. "Novel non-peptidic vinylsulfones targeting the S2 and S3 subsites of parasite cysteine proteases". *Bioorg Med Chem Lett*, 2009. **19**(21): p.6218-6221
103. LEACH, A.R. *Molecular Modelling Principles and Applications*. 2 ed. 2001, Harlow, England: Prentice Hallp.

## Apêndice

## Banco de fragmentos final (227 estruturas)

1. CCc1c(C)nc(N)nc1N1[NH+]=C1SCC#N
2. CCc1c(C)nc(N)nc1N1[NH+]=C1SCC(N)=O
3. CCCCc1c(C)nc(N)nc1N1[NH+]=C1S
4. CCc1c(C)nc(N)nc1N1[NH+]=C1SCC(N)=O
5. Nc1nnc(o1)-c1cccc(c1)[N+]([O-])=O
6. Cn1c(NCCO)nc2n(C)c(=O)[nH]c(=O)c12
7. CC(C)Cn1c2c([nH]c1=NN)n(C)c(=O)[nH]c2=O
8. NC(=[NH2+])NN=C1C(=O)Nc2ccc(Br)cc12
9. [O-]C(=O)CNS(=O)(=O)c1ccc2NC(=O)Cc2c1
10. O=C1Nc2cccc2C1=C(C#N)C#N
11. NC(=[NH2+])NN=C1C(=O)Nc2cccc12
12. NC(=O)NN=C1C(=O)Nc2cccc12
13. NC(=O)C(=O)C1C(=O)Nc2cccc12
14. NC(=O)NN=C1C(=O)Nc2cccc12
15. Cc1sc2nc(CC(=O)NN)[nH]c(=O)c2c1C
16. CC1C(=O)Nc2ccc(cc12)-c1csc(NC(N)=N)n1
17. O=C1N(CCC#N)c2cccc2C1=O
18. Oc1cncc(c1)C(=O)NCCCC([O-])=O
19. COC(=O)c1cc(C(=O)NN)c(O)nc1C
20. OCCNC(=O)c1ccc(C=NO)nc1
21. O=C1[N+](CC#N)=NN=[N+]1c1cccc1
22. O=C1CC(=NN1CCC#N)c1cccc1
23. O=C(CC1=NC(=O)CS1)c1cccc1
24. Oc1cccc1C1=NN=C(C1)C([O-])=O
25. [O-]C(=O)C(=O)CC(=O)c1cccc(c1)[N+](O-)=O
26. OCN=CC1=C([O-])c2cccc2C1=O
27. OCCN=CC1Nc2cccc2C1=O
28. COc1ccc(cc1)N1C(=O)CSC1=S
29. [O-]C(=O)CN1C(=S)SC(=Cc2cccs2)C1=O
30. Cc1ccc(C=C2SC(=S)[N-]C2=O)cc1[N+](O-)=O
31. COc1ccc(C=C2SC(=S)[N-]C2=O)cc1
32. [O-]C(=O)c1ccc(cc1)N1C(=O)CSC1=S
33. Oc1cccc(C=C2SC(=S)N(CC([O-])=O)C2=O)c1
34. Cc1ccc(C=C2SC(=S)N(CC([O-])=O)C2=O)s1
35. O=C1[N-]C(=S)SC1=C1OC(=O)c2cccc12
36. Oc1ccc(C=C2SC(=S)[N-]C2=O)cc1
37. Oc1cccc(C=C2SC(=S)N(CC([O-])=O)C2=O)c1
38. O=C1CSC(=S)N1N1C(=O)c2cccc2C1=O
39. [O-]C(=O)c1cn2nc(NCCC[NH+]3CCOCC3)ccc2[nH+]1
40. O=C1Nc2cccc2C11NN=C(CC#N)O1
41. O=C1Nc2cccc2C11NN=C(CC#N)O1
42. Cc1ccc2N(CCCc2c1)C(=O)CC[n+]1cnc[n-]1
43. [O-]C(=O)C=CC(=O)N1CCCc2cccc12
44. Cc1ccc2N(CCCc2c1)C(=O)C[n+]1cnc[n-]1
45. O=C(Nc1ccc(cc1)C#N)C1CC(=O)Nc2cccc12
46. Cc1ccc2N(CC(=O)NCCC#N)CCCc2c1
47. [O-]C(=O)C1CC(=O)Nc2cc(F)ccc12
48. [O-]C(=O)c1ccc(COc2ccc3NC(=O)CCc3c2)o1
49. [O-]C(=O)c1nc2cccc2nc1C([O-])=O
50. C1c1nc2cccc2nc1C(C#N)C#N
51. Oc1nc2ccc(cc2nc1O)C([O-])=O
52. c1ccc2[n+]3N=NN=c3c3=NN=N[n+]3c2c1
53. Oc1nc2ccc(cc2nc1O)[N+](O-)=O
54. Oc1nc2ccc3nsnc3c2nc1O
55. N=[N+]=Nc1nc2cccc2[n+]2nn[n-]c12
56. N=[N]=Nc1nc2cccc2[n+]2nn[n-]c12

57. NNC(=O)Cc1nc2ccccc2[nH]c1=O  
 58. Cn1c2ccc3nsnc3c2n(C)c(=O)c1=O  
 59. NC(=O)Nc1nc2cc(Cl)cc(Cl)c2[nH]c1=O  
 60. Nc1c(C#N)n(O)c2cc(Cl)c(Cl)cc2[n+]<sub>1</sub>=O  
 61. NC(=O)Nc1nc2cc(Cl)cc(Cl)c2[nH]c1=O  
 62. Cc1ccc2nc(O)c(nc2c1)[NH+]=C(N)S  
 63. Oc1[nH]c2ccccc2n2-c1nnc2=O  
 64. [S-]c1nnc(o1)-c1nc2ccccc2[nH]c1=O  
 65. NC(=O)Nc1nc2cc(Br)ccc2[nH]c1=O  
 66. Nc1c(C#N)[n+](=O)c2ccccc2n1O  
 67. NC(S)=[NH+]c1nc2ccccc2nc1O  
 68. Cn1c2ccccc2nc(-c2nnc(N)o2)c1=O  
 69. Nc1nnc(o1)-c1nc2ccccc2[nH]c1=O  
 70. O=c1oc(nn1CCC#N)-c1nccc2ccccc12  
 71. CN(CC(N)=O)S(=O)(=O)c1cccc2cnccc12  
 72. NC(=O)CSc1nccc2ccccc12  
 73. COc1cccc2c(SCC(N)=O)nccc12  
 74. COc1cccc2c1ccn(CC(N)=O)c2=O  
 75. CCOCCn1ccc2c(OCC(N)=O)cccc2c1=O  
 76. COC(=O)COc1cccc2c1ccn(CC(N)=O)c2=O  
 77. COc1cc2ccnc(CCC[NH3+])c2cc1OC  
 78. NC(=O)CCc1cccc1F  
 79. [O-]C(=O)CSc1nnc([nH]1)-c1ccc(Cl)cc1Cl  
 80. Fc1cccc1-c1nnc(SCC#N)[nH]1  
 81. CC(Sc1nnc(o1)-c1cccc1F)C(N)=O  
 82. Fc1cccc1-c1nnc(SCC#N)o1  
 83. NC(=O)CSc1nnc(o1)-c1cccc1F  
 84. Fc1cccc1-c1nnc(SCC#N)[nH]1  
 85. NC(=O)CSc1n[nH]c(n1)-c1cccc1F  
 86. Clc1cccc1-c1nnc(C[NH+]2CCOCC2)o1  
 87. [O-]C(=O)CSc1nnc(o1)-c1cccc1Cl  
 88. [O-]C(=O)CSc1n[nH]c(n1)-c1ccc(Cl)cc1Cl  
 89. CC(=N)C(C#N)C(=O)C[n+]<sub>1</sub>nc(n[n-]<sub>1</sub>)-c1cccc1Cl  
 90. [O-]C(=O)CCn1nnc(n1)-c1cccc1F  
 91. [O-]C(=O)CSc1nnc(o1)-c1cccc1F  
 92. [O-]C(=O)CSc1nnc(o1)-c1ccc(Cl)cc1Cl  
 93. NC(=O)CSc1nnc(o1)-c1cccc1Cl  
 94. NNC(=O)Cn1nnc(n1)-c1cccc1Cl  
 95. [O-]C(=O)Cn1nnc(n1)-c1cccc1F  
 96. Nn1c([S-])nnc1-c1ccc(Cl)cc1Cl  
 97. Fc1cccc1-c1nnn(CCC#N)n1  
 98. NC(=O)Cn1nnc(n1)-c1cccc1F  
 99. Fc1cccc1-c1nnn(CCC#N)n1  
 100. [O-]C(=O)Cn1nnc(n1)-c1cccc1F  
 101. Fc1cccc1-c1nnn(CC#N)n1  
 102. CC(C)n1c(SCC(N)=O)nnc1-c1cccc1F  
 103. [O-]C(=O)Cn1nnc(n1)-c1cccc1Cl  
 104. NNC(=O)Cn1nnc(n1)-c1cccc1Cl  
 105. NC(=[NH2+])[NH+]=C1NC(=O)C(Cc2ccccc2Cl)S1  
 106. NC(=O)Nc1cccc(c1)C(=O)NCCc1cccs1  
 107. CC(C)CCNC(=O)c1cccc(NC(N)=O)c1  
 108. CN(C)C(=O)CNC(=O)c1cccc(NC(N)=O)c1  
 109. CCCNC(=O)C(C)NC(=O)c1cccc(NC(N)=O)c1  
 110. NC(=O)Nc1cccc(c1)C(=O)NCCC1=CCCCC1  
 111. CCOC(=O)CNC(=O)Nc1cccc(c1)C#N  
 112. CN(Cc1cccs1)C(=O)c1cccc(NC(N)=O)c1  
 113. CCC(C)NC(=O)Nc1cccc(c1)C#N  
 114. CCCCNC(=O)Nc1cccc(c1)C([O-])=O  
 115. Cc1nnc(COC(=O)c2cccc(NC(N)=O)c2)o1  
 116. NC(=O)Nc1cccc(c1)C(=O)NCCc1cccs1  
 117. CN(Cc1ccsc1)C(=O)c1cccc(NC(N)=O)c1  
 118. CC(C)CCNC(=O)c1cccc(NC(N)=O)c1  
 119. CC1CCN(CC1)C(=O)c1cccc(NC(N)=O)c1

120. CN(C)C(=O)CNC(=O)c1cccc(NC(N)=O)c1  
 121. NC(=O)Nc1cccc(c1)C(=O)NCCC1=CCCC1  
 122. CCOC(=O)CNC(=O)Nc1cccc(c1)C#N  
 123. CN(Cc1cccs1)C(=O)c1cccc(NC(N)=O)c1  
 124. CCC(C)NC(=O)Nc1cccc(c1)C#N  
 125. CCCNC(=O)Nc1cccc(c1)C([O-])=O  
 126. Cc1nnc(COC(=O)c2cccc(NC(N)=O)c2)o1  
 127. NC(=O)Nc1cccc(c1)C(=O)NCCc1cccs1  
 128. CN(Cc1ccsc1)C(=O)c1cccc(NC(N)=O)c1  
 129. CC(C)CCNC(=O)c1cccc(NC(N)=O)c1  
 130. CC1CCN(CC1)C(=O)c1cccc(NC(N)=O)c1  
 131. CN(C)C(=O)CNC(=O)c1cccc(NC(N)=O)c1  
 132. NC(=O)Nc1cccc(c1)C(=O)NCCc1cccs1  
 133. CC(C)CCNC(=O)c1cccc(NC(N)=O)c1  
 134. CN(C)C(=O)CNC(=O)c1cccc(NC(N)=O)c1  
 135. CCCNC(=O)C(C)NC(=O)c1cccc(NC(N)=O)c1  
 136. [O-]C(=O)CC(NC(=O)c1cccnc1)C([O-])=O  
 137. [O-]C(=O)CC(NC(=O)c1cccnc1)C([O-])=O  
 138. [O-]C(=O)CCNC(=O)c1cncc(Br)c1  
 139. OCC(NC(=O)c1cncc(Br)c1)C([O-])=O  
 140. OCC(NC(=O)c1cncc(Br)c1)C([O-])=O  
 141. Oc1cncc(c1)C(=O)NCCC([O-])=O  
 142. [O-][N+](=O)OCCNC(=O)c1cccnc1  
 143. Brc1cncc(c1)C(=O)OC1CCOC1=O  
 144. [O-][N+](=O)c1ccc(s1)C(=O)Nc1cccnc1  
 145. CC1=NC(=[NH2+])SC1=NNc1cccnc1  
 146. CN(CC([O-])=O)S(=O)(=O)c1cccnc1  
 147. O=C(Nc1sccc1C#N)c1cccnc1  
 148. [O-]C(=O)CCNS(=O)(=O)c1cccnc1  
 149. [O-]C(=O)CC(NC(=O)c1cccnc1)C([O-])=O  
 150. Cc1ccc[n+](c1)-c1nc(O)sc1C=O  
 151. Cn1nc(nc1Oc1cccnc1)[N+]([O-])=O  
 152. NC(=O)CSc1nnc(o1)-c1cccnc1  
 153. [O-]C(=O)CCNC(=O)c1cncc(Br)c1  
 154. NC(=O)c1nnn(c1N)-c1cccnc1  
 155. OCC(NC(=O)c1cncc(Br)c1)C([O-])=O  
 156. [O-][N+](=O)c1ccc(o1)C(=O)Nc1cccnc1  
 157. [O-]C(=O)C=CC(=O)NCc1cccnc1  
 158. Oc1cncc(c1)C(=O)NCCC([O-])=O  
 159. [O-]C(=O)CCc1nc(no1)-c1cccnc1  
 160. O=C(C[n+]1cccc1)Nc1nccs1  
 161. [O-]C(=O)CCSc1nnc(o1)-c1cccnc1  
 162. [O-]C(=O)Cc1csc(n1)-c1cccnc1  
 163. [O-]C(=O)Cn1nnc(n1)-c1cccnc1  
 164. [O-]C(=O)C=CC(=O)NCc1cccnc1  
 165. O=C(NN=Cc1cccnc1)c1cc[nH]n1  
 166. CC([O-])=C1C(NC(=O)C1=O)c1cccnc1  
 167. O=C1CSC(NN=Cc2cccnc2)=[NH+]1  
 168. O=C(NN=Cc1cccnc1)c1ccn[nH]1  
 169. COC(=O)c1cc(no1)-c1cccnc1  
 170. O[NH+]=C1NC(=O)C(S1)=Cc1cccnc1  
 171. O[NH+]=C1NC(=O)C(S1)=Cc1cccnc1  
 172. Cn1nc(nc1Oc1cccnc1)[N+]([O-])=O  
 173. N#CCSc1nnnn1-c1cccnc1  
 174. [O-][N+](=O)OCCNC(=O)c1cccnc1  
 175. NC(=O)CSc1nnc(o1)-c1cccnc1  
 176. NC(=O)c1nnn(c1N)-c1cccnc1  
 177. Nc1nonc1C(=O)NCc1cccnc1  
 178. [O-][N+](=O)OCCOC(=O)c1cccnc1  
 179. O=C(NN=Cc1cccnc1)c1cc[nH]n1  
 180. [O-]C(=O)Cn1nnc(n1)-c1cccnc1  
 181. OCc1nnn(c1CO)-c1cccnc1  
 182. [O-]C(=O)Cc1nc(n[nH]1)-c1cccnc1

183. [O-][N+](=O)c1ccc(o1)C(=O)Nc1cccnc1  
184. COC(=O)c1concl-c1cccnc1  
185. [O-]C(=O)Cc1n[nH]c(n1)-c1cccnc1  
186. CC(=O)Nc1nnc(o1)-c1cccnc1  
187. Cc1nc(sc1C([O-])=O)-c1cccnc1  
188. O=C(NCn1encn1)c1cccnc1  
189. NNC(=O)c1csc(n1)-c1cccnc1  
190. O[NH+]=CNc1noncl-c1cccnc1  
191. ONC(=O)c1csc(n1)-c1cccnc1  
192. [O-]C(=O)COCC(=O)Nc1cccnc1  
193. Nc1nonclC(=O)NCc1cccnc1  
194. [O-][N+](=O)c1ccsc1C[n+]1cccc1  
195. NC(=[NH2+])[NH+]=C1NC(=O)C(Cc2cccc2Cl)S1  
196. [O-]C(=O)CSc1nnc([nH]1)-c1ccc(Cl)cc1Cl  
197. Fc1cccc1-c1nnc(SCC#N)[nH]1  
198. CC(Sc1nnc(o1)-c1cccc1F)C(N)=O  
199. Fc1cccc1-c1nnc(SCC#N)o1  
200. NC(=O)CSc1nnc(o1)-c1cccc1F  
201. Fc1cccc1-c1nnc(SCC#N)[nH]1  
202. NC(=O)CSc1n[nH]c(n1)-c1cccc1F  
203. Clc1cccc1-c1nnc(C[NH+]2CCOCC2)o1  
204. [O-]C(=O)CSc1nnc(o1)-c1cccc1Cl  
205. [O-]C(=O)CSc1n[nH]c(n1)-c1ccc(Cl)cc1Cl  
206. CC(=N)C(C#N)C(=O)C[n+]1nc(n[n-]1)-c1cccc1Cl  
207. [O-]C(=O)CCn1nnc(n1)-c1cccc1F  
208. [O-]C(=O)CSc1nnc(o1)-c1cccc1F  
209. OC(=O)Cc1nnc(o1)-c1cccc1Cl  
210. NC(=O)CSc1nnc(o1)-c1cccc1Cl  
211. NNC(=O)Cn1nnc(n1)-c1cccc1Cl  
212. [O-]C(=O)Cn1nnc(n1)-c1cccc1F  
213. Nn1c([S-])nnc1-c1ccc(Cl)cc1Cl  
214. Fc1cccc1-c1nnn(CCC#N)n1  
215. NC(=O)Cn1nnc(n1)-c1cccc1F  
216. Fc1cccc1-c1nnn(CCC#N)n1  
217. [O-]C(=O)Cn1nnc(n1)-c1cccc1F  
218. Fc1cccc1-c1nnn(CC#N)n1  
219. NC(N)=Nc1nc(CSCCC#N)cs1  
220. [O-]C(=O)Cn1nnc(n1)-c1cccc1Cl  
221. NNC(=O)Cn1nnc(n1)-c1cccc1Cl  
222. Nc1nc(cs1)-c1cccnc1  
223. N#CCCNCc1cccnc1  
224. OCC[NH2+]Cc1cccnc1  
225. N#Cc1csc(n1)-c1cccnc1  
226. NC(=O)CCc1cccc1F  
227. Cc1cccc(c1)-c1nnc(N)s1