WILLIAM BORGES FERNANDES

**Determinação do Modo de Interação de Inibidores Reversíveis da
Cruzaína de *Trypanosoma cruzi* via Cristalografia de Raios X**

Tese apresentada ao Programa de Pós-Graduação do
Instituto de Química de São Carlos (IQSC), da
Universidade de São Paulo (USP), como um dos pré-
requisitos para concorrer ao título de doutor em ciências.

Área de concentração: Química Orgânica e Biológica
Orientador: Prof. Dr. Carlos Alberto Montanari

São Carlos
2015

AGRADECIMENTOS

A minha noiva Polyana França, meus pais Celismar e Wilma, e a minha irmã Liliam pelo suporte emocional e psicológico durante estes quatro anos de doutorado, especialmente durante o ano em que estive em San Francisco realizando o estágio sanduíche.

Ao professor Dr. Carlos Montanari pela oportunidade de trabalhar neste projeto desafiador em uma das melhores universidades da América Latina.

Aos amigos e colaboradores do grupo NEQUIMED: Malu, Andrei, Patrícia, Jean, Josmar, Helton, Juliana, Emanuella, Ricardo, as Fabianas, Igor, Geraldo, Leandro, Erika, Christian, Nádia, Karen e Daniel.

Ao Instituto de Química de São Carlos (IQSC) pelas dependências, instalações e equipamentos. Aos amigos que fiz no IQSC e na USP São Carlos.

Ao professor Dr. James McKerrow pela oportunidade de realizar a maior parte dos estudos cristalográficos deste projeto na Universidade da Califórnia em San Francisco (UCSF). Aos amigos do laboratório de patologia da UCSF, Judy, Connor Caffrey, Larissa Podust, Juan Engel, Jiri Gut, Alberto, Brian, Thavy, Debalina, Anjan, Potter, Sally, Lili, Chealsea, Cristine, Claudia Calvet, Jair Lage, Gregory Lee e Ana Maria.

Ao professor Dr. Brian Schoichet pelas fantásticas idéias e frutíferas discussões na parte computacional durante o período de estágio na UCSF e aos amigos que fiz no Schoichet Lab: John Irwin, Matthew Merski, Olive Eidam, Marcus Fischer, Hao Fan, Sara Bariler, Magdalena, Nir London.

Ao pesquisador Peter Kenny por sua motivação e inúmeras idéias que tanto nos inspirou durante sua estadia no NEQUIMED, e aos pesquisadores James Holton e George Meigs por compartilharem suas experiências durante as oito viagens para coletas de dados no sincrotron Advanced Light Source (ALS) que tive a oportunidade de realizar.

Aos organizadores das três principais escolas de cristalografia que tive a chance de participar durante o doutorado que contribuíram enormemente para meu aprendizado científico nesta área: Raúl Carbonio, Eduardo Horjales, Richard Garratt, Tom Blundell e em especial o professor Robert Stroud pelos ensinamentos e suporte laboratorial na UCSF.

A CAPES pelo suporte financeiro especialmente durante o período de estágio na UCSF com bolsa do programa PDSE (processo BEX 5985/11-0).

Ao CNPQ e FAPESP (projetos #2011/01893-3 e temático 13/18009-4) também pelo apoio financeiro na compra de materiais, reagentes e equipamento utilizados no laboratório.

A todos vocês, muito obrigado pelo apoio!

EPÍGRAFE

“Não há maior satisfação para uma pessoa justa e bem-intencionada do que saber que ela dedicou suas melhores energias a serviço de uma boa causa.”

Albert Einstein

RESUMO

A cruzaina, principal cisteína protease do *Trypanosoma cruzi*, é um alvo terapêutico validado para o tratamento da doença de Chagas. Grande parte dos inibidores desta enzima é constituída de peptídeo-miméticos do tipo covalente irreversível cujo desenvolvimento, entretanto, tem sido evitado devido ao potencial efeito *off target* e ao perfil farmacocinético indesejável. O grupo de química medicinal NEQUIMED/IQSC/USP vêm utilizando métodos avançados em quimioinformática para a identificação de inibidores reversíveis da cruzaina e de outras cisteína proteases. Contudo, para que estes inibidores se desenvolvam a compostos matrizes mais eficientes, informações computacionais e estruturais de seus Modos de Interação (MOB) com a proteína alvo tornam-se fundamentais. Neste trabalho, a cristalografia de raios X foi utilizada para descrever em detalhes o MOB de três importantes inibidores reversíveis da cruzaina identificados pelo NEQUIMED: a dipeptidil-nitrila Neq0409 e os dois fragmentos moleculares Neq0147 e Neq0176. De modo a evitar a auto-proteólise experimentada pela cruzaina nativa que torna desafiadora a cristalização com inibidores de baixa afinidade como os fragmentos reversíveis, duas novas construções mutantes e inativas da cruzaina (a C25A e a C25S) foram desenvolvidas. A C25S foi validada como modelo cristalográfico representativo dado a coerência do MOB elucidado nesta enzima mutante e na cruzaina nativa para o inibidor mais estudado da cruzaina, o K777. A descrição da presença, orientação, conformação e modo de ação no sítio, além do completo padrão de interações fornecidos por estas estruturas cristalinas, validaram ortogonalmente os MOB preditos e os métodos de planejamento *in silico* usados no NEQUIMED, permitindo a identificação de inibidores muito mais potentes análogos ao Neq0147 e ao Neq0409. O grupo 2-acetamidotiofeno-3-carboxamida do Neq0176 e o anel heterocíclico de cinco membros baseado no 1,3,4-oxadiazol do Neq0147 foram identificados como alternativos aos tradicionais esqueletos peptídeo-miméticos. O impacto do efeito proteolítico na qualidade e na resolução de estruturas cristalográficas na cruzaina foi melhor compreendido a partir de duas estruturas de alta resolução obtidas para a cruzaina nativa em complexo com o metanotiosulfonato de metila (MMTS) e a iodoacetamida. Estes resultados permitiram compreender as bases experimentais para a cristalização de inibidores reversíveis de baixa afinidade. Todos os resultados estruturais obtidos via cristalografia de raios X neste trabalho são úteis para mapear as bases estruturais essenciais para o planejamento de futuros inibidores mais potentes e seletivos da cruzaina.

ABSTRACT

Cruzain, the major *Trypanosoma cruzi* cysteine protease, is a validated therapeutic target for the search of new medicines for the treatment of Chagas disease. A myriad of inhibitors of this enzyme consists of covalent irreversible peptidomimetics whose development has been impaired due to potential *off target* effect and undesirable pharmacokinetic profile. Modern cheminformatic methods employed by The Medicinal Chemistry Group (NEQUIMED/IQSC/USP) were used to identify cruzain reversible inhibitors. Their optimization for more efficient compounds can be accomplished by the use of data and information gathered from computational and structural modes of interaction (MOB). In this doctoral thesis, the X-ray crystallography was used to describe in detail the MOB of three important new cruzain reversible inhibitors: the dipeptidil-nitrile Neq0409 and the two molecular fragmentos Neq0147 and Neq0176. In order to avoid cruzain self-proteolysis during crystallization with low affinity reversible inhibitors such as the identified fragments, two new mutant and inactive constructs of cruzain (the C25S and C25A) were designed to uphold the same properties of the wild type catalytic site. The C25S was validated as representative crystallographic model given the coherence of the MOB elucidated for the best-known and studied cruzain inhibitor, the K777. The description of the presence, orientation, conformation and mode of action at the site, besides the complete pattern of bimolecular interactions, provided by these crystalline structures, orthogonally validated the predicted *in silico* MOB and allowed the identification of other potent inhibitors analogous to the Neq0147 and the Neq0409. The 2-acetamidothiophene-3-carboxamide moiety (Neq0176) and heterocyclic five-membered ring based on the 1,3,4-oxadiazole (Neq0147) were thereby identified as attractive alternatives to traditional peptidomimetics. The impact of proteolytic effect on the quality and resolution of crystallographic structures in cruzain was best understood from two high-resolution structures obtained for the native cruzain in complex with methyl methanethiolsulfonate (MMTS) and iodoacetamide. These results allow the understanding of experimental basis for the crystallization with reversible inhibitors of low affinity such as fragments. All structural results obtained by X-ray crystallography together with the catalytic site depiction using GRID and SuperStar methods are useful for mapping the essential structural basis for the design of future more potent and selective cruzain inhibitors.

LISTA DE FIGURAS

Figura 1.1 - Ciclo de vida do parasito <i>Trypanosoma cruzi</i>	17
Figura 1.2 - Radiografia do tórax mostrando o aumento progressivo do coração e de órgãos do sistema digestivo de um paciente chagásico.....	18
Figura 1.3 - Estruturas moleculares dos fármacos atualmente disponíveis para o tratamento da doença de Chagas.....	19
Figura 1.4 - Inibidores da via metabólica do ergosterol com atividade antichagásica....	22
Figura 1.5 - Sítio catalítico da cruzaina em superfície de <i>Van der Waals</i> colorida de acordo com os subsítios de ligação.....	25
Figura 1.6 - Habilidade do aminoácido Glu208 na base do sítio S2 da cruzaina em modificar sua conformação para interagir com os substratos Phe, Arg e Tyr.....	26
Figura 1.7 - Warheads frequentemente encontrados em inibidores da cruzaina.....	27
Figura 1.8 - Mecanismo de morte do <i>T. cruzi</i> via inibidores da cruzaina.....	28
Figura 1.9 - Estrutura química do K777 em A, MOB em B e padrão de interações com os principais aminoácidos do sítio catalítico em C.....	29
Figura 1.10 - Estágios do processo de P&D de fármacos.....	31
Figura 1.11 - Esquema representativo do espaço químico-biológico.....	34
Figura 1.12 - Vantagens combinatórias da identificação de compostos líderes por métodos baseados em fragmentos em relação aos métodos tradicionais.....	40
Figura 1.13 - Sensibilidade das técnicas biofísicas, bioquímicas e calorimétricas de identificação de <i>hits</i> em relação à afinidade relativa e suas massas.....	41
Figura 1.14 - Cascata de etapas para o desenvolvimento de fármacos por FBDD...	43
Figura 1.15 - Esquema representando o mecanismo de <i>binding</i> de primeira ordem em termos cinéticos e termodinâmicos.....	46
Figura 1.16 - Inibidores reversíveis da cruzaina ensaiados cristalograficamente na UCSF com seus respectivos parâmetros de caracterização.....	48
Figura 1.17 - Inibidores da cruzaina identificados por FBDD que foram ensaiados.	49
Figura 3.1 - Representação do sistema de expressão de proteína heteróloga em plasmídeo bacteriano clássico pET-21a induzido por IPTG.....	62

Figura 3.2 - Representação esquemática do mecanismo de autoindução mínima...	63
Figura 3.3 - Diagrama de fases de cristalização de uma proteína.....	66
Figura 3.4 - Experimento de difusão de vapor com gota suspensa.....	67
Figura 3.5 - Modelo do dano causado por resfriamento rápido sem crioproteção....	69
Figura 3.6 - Diagrama da construção do plasmídeo para expressão da cruzaina em <i>E. coli</i>	70
Figura 3.7 - Sequência proteica codificada pelo gene da procruzaina.....	72
Figura 3.8 - Ativação da cruzaina nativa a partir da autoproteólise do prodomínio N-terminal.....	75
Figura 3.9 - Efeito proteolítico da cruzaina em relação ao tipo de inibidor.....	78
Figura 3.10 - Atividade proteolítica da cruzaina em relação pH de operação.....	79
Figura 3.11 - Estrutura cristalográfica da procatepsina L e sobreposição com a estrutura da C25A após clivagem por trombina.....	82
Figura 3.12 - Cruzaina C25S mutante após purificação IMAC.....	84
Figura 3.13 - Trans-ativação da cruzaina C25S utilizando 10% e 20% de cruzaina nativa e teste de estabilidade monitorado em gel SDS.....	85
Figura 3.14 - Cristais dos complexos com a C25S obtidos nos ensaios iniciais.....	86
Figura 3.15 - Cristais da C25S co-cristalizada com Neq0147 em A e padrão de difração correspondente em C.....	87
Figura 4.1 - Representação esquemática da rede cristalina e dos seis parâmetros de cela unitária correspondentes: a , b , c , α , β e γ	91
Figura 4.2 - Combinação dos elementos de simetria pontuais e translacionais para a formação dos 230 grupos espaciais cristalográficos.....	94
Figura 4.3 - Representação esquemática fora de escala da Lei de Bragg e dependência do vetor espalhamento h em relação à s e s₀	95
Figura 4.4 - Planos cristalográficos em A e representação geométrica das condições de Laue em B.....	97
Figura 4.5 - A definição da esfera de Ewald e as condições que satisfazem as equações de Bragg e Laue da difração.....	98
Figura 4.6 - Representação do padrão de difração para um cristal estacionário em A e a formação de meia-luas concêntricas no padrão de difração mediante a rotação	

do cristal por um ângulo $\Delta\phi$ em B.....	102
Figura 4.7 - Meia-luas concêntricas resultantes da família de planos paralelos do espaço recíproco.....	103
Figura 4.8 - Representação esquemática de uma reflexão F_{hkl} qualquer inteiramente gravada em A , e parcialmente gravada em diferentes imagens de difração em B e C de acordo com o ângulo $\Delta\phi$	105
Figura 4.9 - Mapas de densidade eletrônica (F_{obs} , ϕ_{calc}) calculados com diferentes limites de resolução para o fragmento LVP da lisozima.....	108
Figura 4.10 - MOB do Neq0409 em complexo com a cruzaina nativa covalentemente ligado à cisteína catalítica da cadeia A com mapa $2F_{obs}-F_{calc}$ contornados à 1σ	117
Figura 4.11 - Sobreposição do MOB do Neq0409 com o análogo estrutural ligado a catepsina L (código PDB 3HHA) em superfície de Van der Waals da cruzaina.....	118
Figura 4.12 - Exibição estereoscópica da tentativa de modelar o inibidor Neq0409 na cadeia B do complexo com a cruzaina nativa.....	119
Figura 4.13 - Modos de interação do fragmento Neq0176 em complexo com a cruzaina nativa e suas variantes conformacionais.....	123
Figura 4.14 - Análise detalhada do MOB para as três variantes conformacionais do Neq0176 na cruzaina nativa para as cadeias A, B e C.....	124
Figura 4.15 - Remodelagem da estrutura cristalográfica que descreve o MOB do Neq0176 considerando a presença do MMTS parcialmente ligado à Cys25 para as cadeias B e C do modelo.....	131
Figura 4.16 - Representação em superfície de Van der Waals das cinco cadeias da cruzaina relacionadas por NCS existentes na ASU com destaque para a localização dos inibidores nas cadeias B e C.....	126
Figura 4.17 - Exibição estereoscópica do MOB do fragmento Neq0147 em complexo com a cruzaina C25S mutante obtido na cadeia A.....	129
Figura 4.18 - Variantes conformacionais no MOB do Neq0147 obtidos nas cadeias B, C e D do complexo com a cruzaina C25S mutante.....	130
Figura 4.19 - O MOB do Neq0147 nas cadeias E e F do complexo com a C25S foram influenciadas pelo empacotamento cristalino.....	131
Figura 4.20 - Sobreposição de todas as variantes conformacionais obtidas para as cadeias A, B, C, D, E e a simétrica F em superfície de Van der Waals.....	131

Figura 4.21 - Sobreposição do MOB obtido para o Neq0147 na cadeia B da C25S (magneta) com a pose de docagem (em verde).....	132
Figura 4.22 - MOB do inibidor K777 na cruzaína C25S com mapas de densidade eletrônica, e sobreposição deste MOB com o obtido para a cruzaína nativa.....	134
Figura 5.1 - Mecanismo de inibição irreversível da cruzaína por Iodoacetamida....	140
Figura 5.2 - Mecanismo de inibição covalente reversível da cruzaína por MMTS...140	
Figura 5.3 - Estrutura da cruzaína nativa mostrando o MOB do inibidor reversível B95 com 70% de ocupação, com destaque para a molécula de MMTS ligada covalentemente com a Cys25B com 30% de ocupação.....	141
Figura 5.4 - Estrutura cristalográfica da cruzaína nativa em complexo com o MMTS obtido a 1,95 Å de resolução.....	144
Figura 5.5 - Estrutura do complexo cruzaína nativa ligada covalentemente a iodoacetamida em A, e sobreposição das estruturas para os complexos com MMTS, iodoacetamida e 1ME4 em B.....	146
Figura 5.6 - Sobreposição do MOB dos inibidores Neq0147, Neq0176 e Neq0409 em superfície de Van der Waals da cruzaína.....	148
Figura 5.7 - Inibidores da cruzaína do tipo fragmentos planejados com base no MOB do Neq0147 e Neq0409 que foram utilizados nos ensaios cristalográficos.....	149
Figura 5.8 - Inibidores da cruzaína do tipo dipeptidil-nitrilas planejadas a partir do composto protótipo Neq0409 que foram cristalograficamente ensaiadas.....	150
Figura 5.9 - Cristais obtidos nos ensaios de cristalização da cruzaína nativa com a segunda geração de inibidores.....	154
Figura 5.10 - Exibição estereoscópica do sítio catalítico da cadeia A da cruzaína em complexo com iodoacetamida cujo cristal foi submetido à <i>soaking</i>	156
Figura 5.11 - Conteúdo da ASU na estrutura da cruzaína nativa em complexo com a iodoacetamida e empacotamento cristalino visto na direção dos eixos x, y e z.....	157
Figura 5.12 - Sondagem do sítio catalítico da cruzaína com o programa GRID utilizando as sondas H ₂ O ou hidrofílica em A , dry ou hidrofóbica em B , C3 em C , C1 em D , N1 em E e O _{sp} em F	159
Figura 5.13 - Sondagem do sítio catalítico da cruzaína com o programa GRID utilizando as sondas H ₂ O ou hidrofílica em A , dry ou hidrofóbica em B , C3 em C , C1 em D , N1 em E e O _{sp} em F	160

Figura 5.14 - Sobreposição do MOB do Neq0147 (cadeia C) com a superfície de nível de energia à -5 kcal/mol (em verde) obtida para a sonda cloro.....	160
Figura 5.15 - Sobreposição das estruturas da cruzaina (complexo com iodoacetamida) e catepsina L (PDB 3HHA) com destaque para os principais aminoácidos do sítio.....	162
Figura 5.16 - Sondagem do sítio catalítico da cruzaina e da catepsina L com as sondas OH e C=O no SuperStar.....	163
Figura 5.17 - Sondagem do sítio catalítico da cruzaina e da catepsina L com a sonda NH amino do SuperStar.....	164
Figura 5.18 - Sondagem do sítio catalítico da cruzaina e da catepsina L com sondas CH alifático e aromático no SuperStar.....	164
Figura 5.19 - Análise do sítio da cruzaina nativa em complexo com o Neq0409 (PDB 4QH6) e catepsina L (PDB 3HHA) no DS Vizualiser.....	165

LISTA DE TABELAS

Tabela 3.1 - Estruturas cristalográficas da cruzaína disponíveis no PDB no início deste projeto mostrando o precipitante e o pH de cristalização, fonte de coleta de dados, mecanismo de inibição, resolução, ASU e grupo espacial.....	76
Tabela 4.1 - Os sete sistemas cristalinos e suas respectivas simetrias pontuais essenciais, restrições dos parâmetros de rede e possíveis centragens.....	92
Tabela 4.2 - Estatísticas relacionadas à coleta de dados e ao refinamento do modelo para a cruzaína nativa em complexo com a dipeptidil-nitrila Neq0409.....	115
Tabela 4.3 - Estatísticas relacionadas à coleta de dados e ao refinamento do modelo para a cruzaína nativa em complexo com o inibidor reversível Neq0176.....	121
Tabela 4.4 - Estatísticas relacionadas à coleta de dados e ao refinamento do modelo para a cruzaína C25S em complexo com o fragmento Neq0147.....	128
Tabela 4.5 - Estatísticas da coleta de dados e de refinamento do modelo para a cruzaína C25S mutante em complexo com o K777.....	133
Tabela 5.1 - Estatísticas da coleta de dados e de refinamento do modelo para a cruzaína nativa em complexo com MMTS.....	143
Tabela 5.2 - Estatísticas da coleta de dados e de refinamento do modelo para a cruzaína nativa em complexo com Iodoacetamida.....	146
Tabela 5.3 - Condições de cristalização identificadas nas triagens iniciais com <i>kits</i> da Hampton para os complexos com os inibidores de segunda geração.....	153
Tabela 5.4 - Estatísticas da coleta e refinamento do modelo para a cruzaína nativa em complexo com iodoacetamida em grupo espacial de maior simetria.....	156

LISTA DE ABREVIATURAS E SIGLAS

Å	Angstroms
ADME	Absorção, Distribuição, Metabolismo e Excreção
Ala	Alanina
ALS	<i>Advanced Light Source</i>
ANVISA	Agência Nacional de Vigilância Sanitária
Arg	Arginina
ASU	Unidade assimétrica, do inglês <i>Asymmetric Unit</i>
BMCD	<i>Biological Macromolecules Crystallization Database</i>
CAS	<i>Chemical Abstracts Service</i>
CDIPD	<i>Center for Discovery and Innovation in Parasitic Diseases</i>
C.I.	Covalente Irreversível
ClogP	Coeficiente de partição octanol / água
CoMFA	<i>Comparative Molecular Field Analysis</i>
CoMSIA	<i>Comparative Molecular Similarity Indices Analysis</i>
C_P	<i>Capacidade calorífica</i>
CSD	<i>Cambridge Structural Database</i>
Cys	Cisteína
ΔG	Energia livre de Gibbs
Da	Dalton, unidade de massa molecular
DFS	<i>Differential Scanning Fluorimetry</i>
d_{hkl}	Distância interplanar
DHODH	Dihidroorotato Desidrogenase
DLS	<i>Differential Static Light Scattering</i>
DRX	Difração de Raios X
DSC	Calorimetria diferencial de varredura ou <i>Differential Scanning Calorimetry</i>
E.C.	<i>Enzyme Commission</i>
<i>E. coli</i>	<i>Escherichia coli</i>
e.g.	Exemplo Geral
ELBOW	<i>Electronic Ligand Building and Optimization Workbench</i>
EMA	<i>European Medicines Agency</i>
Fator B	Fator de deslocamento atômico, do inglês <i>B-factor</i>

FBDD	<i>Fragment Based Drug Design</i>
FDA	<i>Food and Drug Administration</i>
F_{hkl}	Fator estrutura
FPLC	<i>Fast Protein Liquid Chromatography</i>
G3P	Gliceraldeído-3-fosfato
GAPDH	Gliceraldeído-3-fosfato desidrogenase
GeV	Giga eletron-volts
Gln	Glutamina
Glu	Glutamato
Gly	Glicina
GST	Glutathione S Transferase
H	Entalpia
HAC	Número de átomos diferentes do hidrogênio
HCS	Ensaio enzimáticos
His	Histidina
<i>Hits</i>	Ligantes
HTC	<i>High Throughput Crystallography</i>
HTS	Ensaio de triagem em massa, do inglês <i>High Throughput Screening</i>
IC_{50}	Concentração inibitória 50%
IMAC	Cromatografia de Afinidade por Metal Imobilizado
IND	<i>Investigational New Drug application</i>
IPTG	isopropil β -D-1-tiogalactopiranósido
IQSC	Instituto de Química de São Carlos
ITC	Calorimetria de titulação isotermal ou <i>Isothermal Titration Calorimetry</i>
IUPAC	<i>International Union of Pure and Applied Chemistry</i>
JCSG	<i>Joint Center for Structural Genomics</i>
K_d	Constante de dissociação
K_i	Constante inibição enzimática
k_{on}	Taxa de associação
k_{off}	Taxa de dissociação
k_{obs}	Constante aparente de primeira ordem
LB	<i>Luria Broth</i>

LBDD	<i>Ligand Based Drug Design</i>
LE	Eficiência do ligante, do inglês <i>Ligand Efficiency</i>
Leu	Leucina
$LL E_{Astex}$	Eficiência lipofílica, do inglês <i>Ligand Lipophilicity Efficiency Astex</i>
LLG	<i>Log-Likelihood Gain</i>
logP	Coeficiente de partição octanol / água.
MAD	<i>Multiple wavelength Anomalous Dispersion</i>
MD	Simulação por dinâmica molecular, do inglês <i>Molecular Dynamics</i>
Met	Metionina
MIFs	Campos de interações moleculares ou <i>Molecular Interaction Fields</i>
MIR	<i>Multiple Isomorfic Replacement</i> ou substituição isomórfica múltipla
ML	Máxima verossimilhança, do inglês <i>Maximum Likelihood</i>
MOB	Modo de interação, do inglês <i>Mode of Binding</i>
MMTS	Metil Metano Tio-Sulfonato
MW	Massa molecular, do inglês <i>Molecular Weight</i>
NAD ⁺	Nicotinamida Adenina Dinucleotídeo
NCEs	Novas entidades químicas, do inglês <i>New Chemical Entities</i>
N.C.R.	Não Covalente Reversível
NCS	Simetria não cristalográfica, do inglês <i>Non Crystallographic Symmetry</i>
NDA	Registro de novo fármaco, do inglês <i>New Drug Application</i>
NEQUIMED	Grupo de Química Medicinal do IQSC / USP
OMS	Organização Mundial da Saúde
P&D	Pesquisa e Desenvolvimento
PDB	Banco de dados de proteínas, do inglês <i>Protein Data Bank</i>
pIC_{50}	Menos o logaritmo na base dez do IC_{50} na unidade molar
Phe	Fenilalanina
PLS	<i>Partial Least Squares</i> ou regressão por mínimos quadrados parciais
QSAR	<i>Quantitative Structure Activity Relationship</i>
RFZ	<i>Rotation Function Z-score</i>
RMN	Ressonância Magnética Nuclear
RMSD	Desvio médio quadrático, do inglês <i>Root Mean Square Deviation</i>
rpm	Rotações por minuto

Ro5	<i>Ruler of five</i> ou regra dos cinco de Lipinski
ROCS	<i>Rapid Overlay for Chemical Structures</i>
S	Entropia
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SAD	<i>Single Wavelength Anomalous Dispersion</i>
SAR	Relação estrutura atividade, do inglês <i>Structure Activity Relationship</i>
SBDD	<i>Structure-Based Drug Design</i>
SDS-PAGE	Gel de poliacrilamida-docecil sulfato de sódio para eletroforese
Ser	Serina
SPR	Ressonância plasmônica de superfície ou <i>Surface Plasmon Resonance</i>
SSRL	<i>Stanford Synchrotron Radiation Lightsource</i>
$t_{1/2}$	Tempo de meia vida
TcCYP51	Enzima esterol 14- α -demetilase
TFZ	<i>Translation Function Z-score</i>
TLS	<i>Translation, libration, screw</i> – parâmetros de deslocamento atômico
t_R	Tempo de residência
TR	Tripanotona Redutase
Trp	Triptofano
TSA	Análise térmica diferencial, do inglês <i>Thermal Shift Analysis</i>
Tyr	Tirosina
UCSD	Universidade da Califórnia em San Diego
UCSF	Universidade da Califórnia San Francisco
USP	Universidade de São Paulo
V_M	Coeficiente de Matthews
VS	Triagem virtual, do inglês <i>Virtual Screening</i>
WDI	Índice mundial de fármacos, do inglês <i>World Drug Index</i>

SUMÁRIO

CAPÍTULO I	17
<i>Introdução</i>	18
1.1 A doença de Chagas e o agente etiológico <i>Trypanosoma cruzi</i>	18
1.2 Alvos terapêuticos para o desenvolvimento de novas quimioterapias	21
1.3 Cruzaína: alvo validado para a identificação de agentes tripanocidas	23
1.4 Métodos da química medicinal na gênese planejada de fármacos	30
1.4.1 <i>Métodos virtuais para o planejamento e a otimização de inibidores</i>	35
1.4.2 <i>Planejamento de fármacos baseado em fragmentos</i>	40
1.5 Inibidores da cruzaína utilizados nos ensaios cristalográficos iniciais	47
1.6 Referências bibliográficas	49
CAPÍTULO II	58
Objetivos	58
CAPÍTULO III	60
Expressão, Purificação e Cristalização da Cruzaína Nativa e das Mutantes C25S e C25A para os Ensaios Cristalográficos	60
3.1 Da obtenção da proteína pura até os ensaios de cristalização	61
3.1.1 <i>Métodos em biologia molecular para expressão e purificação de proteínas</i>	61
3.1.2 <i>Cristalização de proteínas</i>	64
3.2 Produção da cruzaína nativa e ensaios de cristalização	69
3.2.1 <i>Material e métodos</i>	72
3.2.2 <i>Resultados e discussão</i>	74
3.3 Desenvolvimento da cruzaína mutante C25S e C25A	81
3.3.1 <i>Material e métodos: expressão, purificação e cristalização da C25S</i>	82
3.3.2 <i>Resultados e discussão</i>	83
3.4 Referências bibliográficas	87
CAPÍTULO IV	90
Coleta de Dados, e Elucidação Estrutural do MOB de Inibidores na Cruzaína Nativa e C25S Mutante	90
4.1 Fundamentos da cristalografia de raios X em macromoléculas	91
4.1.1 <i>Cristais e a simetria cristalina</i>	91
4.1.2 <i>Difração de raios X por monocristais</i>	92

4.1.3 <i>Função densidade eletrônica e o problema das fases</i>	98
4.2 Método cristalográfico: da coleta de dados até a validação do modelo.....	100
4.2.1 <i>Coleta de dados cristalográficos</i>	100
4.2.2 <i>Processamento e redução de dados coletados</i>	104
4.2.3 <i>Faseamento e solução estrutural</i>	106
4.2.4 <i>Refinamento e validação do modelo cristalográfico</i>	109
4.3 Material e métodos: coleta de dados e elucidação estrutural.....	113
4.4 Resultados: elucidação estrutural e análise do MOB de inibidores.....	114
4.4.1 <i>Análise do MOB do Neq0409 em complexo com a cruzaina nativa</i>	114
4.4.2 <i>Análise do MOB do Neq0176 em complexo com a cruzaina nativa</i>	120
4.4.3 <i>Análise do MOB do Neq0147 em complexo com a cruzaina C25S</i>	127
4.4.4 <i>Validação da cruzaina C25S como modelo cristalográfico</i>	132
4.5 Referências bibliográficas.....	135
CAPÍTULO V	139
Estruturas Cristalográficas de Alta Resolução, Ensaios de Novos Inibidores e Mapeamento do Sítio Catalítico da Cruzaina.....	139
5.1 A estabilização da auto-proteólise leva a estruturas de alta qualidade e resolução para complexos com a cruzaina nativa.....	140
5.2 Ensaios de cristalização de novos inibidores planejados com base nos modos de interação elucidados para o Neq0147 e para o Neq0409.....	147
5.2.1 <i>Proposição e teste de novas abordagens de cristalização para obtenção do MOB dos inibidores de segunda geração</i>	151
5.2.2 <i>Resultados preliminares e cristais da cruzaina crescidos no Brasil</i>	152
5.3 Sondagem do sítio catalítico da cruzaina com GRID e SuperStar em busca de pontos para incremento de potência e seletividade.....	157
5.4 Referências bibliográficas.....	166
CAPÍTULO VI	168
Conclusões e Perspectivas.....	168
6.1 Conclusões.....	169
6.2 Perspectivas.....	171
ANEXO I Paper publicado na Plos NTD.....	173
ANEXO II Paper “under consideration” para publicação no periódico Plos NTD.....	174

Capítulo 1

Introdução

Este capítulo aborda aspectos fundamentais da doença de Chagas e do seu agente etiológico *Trypanosoma cruzi*, seguidos da quimioterapia atualmente disponível e dos alvos terapêuticos empregados para a gênese planejada de fármacos. Na sequência, é feita uma breve descrição bioquímica e estrutural do alvo enzimático utilizado neste projeto, a cruzaina, e alguns de seus inibidores. Por último, os métodos de química medicinal utilizados para a seleção de compostos bioativos são abordados, seguidos da descrição de uma série de compostos bioativos desenvolvidos no laboratório NEQUIMED que foram objetos de estudo para o desenvolvimento inicial deste trabalho.

1 INTRODUÇÃO

1.1 A doença de Chagas e o agente etiológico *Trypanosoma cruzi*

A doença tropical parasitária Tripanossomíase Americana ou doença de Chagas como é conhecida, é a principal responsável por mortes devido à insuficiência cardíaca em regiões endêmicas da América Latina segundo a Organização Mundial da Saúde (OMS). Estima-se que cerca de oito milhões de pessoas estejam infectadas com outras 100 milhões correndo risco em regiões não endêmicas do mundo devido à imigração.^{1,2} A descrição do agente etiológico da doença, o protozoário kinetoplastídeo *Trypanosoma cruzi*, além do complexo ciclo de vida do parasito foi realizada pelo médico sanitarista brasileiro Carlos Chagas em 1909.³ Este ciclo envolve uma fase de multiplicação intracelular obrigatória no hospedeiro vertebrado (mamíferos) e extracelular no inseto vetor (triatomíneos) em três estágios de diferenciação celular: o tripomastigota (eliminado pelo inseto vetor e transmissível ao homem) e as formas replicativas amastigota (hospedeiro vertebrado) e epimastigota (inseto vetor), conforme descrito na Figura 1.1.^{2,4}

Figura 1.1 - Ciclo de vida do parasito *Trypanosoma cruzi*.

Fonte: Adaptado de BERN, C., 2011, p. 2529.⁴

A doença se desenvolve em uma fase aguda que ocorre nos primeiros meses após a infecção pelo parasito com poucos sintomas tais como febre, vômito, dor de cabeça. É possível conviver com a doença durante anos sem sentir seus efeitos, que em geral, se manifestam no final da fase crônica podendo causar a morte lenta do hospedeiro se não tratado devido a alterações no miocárdio e no sistema digestivo, conforme apresentado na Figura 1.2. Estas alterações levam à cardiomiopatia, megaesôfago, megacólon, meningoencefalite aguda, dentre outros efeitos.^{3,5} Este quadro clínico da doença durante a fase crônica é decorrente da ação direta do parasito e indireta da resposta autoimune do hospedeiro que ataca e destrói tecidos saudáveis do corpo por engano levando a um processo inflamatório autossustentável que persiste mesmo após o tratamento antiparasitário.^{2,6} Desta forma, a evolução da doença diminui drasticamente a qualidade de vida e a capacidade produtiva do paciente chagásico, levando a grande morbidade, mortalidade e vários outros problemas socioeconômicos quando comparados com outras doenças como malária, esquistossomose e leishmaniose.^{1,2}

Figura 1.2 - Radiografia do tórax mostrando o aumento progressivo do coração e de órgãos do sistema digestivo de um paciente chagásico em 5 (A), 10 (B), 15 (C) e 20 (D) anos de evolução do estágio crônico da doença.

Fonte: Adaptado de MUÑOZ, S. S., 2012, p. 50.⁵

Esta doença atinge principalmente regiões pobres e rurais da América Latina, devido a maior incidência do inseto transmissor mais popular, o barbeiro. Contudo o parasito também pode ser transmitido por transfusão sanguínea, transmissão congênita durante a gestação, sendo difundido para outras regiões não endêmicas do globo devido as constantes imigrações tornando-se assim um problema de saúde mundial.^{2,5} Mais de cem anos já se passaram desde as descobertas de Carlos Chagas e pouco progresso foi realizado no tratamento da doença, parte devido ao desinteresse das indústrias farmacêuticas que veem poucos atrativos econômicos para o desenvolvimento de novas quimioterapias sendo considerada deste modo, uma doença negligenciada.^{3,7} Na Figura 1.3 são apresentadas as estruturas

químicas dos dois fármacos atualmente disponíveis para o tratamento desta doença. O Benzonidazol (Rochagan[®], Roche) e Nifurtimox (Lampit[®], Bayer) foram introduzidos há mais de 40 anos e têm demonstrado uma eficiência limitada contra o estágio crônico da doença.⁶ A estes fármacos são ainda associados severos efeitos colaterais, grande variabilidade na eficiência terapêutica para pacientes vindos de diferentes regiões geográficas que aliados ao longo período de tratamento (meses), levam ao frequente abandono da quimioterapia pelos pacientes.^{4,7,8}

Figura 1.3 - Estruturas moleculares dos fármacos atualmente disponíveis para o tratamento da doença de Chagas.

Fonte: Autoria própria.

Os mecanismos de ação do Benzonidazol e do Nifurtimox são ainda pouco conhecidos. Estudos têm demonstrado que o grupo nitro em ambos os fármacos é reduzido a grupos amino pela ação de nitroredutases, levando a formação de vários radicais livres intermediários e metabólitos eletrofílicos que causam modificações covalentes de macromoléculas do parasito como o DNA.^{8,9} O Nifurtimox também atua pela produção de espécies de oxigênio reativas, tais como superóxidos e peróxidos, pelos quais o parasito tem pouca capacidade de desintoxicação.¹⁰ Todos estes intermediários reativos são responsáveis pelos mecanismos de estresse redutivo que promovem as ações deletérias observadas no *T. cruzi* e consequente ação farmacológica.^{9,10} Como estes intermediários também interagem com DNA, proteínas e lipídeos dos tecidos do hospedeiro, estes são igualmente responsáveis pelos efeitos toxicológicos observados em mamíferos.^{8,10}

A doença de Chagas é uma condição complexa resultante do sucesso na invasão e estabelecimento do *T. cruzi* em células humanas de tipo variado incluindo nervosas, musculares e endoteliais. O parasito então passa a utilizar toda a maquinaria celular do hospedeiro para produzir centenas de novos parasitos que infectarão os tecidos vizinhos.^{2,6,7} O fato do *T. cruzi* ser um parasito intracelular representa, portanto, um desafio extra para o desenvolvimento de novos fármacos, uma vez que estes devem ser capazes de atravessar a membrana plasmática

celular do hospedeiro antes de atuarem no parasito.^{2,3} O mesmo não ocorre com outros parasitos da família *trypanosomatidae* como os que causam a leishmaniose e o *Trypanosoma brucei* que causa a tripanossomíase africana ou doença do sono. Estes dois parasitos permanecem no sangue e são potencialmente mais acessíveis aos fármacos.³ Em virtude das limitações da quimioterapia atual e das dificuldades para se eliminar o *T. cruzi* sem prejudicar o hospedeiro humano, novos fármacos mais eficientes precisam ser urgentemente desenvolvidos para o tratamento desta doença negligenciada que representa enorme problema de saúde mundial.

1.2 Alvos terapêuticos para o desenvolvimento de novas quimioterapias

Recentes progressos na compreensão da biologia e bioquímica do *T. cruzi*, têm possibilitado a identificação de múltiplos alvos enzimáticos para o planejamento racional de quimioterapias mais eficientes contra esta doença.¹¹ Dentre os alvos promissores identificados, destaca-se a **esterol 14- α -demetilase (TcCYP51)** que é uma enzima chave da via de biossíntese de ergosterol, sendo este um componente estrutural essencial da membrana celular que modula o desenvolvimento do parasito.^{12,13} Estudos realizados na última década, demonstraram que o *T. cruzi*, assim como fungos e leveduras, requerem esteróis específicos para a proliferação e a viabilidade de células em todos os estágios dos seus ciclos de vida.^{6,14} Desta forma, são suscetíveis a inibidores que atuam na via de biossíntese de ergosteróis, levando ao reposicionamento de fármacos do tipo azóis e triazóis antifúngicos para a terapia da doença de Chagas.¹³ Alguns destes inibidores comercialmente disponíveis e altamente eficientes no tratamento de doenças antifúngicas tais como o cetoconazol, intraconazol (Figura 1.4), exibiram atividade supressora, porém não curativa contra as infecções de *T. cruzi* em modelos animais.^{6,11}

O posaconazol é um recente membro desta família de antifúngicos que tem demonstrado potencial antichagásico tendo atualmente concluído a fase II das pesquisas clínicas desenvolvidas pela indústria farmacêutica Merck.^{7,14} Outros dois inibidores do citocromo P-450 de fungos que foram potentes inibidores da TcCYP51 e também concluíram a fase clínica II de seu desenvolvimento, foi o pro-fármaco do antifúngico ravocunazol (o E12-24) e o antifúngico Tak-187, ambos desenvolvido no Japão pelos laboratórios Eisai e Takeda respectivamente (Figura 1.4).^{7,14} Estes são os primeiros compostos a exibirem atividades curativas em ambos os estágios da

doença e em cepas infectadas com *T. cruzi* oriundos de diferentes regiões geográficas.¹⁴ As notáveis atividades antiparasitárias *in vivo* observadas para estes derivados triazólicos decorrem da combinação da potência e seletividade intrínsecas anti *T. cruzi*, e suas especiais propriedades farmacocinéticas tais como o longo tempo de meia vida e amplo volume de distribuição.^{6,14} Contudo, estes fármacos contendo azóis não serão aprovados para o uso contra a doença de Chagas devido aos resultados insatisfatórios obtidos na fase clínica II de seu desenvolvimento.¹³

Figura 1.4 - Inibidores da via metabólica do ergosterol com atividade antichagásica.

Fonte: Adaptado de DUSCHAK, V. G., 2007, p. 22.¹¹

Grande parte da energia utilizada pelo parasito é obtida pela via glicolítica, fazendo das enzimas desta rota bioquímica potenciais alvos terapêuticos. Uma enzima chave nesta rota é a **gliceraldeído-3-fosfato desidrogenase (GAPDH, EC 1.2.1.12)**. Este homo tetrâmero é responsável por catalisar a fosforilação oxidativa do D-gliceraldeído-3-fosfato (G3P) em 1,3-difosfoglicerato mediante a redução reversível da Nicotinamida Adenina Dinucleotídeo (NAD⁺) para a forma reduzida NADH sendo essencial para a obtenção de energia pelo parasito na sua forma amastigota.^{11,15} Grande identidade na sequencia primária (cerca de 90%) é notada para a GAPDH nos parasitos patogênicos *T. cruzi*, *T. brucei* e *Leishmania mexicana*, possuindo, no entanto, diferenças estruturais importantes para o desenho de drogas quando comparadas à proteína homóloga em humanos, cuja identidade é de 45%.¹⁵ Esta diferença tem explicação evolutiva. Os tripanossomos pertencem à ordem dos cinetoplastídeos, sendo esta uma das linhagens mais antigas de protozoários que se divergiram evolutivamente muito cedo em relação às linhagens comuns de eucariotos.¹⁶ Portanto, várias rotas bioquímicas divergem daquelas observadas em células dos mamíferos, promovendo alvos seletivos para o desenvolvimento racional

de fármacos. Avanços no entendimento bioquímico e estrutural da GAPDH no *T. cruzi*, tem possibilitado a seleção de potentes inibidores desta enzima.^{11,17,18}

Outro alvo validado é a **dihidroorotato-desidrogenase (DHODH, EC 1.3.99.11)**. Esta é a quarta enzima da rota de biossíntese *de novo* de nucleotídeos de pirimidina, catalisando a produção de orotato a partir do dihidroorotato mediante a redução de fumarato a succinato.^{11,19} Estudos de *knockout* mostraram que na ausência desta enzima, o *T. cruzi* não é capaz de sobreviver mesmo na presença de nucleosídeos de pirimidina, confirmando assim a dependência desta rota para o desenvolvimento do parasito.²⁰ Estruturas cristalográficas da DHODH de *T. cruzi* no seu estado nativo e em complexo com succinato, fumarato, orotato e dihidroorotato (código PDB: 2DJX, 2DJL, 2E6D, 2E6A e 2E68 respectivamente), possibilitaram que novos inibidores fossem planejados usando métodos *in silico* baseados na estrutura do alvo.^{19,21} A **tripanotiona redutase (TR)** é outro alvo atraente que atua na via de síntese de tripanotiona sendo uma enzima chave na defesa antioxidante e na manutenção do ambiente intracelular, cuja inibição causa uma condição de estresse oxidativo no parasito.¹¹ Contudo, são proteases tais como a **oligopeptidase B** (uma serino protease envolvida na sinalização de Ca^{2+} durante a invasão no hospedeiro) e a cisteíno protease **cruzaína** (usada como alvo neste projeto) que têm se destacado devido as múltiplas funções que desempenham na patogenicidade da doença.^{11,22}

1.3 Cruzaína: alvo validado para a identificação de agentes tripanocidas

A cruzaína, forma recombinante da cruzipaína (EC 3.4.22.51), é o membro mais abundante da família da papaína de cisteíno proteases encontradas no *T. cruzi*. Esta enzima está envolvida no processo de infecção da célula do hospedeiro, replicação, diferenciação, metabolismo, evasão de mecanismos de defesa sendo essencial na regulação da interação entre parasito e hospedeiro. A vulnerabilidade do parasito a inibidores da cruzaína somados a falta de sistemas de redundância para esta enzima, fazem desta um alvo terapêutico validado importante no combate a doença de Chagas.²³⁻²⁵ A cruzaína é expressada em todos os estágios do ciclo de vida do *T. cruzi* como uma mistura complexa de isoformas, estando localizada em diferentes compartimentos na célula do parasito. No estágio epimastigota, onde são observados níveis de expressão até dez vezes maiores em relação aos outros estágios, a protease está confinada em uma organela pré-lisossomal conhecida

como reservossoma.^{26,27} Para o estágio intracelular amastigota, chave no tratamento da doença, isoformas da protease são encontradas no reservossoma e ainda ligado à face externa da membrana plasmática diretamente em contato com o citoplasma da célula do hospedeiro.²⁶ Portanto, a cruzaina pode operar em dois microambientes químicos: no reservossoma (pH 6,3) e na superfície do parasito (pH 7,4).²⁸

A cruzaina é uma glicoproteína expressa como uma proenzima com três domínios distintos: o pró-domínio N-terminal com 104 aminoácidos, o domínio catalítico com 215 e uma extensão carboxi- ou C-terminal com 130. Existem três potenciais sítios de glicosilação sendo um na porção C-terminal, e outros dois no domínio catalítico desta enzima.^{27,29} Acredita-se que seja pela região de glicosilação da porção C-terminal que a cruzaina se liga à superfície do parasito na forma amastigota.^{30,31} Esta extensão C-terminal é incomum em outros membros da família da papaína sendo encontrado apenas na homóloga proveniente do parasito *T. brucei rodesaiense*, a rodesaina.³² Como outras proteases da família da papaína, a cruzaina é inicialmente expressa como um zimogênio inativo que é auto ativado em um processo dependente do pH, da temperatura e da concentração, transformando-se na enzima ativa mediante quebra e digestão do pró-domínio em cadeias menores de aminoácidos.^{33,34} Este processo é realizado no complexo de Golgi do parasito por um processo ainda pouco conhecido. Estudos têm demonstrado que o pró-domínio N-terminal da pró-cruzaina é ligado à membrana deste complexo via receptores análogos aos identificados para as pró-catepsinas L e D em humanos. Ainda por analogia, é sugerido que uma mudança conformacional do pró-domínio ocorra na região que interage com o sítio catalítico, possivelmente induzido pela entrada no compartimento de menor pH do complexo de Golgi, resultando na auto ativação de algumas moléculas da cruzaina que desencadeia uma cascata de trans-ativações para os demais zimogênios.^{26,35} Uma vez processada e ativada, a enzima livre sem o pró-domínio é então liberada para o reservossomo. Estudos confirmaram que a extensão C-terminal ainda continua ligado ao domínio catalítico da cruzaina mesmo após o transporte para o reservossomo em algumas células do amastigota. O mesmo não foi observado na forma epimastigota, onde o pró-domínio e a extensão C-terminal são proteoliticamente removidos da cruzaina ativa.^{26,30}

Estruturalmente, o domínio catalítico é constituído de um monômero contendo 215 aminoácidos (22,5 kDa) cujo sítio catalítico apresenta alto grau de identidade tanto sequencialmente quanto estruturalmente com as enzimas catepsinas B e L de

humanos.^{33,34} O sítio catalítico está localizado em uma fenda entre dois sub domínios principais (α hélices e folhas β), onde está posicionada a tríade catalítica formada pelos resíduos Cys25, His162, e Asn182. Um quarto resíduo, Gln19, está envolvido na estabilização da cavidade do oxiânion que se forma.³³ Assim como em outros membros da família da papaína o sítio catalítico contém sete subsítios onde ocorrem interações com os aminoácidos do substrato cuja ligação peptídica será hidrolisada. Quatro subsítios na direção C-terminal da enzima a partir da tríade catalítica (S1, S2, S3, S4) interagem com os aminoácidos das posições P1, P2, P3, P4 do substrato, enquanto três subsítios na direção N-terminal (S1', S2' e S3') interagem com os respectivos aminoácidos, conforme exibido na Figura 1.5. Apenas quatro destes subsítios são atualmente explorados no planejamento de fármacos (S3, S2, S1, S1') e somente dois mostram significativa seletividade para interações (S2 e S3).^{36,37}

Figura 1.5 - Sítio catalítico da cruzaina em superfície de *Van der Waals* colorida de acordo com os subsítios de ligação. O peptídeo mostrado foi retirado da estrutura da catepsina K (PDB: 1BY8), uma vez que ainda não existem estruturas cristalográficas da cruzaina ligada a substratos peptídicos.

Fonte: DURRANT, J., 2010, p. 676.³⁷

Como na maioria das cisteíno proteases da família da papaína, as interações no sítio S2 da enzima com o resíduo P2 complementar são chave nos fatores que definem sua especificidade. O subsítio S2 é mais profundo e melhor definido para interações, aceitando uma gama maior de substratos em comparação a proteases homólogas de outros parasitos tais como a rodesaína, sendo ainda responsável pela seletividade da enzima em relação às catepsina L e S de humanos.^{38,39} Grande parte desta especificidade e seletividade é devido ao resíduo Glu208 localizado na base

do bolsão, que é capaz de mudar de conformação para interagir de maneira eficiente tanto com resíduos de aminoácidos hidrofóbicos quanto com resíduos básicos carregados em um mecanismo dependente do pH. Logo, o Glu208 pode adotar uma conformação dirigida ao solvente quando a fenilalanina (Phe) está no S2, ou adotar uma conformação orientada para o substrato quando uma arginina (Arg) ocupa este subsítio, ocorrendo à formação de uma ponte salina com o grupo guanidino carregado, podendo ainda realizar uma ligação de hidrogênio na presença de tirosina (Tyr) conforme representado na Figura 1.6.⁴⁰ Compostos contendo Phe em P2 são até 15 vezes mais potentes em relação aos mesmos compostos contendo Arg em ensaios feitos no pH 6,0, provavelmente porque o resíduo Phe realiza melhores contatos hidrofóbicos quando o Glu208 está orientado para o solvente.³⁹

Figura 1.6 - Habilidade do aminoácido Glu208 na base do sítio S2 da cruzaina em modificar sua conformação para interagir com os diferentes substratos Phe, Arg e Tyr.

Fonte: GILLMOR, S. A., 1997, p. 1607.⁴⁰

A maioria dos atuais inibidores da cruzaina mimetiza substratos naturais peptídicos (peptídeo-miméticos), sendo divididos em irreversíveis e reversíveis.¹¹ Dentre os inibidores que interagem por mecanismo covalente irreversível, destacam-se os que possuem os grupos de reconhecimento químico (*warheads*) baseados em diazometano, diazometil-, fluorometil-cetonas, dipeptidil- α,β -epóxi-cetonas, epóxidos peptídicos (dentre eles os conhecidos E-64, E-64c, E-64-d), alil- e vinil sulfonas.⁴¹⁻⁴⁴ Vinil sulfonas e vinil sulfonamidas são conhecidas como excelentes aceitadores de Michael sendo potentes inibidores do tipo tempo-dependente de várias cisteíno proteases. Outra vantagem é que estas são praticamente inertes a serino proteases.

^{45,46} Centenas de vinil sulfonas dipeptídicas foram sintetizadas e avaliadas bioquimicamente, tendo como objetivo sondar o sítio catalítico da cruzaina em busca de potenciais incrementos de potência e seletividade.^{42,45-47} Dentre os inibidores que atuam por mecanismos reversíveis destacam-se os que possuem *warheads* do tipo aldeídos peptídicos, α -cetonas, diamino-cetonas e nitrilas que formam um complexo covalente reversível com a cisteína catalítica do tipo hemitioacetal, cetel e tioimidato respectivamente, mimetizando o adulto covalente inicial que é formado durante o mecanismo de hidrólise proteolítica.⁴⁸ Derivados de α -cetonas tais como as α -ceto-amidas, α -ceto-ácidos e α -ceto-éster, acil-hidrazidas e aril-uréias, também exibiram potência de inibição na faixa nanomolar nos ensaios *in vitro*, sendo porém, menos potentes quando comparados aos inibidores irreversíveis previamente exibidos.⁴⁹⁻⁵¹ A estrutura química destes *warheads* que também são frequentemente encontrados em inibidores de cisteíno proteases, estão apresentadas na Figura 1.7.

Figura 1.7 - *Warheads* frequentemente encontrados em inibidores da cruzaina. As setas apontam para o átomo eletrofílico susceptível ao ataque nucleofílico pelo grupo tiol ativado do resíduo catalítico Cys25.

Fonte: Adaptado de MALVEZZI, A., 2009, p. 783.⁵²

Da pequena fração de inibidores não peptídeo-miméticos identificados, destacam-se as tiossemicarbazonas e semicarbazonas que tem exibido potência de inibição na faixa nanomolar. Muitos destes são pequenos e apresentam atividades tripanosomicidas contra amastigotas intracelulares *in vivo*, sendo candidatos atraentes para desenvolvimento de novos fármacos.⁵³⁻⁵⁷ Estes compostos também têm apresentado grande afinidade por proteases homólogas tais como a falcipaina e

a rodesaína provenientes de outros parasitos.^{55,56} Recentemente uma série de inibidores não peptídicos, reversíveis foi identificada utilizando métodos *in silico* de triagem, ensaios bioquímicos e cristalografia de raios X, inibindo a cruzaina na faixa sub-micromolar. Cinco destes inibidores foram ativos nos ensaios celulares contra formas tripomastigotas infectivas do *T. cruzi* exibindo potência comparável ao Benzonidazol, utilizado como controle.^{58,59} Inibidores da cruzaina causam a morte do parasito devido à interrupção do processamento auto catalítico natural da cruzaina e subsequente transporte para os lisossomos levando a um acúmulo massivo de procruzaina não processada no complexo de Golgi, conforme esquema da Figura 1.8. As anormalidades que ocorrem no complexo de Golgi são seguidas da distensão do retículo endoplasmático, da membrana nuclear e da mitocôndria além da proliferação de vesículas anormais para o citoplasma. Todas estas alterações promovem um estresse fisiológico e consequente morte do parasito.^{26,28,60}

Figura 1.8 - Mecanismo de morte do *T. cruzi* via inibidores da cruzaina. Neste modelo, a procruzaina é glicosilada no complexo de Golgi sendo ativada e liberada de receptores de membrana. A cruzaina matura é então acumulada nos reservossomos e parcialmente secretada em epimastigotas (CA-I/72). Na presença do inibidor, ocorre acúmulo de protease não processada no complexo de Golgi causando a morte do parasito (CA-I/72+CPI).

Fonte: ENGEL, J. C., 2000, p. 1353.²⁸

A N-piperazina-fenil-homofenil-vinilsulfona (K777, Figura 1.9A) é atualmente o inibidor mais estudado desta enzima, tendo contribuído para a validação da cruzaina como alvo terapêutico e entendimento do mecanismo de morte do parasito por inibidores da cruzaina.^{23,26,61} O K777 é também o inibidor mais avançado na gênese planejada de fármacos estando no final da fase pré-clínica de seu desenvolvimento.⁷

Estudos nesta fase tem demonstrado eficácia em modelos animais da doença de Chagas, sendo não mutagênico e bem tolerado tanto na fase aguda quanto crônica da doença.⁶¹⁻⁶³ Seu modo de interação foi determinado por cristalografia de raios X (Figuras 1.9B e C), confirmando que esta potente vinilsulfona inibe a cruzaina via ligação covalente irreversível que é formada com a cisteína catalítica (Cys25).⁴⁶ Sua considerável potência também está relacionada à complementaridade ao sítio catalítico que maximiza interações eletrostáticas e, conseqüentemente, o efeito hidrofóbico (Figura 1.9B), além das ligações de hidrogênio favoráveis realizadas entre o esqueleto peptídico do inibidor e importantes aminoácidos do sítio catalítico (Figura 1.9C).⁴⁶ Este inibidor também tem demonstrado potencial de inibição para várias outras cisteíno proteases, incluindo a rodesaina e a falcipaina existentes nos parasitos *Trypanosoma brucei* e *Plasmodium falciparum* respectivamente.⁴⁶

Figura 1.9 - Estrutura química do K777 (A), modo de interação em superfície hidrofóbica (B) e interações intermoleculares com os principais aminoácidos do sítio catalítico (C).

Fonte: KERR, I. D., 2009, p. 25701.⁴⁶

Apesar de várias propriedades favoráveis, alguns aspectos do peptídeo-mimético K777 são sub otimizados do ponto de vista da gênese de fármacos tais como eficácia limitada em estudos farmacocinéticos e potenciais efeitos *off target* em que o inibidor poderia interagir de maneira irreversível com proteases homólogas em humanos como as catepsinas B e L, causando efeitos colaterais.^{61,64} O K777 também é inibidor irreversível da CYP3A4, uma enzima responsável pelo metabolismo de muitos fármacos incluindo si mesmo.⁶⁴ Além disto, a seleção de inibidores peptídeo-miméticos também tem sido evitada devido à possibilidade de serem facilmente hidrolisados e metabolizados por outras enzimas antes mesmo de atuarem na enzima alvo.⁶¹ Um inibidor potente e seletivo que interaja por mecanismos competitivos reversíveis, apresentando ainda um perfil farmacocinético do tipo fármaco-similar (*drug-like*) seria o candidato a fármaco ideal.⁴¹

1.4 Métodos da química medicinal na gênese planejada de fármacos

Até meados do século XX, informações bioquímicas e estruturais disponíveis para identificação de moléculas bioativas eram escassas quando comparada a infinidade de detalhes promovidos pela moderna química experimental atual. Assim, os primeiros fármacos eram isolados principalmente de plantas e identificados via métodos casuais. Desta forma, vários fármacos tais como o antibiótico Penicilina, o ansiolítico Clordiazepóxido e o neuroléptico Clorpromazina, foram introduzidos na terapêutica.⁶⁵⁻⁶⁷ Métodos empíricos pouco racionais também levaram a identificação do Nifurtimox e do Benzonidazol como agentes anti-chagásicos há mais de quatro décadas, cujos mecanismos de ação farmacológica e de efeitos toxicológicos só foram elucidados vários anos após o início da sua comercialização.^{6,8} Atualmente, o processo de obtenção de novos fármacos está se tornando cada vez mais racional, apoiado em sólida fundamentação teórica e em métodos que fornecem informações valiosas dos processos bioquímicos e fisiológicos de uma determinada doença, e do mecanismo de ação dos fármacos a nível molecular.⁶⁷ Além disto, agências reguladoras tais como a Agência Nacional de Vigilância Sanitária (Anvisa), a americana *Food and Drug Administration* (FDA) e a *European Medicines Agency* (EMA) regulamentam e fiscalizam o lançamento, fabricação e comercialização de fármacos.⁶⁷ Todos estes fatores têm contribuído para que fármacos mais seguros e eficazes sejam lançados. Contudo, a crescente complexidade na regulamentação por parte destas agências têm sido apontada pelas indústrias farmacêuticas como um motivo pelo longo tempo necessário para a aprovação de um novo produto.^{67,68}

A Química Medicinal desempenha um papel fundamental dentro do atual processo de **descoberta** e **desenvolvimento** de novos fármacos.^{68,69} Segundo a União Internacional de Química Pura e Aplicada (IUPAC), a química medicinal constitui uma área inter e multidisciplinar do conhecimento baseado essencialmente na Química, mas que também envolvem princípios das Ciências Biológicas, Médicas e Farmacêuticas, que se preocupam com a descoberta, planejamento, identificação e preparo de compostos biologicamente ativos, bem como, com o estudo de seu metabolismo, a interpretação de seu modo de ação a nível molecular e a construção de relações entre estrutura química e atividade biológica.⁶⁷ Este processo de descoberta e desenvolvimento de novos fármacos englobado pela química medicinal é dividido nas fases pré-clínica e clínica, conforme exibido na Figura 1.10.⁶⁵

Figura 1.10 – Estágios do processo de descoberta e desenvolvimento de fármacos.

Fonte: Adaptado de LOMBARDINO, J. G., 2004, p. 854.⁶⁵

O processo de descoberta de um fármaco se inicia com a identificação da necessidade de uma nova terapia ou adequação das terapias existentes. Após obter conhecimentos necessários sobre a doença alvo, são iniciados ensaios em modelos biológicos que visam detectar compostos bioativos *in vitro* (**hits**) e a detecção de um composto com uma atividade relacionada no modelo *in vivo* correspondente.⁶⁵ Estruturas análogas são sintetizadas para a maximização desta atividade e por fim são selecionados os **compostos matrizes** (*lead-like*) como candidatos a seguirem no desenvolvimento. Um composto matriz representa um ponto de partida sobre o qual incidirá futuros ensaios que visam refinar a resposta biológica para findar em um fármaco seguro com ação terapêutica desejada.^{67,68} Estes candidatos são submetidos a avaliações toxicológicas em animais e se forem aprovados em todos os testes, terão os dados da pesquisa encaminhados para a solicitação da licença para investigação de um novo fármaco (IND) em agência reguladora como o FDA. Quando esta autorização é concedida, estes candidatos são sintetizados em larga escala para a realização de todos os estudos da fase clínica.^{65,67}

A proposta da fase clínica é aperfeiçoar as propriedades farmacocinéticas e farmacêuticas (formulação química) para que estes candidatos se tornem aptos ao uso clínico. Desta forma, são realizadas avaliações sequenciais da tolerância em voluntários humanos saudáveis (Fase I), da eficácia, dosagem e segurança em pacientes (Fase II) seguidos de uma ampla avaliação em milhares de pacientes de

diferentes regiões geográficas para a compilação de dados acerca da eficácia, segurança e propriedades farmacocinéticas (Fase III).^{65,67} Para os poucos candidatos a fármacos que sobrevivem a esta série de ensaios de desenvolvimento (4-7%), todos os dados acumulados durante a pesquisa são submetidos a ANVISA na forma de um pedido de registro do novo fármaco, ou uma NDA (*New Drug Application*) no caso do FDA. Somente com a aprovação do órgão regulamentador, o novo fármaco poderá ser prescrito pelos médicos para tratar a doença a qual foi designado. Monitoramentos pós-comercialização dos novos fármacos ocorrem na fase IV, sendo esta também conhecida como farmacovigilância.^{65,67}

Este processo de descoberta e desenvolvimento (P&D) de novos fármacos é longo, complexo e caro. P&D para a maioria dos fármacos atualmente disponíveis demandaram entre 12 e 24 anos desde o início do projeto até o lançamento, tendo um custo estimado em cerca de 1,4 bilhões de dólares para cada novo fármaco.^{65,70} Contudo, este valor pode ser muito maior quando também contabilizados os custos envolvidos no desenvolvimento de potenciais fármacos que falharam nos ensaios clínicos. A Astra Zeneca por exemplo, tem gasto cerca de 12 bilhões de dólares para cada novo fármaco aprovado (dados de 2013). Este é um valor muito maior do que a receita anual obtida pelo seu fármaco mais vendido, o Esomeprazol (Nexium®) utilizado no tratamento de úlceras gastrointestinais, cuja receita em 2013 foi de 8,3 bilhões de dólares.⁷¹ Atualmente o fármaco mais vendido no mundo é a atorvastatina (Lipitor® da Pfizer), utilizada para controlar os níveis de colesterol no sangue, que só em 2013 vendeu 12,5 bilhões de dólares.⁷¹ A fase clínica é a mais onerosa de todo o processo de P&D, devido aos rigorosos testes clínicos em humanos que devem ser realizados, sendo em geral, financiada por indústrias farmacêuticas. Estudos têm demonstrado que somente um a cada 25 candidatos a fármacos sobrevive aos rígidos testes de segurança e eficácia requeridos para que estes se tornem produtos comerciais, sendo que ao final vários destes não conseguirão recuperar os custos de desenvolvimento no competitivo mercado farmacêutico.⁶⁵

O reposicionamento de fármacos tem sido uma abordagem alternativa usada por indústrias farmacêuticas para reduzir tempo, custos e riscos em P&D. Este método consiste em encontrar novos usos terapêuticos para fármacos já existentes. Riscos e tempo de P&D são reduzidos porque candidatos ao reposicionamento já passaram pelos estágios clínicos do desenvolvimento e, portanto, seus perfis de segurança e farmacocinéticos já estão bem estabelecidos. Além disto, otimizações

químicas, toxicológicas, processo de fabricação e formulação já foram completados.^{72,73} A Talidomida representa um exemplo de reposicionamento. Este fármaco foi comercializado como um sedativo e hipnótico a partir de 1957 sendo prescrito para milhares de mulheres grávidas para combater enjoos matinais. Os procedimentos de ensaios de fármacos naquela época eram menos rígidos e por isto, testes realizados não revelaram seus efeitos teratogênicos que foram responsáveis por milhares de casos de focomelia onde crianças passaram a nascer com malformação congênita. A Talidomida era vendida como uma mistura racêmica ficando constatado após alguns anos, que o efeito teratogênico observado era causado pelo enantiômero S.⁷⁴ Este incidente foi um marco para a importância do controle estereoquímico na produção de fármacos. Embora este fármaco esteja associado aos mais terríveis acidentes médicos da história, novos usos terapêuticos foram encontrados para o tratamento da lepra, lúpus sistêmico e alguns sintomas da AIDS.^{75,76}

Apesar de menos onerosa, a fase pré-clínica envolve uma enorme quantidade de compostos e testes mais elaborados que exigem conhecimentos especializados e o desenvolvimento de tecnologias que muitas vezes só estão disponíveis em centros acadêmicos de pesquisa. Neste caso, a maior parte do financiamento provém de agências governamentais e de doações da iniciativa privada que visam proporcionar desenvolvimento do conhecimento em benefício de pacientes que sofrem com graves doenças, ao contrário das indústrias farmacêuticas que visam e necessitam do lucro.^{70,77} De maneira geral, as etapas do processo pré-clínico realizadas em centros acadêmicos concentram-se basicamente na (i) identificação, seleção e validação do alvo biológico, (ii) identificação do ligante (*hit*), (iii) otimização dos ligantes a compostos líderes e (iv) a otimização dos compostos líderes.⁷⁸

Um dos aspectos mais importantes na descoberta de novos fármacos é a busca por compostos líderes.⁷⁹ Assim, os tradicionais métodos automatizados de triagem em massa (HTS ou *High Throughput Screening*,) e os modernos métodos de triagem virtual (VS ou *Virtual Screening*) têm sido empregados para varrer o vasto espaço químico na busca por substâncias bioativas que causem o efeito terapêutico desejado sem os intoleráveis efeitos colaterais em humanos.^{58,78-81} O termo espaço químico é dado ao conjunto de todos os compostos químicos teoricamente possíveis de serem obtidos, cujas estimativas são de que este número ultrapasse a 10^{200} substâncias. Mesmo com o extenso conhecimento em química orgânica atual, apenas uma pequena fração destes compostos é plausível de ser sintetizada.⁸²

Segundo o *Chemical Abstract* (CAS), 90 milhões de compostos foram identificados sendo estes obtidos tanto por rotas sintéticas ou por purificações de produtos naturais.⁸³ Define-se como **espaço químico-biológico** a pequena fração deste conjunto que exibe alguma atividade biológica, conforme exibido na Figura 1.11.⁸³

Figura 1.11 - Esquema representativo do espaço químico-biológico. O contínuo espaço químico em azul claro apresenta áreas discretas ocupadas por compostos com afinidade específica para moléculas biológicas como proteases (roxo), GPCRs lipofílicas (azul) e cinases (vermelho). A região de intersecção independente destes compostos com propriedades fármaco-similares é mostrada em verde.

Fonte: LIPINSKI, C., 2004, p. 855.⁸³

No entanto, nem todo composto biologicamente ativo tem propriedades físico-químicas desejadas para ser um candidato a fármaco.⁸⁴ Os fatores que regem o caminho aleatório percorrido pelo fármaco desde sua administração (que pode acontecer por várias vias como a oral, intravenosa e nasal) até o seu sítio de ação e a sua eliminação após ter produzido o efeito farmacológico, são controlados pelas propriedades farmacocinéticas.⁶⁷ Os eventos experimentados pelo fármaco na fase farmacocinética são respectivamente a Absorção, Distribuição, Metabolismo e Excreção ou simplesmente **ADME**.⁸⁵ Portanto um fármaco precisa atravessar membranas biológicas e organelas durante a absorção, ser distribuído pela corrente sanguínea e metabolizado após a sua ação farmacológica em derivados solúveis em água para, assim, ser facilmente eliminados do organismo com uma cinética apropriada. Em alguns casos, o fármaco pode ser metabolizado em intermediários altamente reativos que podem induzir a mecanismos de toxidez. Uma eliminação lenta deste fármaco também pode produzir um acúmulo progressivo de efeitos tóxicos.⁸⁶ Estes processos de ADME estão relacionados com a biodisponibilidade dos fármacos, sendo expressada como a porcentagem do composto que alcança a corrente sanguínea em um determinado tempo.⁶⁷

Lipinski e colaboradores realizaram análises no *World Drug Index* (WDI) em 1997 identificando propriedades determinantes entre compostos considerados com boa biodisponibilidade induzindo assim o conceito de **fármaco-similar**. Estas são substâncias químicas que apresentam características de fármacos, mas que ainda estão em fase de descoberta.^{84,85} A análise no WDI mostrou que 90% dos fármacos oralmente absorvidos exibiam até 5 doadores e 10 aceitadores de ligação de hidrogênio, massa molecular menor que 500 Da e valores de logP (relacionada a lipofilia) menor do que 5.⁸⁷ Todas estas observações conduziram a **regra dos cinco de Lipinski (Ro5)** que restringe o espaço químico-biológico a um espaço de propriedades ADME / Tox compatíveis com fármacos oralmente administrados.^{84,88}

Com o desenvolvimento da química combinatória e dos métodos de triagem baseados em HTS em meados dos anos 80, foram transpostos os desafios da descoberta de compostos líderes no desenvolvimento de fármacos.⁸⁹ No HTS são realizados ensaios bioquímicos em larga escala de maneira automatizada que possibilitam testar milhares de moléculas contra vários alvos biológicos, tendo como inconvenientes, o alto custo envolvido e a baixa taxa de obtenção de moléculas bioativas.^{80,83,90} Além disto, existe grande probabilidade de selecionar inibidores falso-positivos (compostos sem afinidade pelo alvo identificados como promissores), e falso-negativos (compostos que interagem com o alvo biológico que não foram identificados no ensaio) devido a artefatos tais como autofluorescência, reatividade e agregação coloidal que podem interferir nos ensaios bioquímicos.^{91,92}

1.4.1 Métodos virtuais para o planejamento e a otimização de inibidores

No início dos anos 90, o computador passou a ser usado como ferramenta poderosa e alternativa na descoberta de compostos líderes através dos métodos virtuais que devido ao constante desenvolvimento, tem permitido uma expressiva redução nos custos e aumento na taxa de recuperação de moléculas bioativas em relação ao HTS.⁸¹ Dentre os fatores que contribuíram para isto, destacam-se o aumento crescente de informações estruturais de macromoléculas disponíveis em bases de dados como o *Protein Data Bank* (PDB) e o avanço de métodos, *softwares* e *hardwares* para a análise destas informações. Ao contrário do HTS em que uma infinidade de compostos é testada contra o alvo biológico, na triagem virtual (VS) somente uma pequena fração de moléculas melhores ranqueadas nos estudos *in*

silico devem ser ensaiadas experimentalmente para confirmar suas afinidades.⁸⁰ Os ensaios de VS consistem na aplicação de modelos *in silico* para filtrar, selecionar e priorizar as diversas moléculas contidas em coleções virtuais (como o PubChem e o ZINC) de acordo com a estimativa de suas afinidades por um determinado alvo.⁹³

Dentro do processo de descoberta, compostos líderes podem ser planejados identificados e otimizados virtualmente a partir de estratégias que se baseiam na estrutura do ligante (LBDD, do inglês *Ligand-Based Drug Design*) ou na estrutura do receptor macromolecular (SBDD, do inglês *Structure-Based Drug Design*).^{67,94} Em termos de identificação por VS, estas estratégias são aplicadas a coleções virtuais contendo milhares de moléculas que são filtradas em sub-coleções de compostos com as propriedades de interesse. Dentre os filtros moleculares que podem ser aplicados, inclui-se a possibilidade de estabelecer similaridade química com algum inibidor já conhecido, uso de modelos farmacofóricos, parâmetros farmacocinéticos (como a Ro5 de Lipinski e a regra dos três para fragmentos) e a disponibilidade comercial ou viabilidade sintética para as estas moléculas. Estas sub-coleções podem ser avaliadas contra o alvo usando métodos tais como docagem, dinâmica molecular e a relação quantitativa entre estrutura e atividade (QSAR ou *Quantitative Structure Activity Relationship*), antes de serem ensaiadas experimentalmente. Combinações entre SBDD e LBDD, e entre VS e HTS, tem sido atualmente usadas no processo de descoberta de inibidores da Cruzaína.^{58,59,80,94-97}

As estratégias LBDD não exigem informações sobre a estrutura tridimensional (3D) do alvo estando fundamentadas na análise comparativa entre as moléculas bioativas conhecidas. Buscas por similaridade, modelo farmacofórico e por QSAR são os principais métodos usados na LBDD. A busca por similaridade se baseia na premissa de que moléculas similares devem exibir propriedades farmacológicas similares. Assim, características moleculares essenciais chamadas de “impressões digitais” são usadas para classificar compostos como similares ou diferentes de um determinado ligante bioativo. A função mais usada para medir o grau de similaridade estrutural entre duas moléculas é o coeficiente de Tanimoto que pode variar entre zero e um, onde zero refere-se a duas moléculas completamente diferentes e um, a duas moléculas iguais.^{67,98} Programas de busca por similaridade 3D como o ROCS e o EON da OpenEye *Inc.*, usam um método de superposição baseado na forma 3D e no potencial eletrostático, respectivamente, entre a molécula bioativa de referência e cada possível confômero de moléculas contidas em banco de dados.^{99,100}

A LBDD pode se basear em padrões farmacofóricos comuns entre conjuntos de moléculas bioativas. Entende-se por farmacóforo, o conjunto de características estéreo-eletrônicas essenciais para o reconhecimento molecular e consequente resposta biológica.⁶⁷ Os descritores farmacofóricos mais usados incluem aceitadores e doadores de ligação de hidrogênio, centros de interações eletrostáticas e outros. Abordagens utilizando modelos farmacofóricos têm sido extensivamente utilizadas como filtros em VS, no desenho *de novo* de ligantes e na otimização de compostos líderes.¹⁰¹ Dentre os *softwares* comercialmente disponíveis, destacam-se o HypoGen (da Accelrys *Inc.*), DISCO e GASP (Tripos *Inc.*), PHASE e MOE (Schrodinger *Inc.*), cujos algoritmos diferem-se sobretudo pela maneira como lidam com a flexibilidade e o alinhamento de ligantes.¹⁰¹ Modelos farmacofóricos trazem valiosas informações quando em presença da estrutura 3D da macromolécula alvo em complexo com ligante, ou na forma apo. No primeiro caso, são definidos os pontos de interações supramoleculares chaves entre o ligante e a macromolécula que são então incorporados nos modelos farmacofóricos para busca por novas moléculas bioativas. O programa *LigandScout* é um representante desta estratégia. Na presença apenas da estrutura apo, o software GRID pode ser utilizado para mapear o sítio catalítico em busca de regiões favoráveis para interações. Neste programa campos de interação moleculares (MIFs, do inglês *Molecular Interaction Fields*) são calculados utilizando-se diferentes sondas que mapeiam o sítio catalítico.^{99,101-103}

Uma das estratégias mais empregadas na LBDD são os métodos de QSAR que tem como objetivo, planejar moléculas com propriedades superiores em relação àquelas do conjunto de dados que deu origem ao processo de modelagem.^{67,96} Estes métodos identificam e quantificam as relações dominantes entre a estrutura química e atividade biológica de um conjunto treinamento de moléculas quimicamente relacionadas (que atuam pelo mesmo mecanismo e sítio de ligação) através do uso de descritores moleculares especializados. Métodos estatísticos como a regressão por mínimos quadrados parciais (PLS) ou a regressão linear multivariada são utilizados para produzir uma relação que descreve as variações na propriedade alvo (geralmente afinidade) em função dos descritores moleculares que podem ser propriedades moleculares tais como campos eletrostáticos e estéreo-químicos.^{67,96} Os métodos de QSAR também são muito utilizados nas fases de otimização dos compostos líderes quando a estrutura 3D do alvo é disponível.^{104,105} Nesse contexto se destacam os métodos de (i) análise comparativa de campos moleculares (CoMFA

ou *Comparative Molecular Field Analysis*) e (ii) dos índices de similaridade molecular (CoMSIA ou *Comparative Molecular Similarity Analysis*), que se baseiam na ideia de que a atividade biológica está diretamente relacionada com as interações formadas no complexo fármaco-receptor.^{106,107} Programas como o HQSAR e o CoMFA da plataforma SYBYL-X (Tripos, *Inc.*) que realizam, respectivamente, análises QSAR 2D e 3D, podem ajudar a selecionar quais compostos sintetizar dentre um grande número de possibilidades durante a otimização de um composto líder.^{59,67}

A estratégia SBDD é prioritariamente utilizada quando informações 3D detalhadas da macromolécula alvo estão disponíveis. A cristalografia de raios X é o método mais utilizado para a determinação da estrutura 3D de macromoléculas devido à riqueza de detalhes fornecida. Dependendo do tamanho da macromolécula, estruturas 3D também podem ser obtidas via Ressonância Magnética Nuclear (RMN), ou via modelagem por homologia quando uma estrutura homóloga é disponível. Tais estruturas encontram-se depositadas em bases de dados públicas tais como o PDB, ExPASy e Relibase, que podem ser acessadas livremente.^{108,109} A docagem molecular é um dos principais métodos de SBDD empregados tendo expressivas contribuições nas fases de descoberta e de otimização de compostos líderes em relação a uma série de propriedades como potência, propriedades farmacocinéticas, e outras.¹¹⁰⁻¹¹² Utilizando programas tais como o AutoDock, DOCK, FlexX, FRED, GOLD e Glide, milhões de compostos podem ser avaliados e ranqueados *in silico* de acordo com suas propensões de interagirem com o alvo.¹¹² O objetivo da docagem é prever o modo de interação do complexo bi-molecular formado entre o ligante e o receptor cujo processo pode ser dividido em duas etapas: (i) geração das poses de docagem e (ii) predição da afinidade do ligante pelo sítio ativo. Na primeira etapa deste processo, diferentes métodos de amostragem que atribuem flexibilidade à molécula do ligante são utilizados sendo estes classificados em três categorias principais: (i) Métodos Aleatórios ou Estocásticos (e.g., Método de Monte Carlo, Algoritmo Genético); (ii) Métodos Sistemáticos (e.g., Construção Incremental, Amostragem Conformacional); e (iii) Métodos de Simulação (e.g., Dinâmica Molecular e Minimização de Energia).^{59,113,114}

A segunda etapa da docagem baseia-se em funções de pontuação que avaliam e classificam as energias de interação das diferentes poses adotadas pelo ligante ao interagir com o receptor de acordo com sua complementaridade estéreo-eletrônica, sendo agrupadas em três categorias: (i) campos de força, geralmente

baseados em métodos de mecânica molecular (e.g., G-score e D-score do pacote de programas SYBYL; Amber e GOLDscore, dos programas AUTODOCK e GOLD respectivamente); (ii) empíricas, são funções baseadas no ajuste teórico de dados experimentais através do uso de equações como regressão linear (e.g., LigScore, PLP, LUDI, F-Score); (iii) funções baseadas no conhecimento, que utilizam dados estatísticos dos potenciais de pares atômicos de interação, os quais são derivados de dados proveniente de complexos cristalográficos entre proteínas e ligantes (e.g., PMF e DrugScore).^{110,115} No processo de reconhecimento molecular, a energia de ligação (ΔG) é determinada pela combinação das contribuições entálpicas e entrópicas. Esse conceito é implementado nas funções de pontuação de maneira aproximada, uma vez que o componente entálpico está razoavelmente caracterizado e descrito, entretanto, as contribuições entrópicas ainda são difíceis de descrever, pois dependem de um número significativo de fatores que ainda não são totalmente compreendidos tais como a área de superfície hidrofóbica, a dessolvatação de moléculas de água ligadas ao sítio ativo e as restrições na conformação da molécula do ligante impostas durante a ligação.¹¹⁶ Estes são alguns dos motivos pelos quais os métodos de docagem ainda apresentam relativamente baixas taxas de identificação de moléculas bioativas, apesar de suas notáveis qualidades.¹¹⁷

A simulação por dinâmica molecular (DM) é um método bem estabelecido para o entendimento da dinâmica de proteínas, que no âmbito da SBDD, possibilita simular o processo de reconhecimento molecular entre o ligante e o receptor com melhor predição das poses, trazendo informações valiosas para o processo de modelagem molecular.^{118,119} Na DM, o comportamento dinâmico de um sistema bi molecular 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 que especificam como as posições e velocidades das partículas no sistema variam com o tempo. Um dos maiores desafios da DM 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, sendo ainda mais difícil para proteínas. Softwares como o GROMACS e o AMBER se destacam nestes métodos.^{118,120}

1.4.2 Planejamento de fármacos baseado em fragmentos

O método de planejamento de fármacos baseado em fragmentos (FBDD, do inglês *Fragment Based Drug Design*) tem resolvido o problema da baixa taxa de descoberta de *hits* do HTS e do VS, permitindo ainda que estes sejam facilmente otimizados a compostos com propriedades fármaco-similares.¹²¹⁻¹²³ Fragmentos são compostos de baixa massa molecular ($MW < 250$ Da), que por serem menos complexos e funcionalizados do que compostos líderes, oferecem vantagens combinatórias na exploração do espaço químico-biológico.^{83,124} Existe uma relação inversa entre a complexidade molecular dos compostos ensaiados e a capacidade destes possuírem uma boa complementaridade com o alvo proteico.¹²⁵ Assim, uma pequena coleção de fragmentos é necessária para cobrir uma maior diversidade química, conforme esquema representado da Figura 1.12.^{126,122}

Figura 1.12 - Vantagens combinatórias da identificação de compostos líderes por métodos baseados em fragmentos em relação aos métodos tradicionais. Neste caso hipotético, seria necessário sintetizar e testar um número 10000 vezes menor de compostos para se encontrar os mesmos 100 *hits* ativos por métodos FBDD em relação ao HTS, partindo-se da mesma biblioteca de 10^4 fragmentos.

Fonte: ERLANSON, D. A., 2004, p. 402.¹²⁷

Um dos desafios encontrados pela FBDD é que os fragmentos tipicamente exibem baixa potência *in vitro* ($100 \mu\text{M}$ - mM) devido ao menor número de interações realizadas com o alvo enzimático tornando difícil sua detecção pelos ensaios bioquímicos tradicionais. Além disto, como estes compostos são testados em altas concentrações existe a possibilidade de que estes inibam a enzima de maneira promíscua e não seletiva por agregação coloidal.¹²⁸ Estes falso positivos são melhor identificados quando técnicas biofísicas de detecção mais sensíveis tais como o RMN, a Ressonância Plasmônica de Superfície (SPR, *Surface Plasmon*

Ressonance) e a cristalografia de raios X são utilizados.^{122,129-132} Um esquema representando o espectro de sensibilidade destas técnicas biofísicas em relação a ensaios enzimáticos (HCS) e calorimétricos está apresentado na Figura 1.13.

Figura 1.13 – Sensibilidade das técnicas biofísicas, bioquímicas e calorimétricas de identificação de *hits* em relação à afinidade relativa e suas massas moleculares.

Fonte: HUBBARD, R. E., 2011, p. 520.¹³³

Apesar de menos potentes, fragmentos interagem com o alvo de maneira mais eficiente exibindo alta eficiência de ligante (LE). Este parâmetro é dado pela razão entre a energia livre de Gibbs para a interação entre o ligante e uma proteína específica (ΔG) e o número de átomos diferentes do hidrogênio (HAC) conforme Eq. (1.1).¹³⁴ A LE também pode ser relacionada com a constante de inibição enzimática K_i através da substituição do termo ΔG pelo produto da constante universal dos gases R , da temperatura absoluta T e do logaritmo de K_i . Outro parâmetro usado para medir a qualidade de um fragmento é sua eficiência lipofílica (LLE_{Astex}) dada pela Eq. (1.2), onde $\log P$ é o coeficiente de partição octanol / água.^{135,136}

$$LE = -\frac{\Delta G}{HAC} = -\frac{RT \ln(K_i)}{HAC} \quad \text{Eq. (1.1)}$$

$$LLE_{Astex} = \frac{0,11 \ln(10) RT (\log P - \log IC_{50})}{HAC} \quad \text{Eq. (1.2)}$$

Os parâmetros LE e LLE_{Astex} possibilitam que ligantes de diferentes tamanhos sejam comparados e selecionados para otimização, sendo este, um processo que em geral resulta no aumento do tamanho e da lipofilia do *hit* original, afetando diretamente a biodisponibilidade do composto.¹³⁴ Para se obter um padrão ADME adequado, grupos químicos precisam ser modificados de maneira a modular lipofilia, polaridade e outros, o que comumente leva a perda de interações favoráveis e por

consequência, a potência.¹²⁴ Os *hits* provenientes da FBDD em geral apresentam massa molecular entre 120 e 250 Da e uma baixa potência (2mM a 30 μ M). Para se tornarem compostos líderes e talvez fármaco-similares atraentes, estes precisarão ter sua potência aumentada o que na maioria das vezes traz como consequência, o aumento de massa molecular. Por outro lado, *hits* advindos do HTS são maiores (250 a 600 Da) e mais potentes (μ M a nM) cuja otimização usualmente requer a atenção na manutenção da massa mediante aumento da potência. Experiências em desenvolvimento de fármacos tem demonstrado que é mais fácil aumentar as moléculas (por meio da adição de grupos funcionais) do que diminuí-las, de maneira a conseguir biodisponibilidade adequada. *Hits* e compostos líderes de alta qualidade tendem a serem potentes em relação ao seu tamanho relativo implicando em elevada LE.¹²⁴ Publicações recentes de fragmentos ativos identificados por triagens em RMN e cristalografia de raios X para um grande número de alvos, têm apontado que fragmentos obedecem, na média, a **regra dos 3** em que a massa molecular é menor do que 300 Da, o número de doadores e aceptadores de ligação de hidrogênio é menor ou igual à 3, com cLogP também 3.^{95,137,138} Esta regra tem sido utilizada na construção de bibliotecas de fragmentos que poderão “crescer” e tornar-se compostos líderes com propriedades fármaco-similares baseados na Ro5 de Lipinski.^{87,139} A cascata de etapas formalmente utilizadas para o planejamento de compostos líderes baseado em FBDD está representada na Figura 1.14.

A triagem primária de uma coleção de fragmentos é feita por métodos biofísicos tais como SPR, análise térmica diferencial (TSA, do inglês *Thermal Shift Assay*) acoplada a fluorimetria de escaneamento diferencial (DFS) ou espalhamento de luz estático diferencial (DLS), e RMN baseado na proteína ($^{15}\text{N}/^1\text{H}$ HSQC) ou no ligante (*water*-LOGSY, STD).^{121,131,132} Cada um destes métodos apresentam pontos positivos e limitações. Métodos TSA são baratos e rápidos, porém são incompatíveis com proteínas hidrofóbicas sendo obtido grande número de falsos negativos. No RMN, mudanças no sinal da proteína são observadas após a ligação do fragmento fornecendo as primeiras informações estruturais das interações existentes. No entanto, esta técnica apresenta limitações de tamanho da amostra e requer grandes quantidades de proteínas isotopicamente marcadas, sendo um experimento caro que demanda equipamento especializado. SPR fornece alta capacidade de triagem e dados cinéticos do processo de reconhecimento molecular como o k_{on} e k_{off} , porém necessita que a proteína seja imobilizada.¹²¹ Métodos virtuais como os descritos na

Seção 1.4.1 também vem sendo utilizados com sucesso em triagens primárias para a identificação de fragmentos bioativos.^{140,141} Neste caso porém, as coleções virtuais de compostos são reduzidas a sub coleções contendo fragmentos que obedecem a regra dos 3, garantindo assim que estes sejam bons candidatos a otimização.

Figura 1.14 - Fluxograma esquemático para o desenvolvimento de fármacos por FBDD. Uma coleção de fragmentos é ensaiada por métodos biofísicos que detectam *hits* que são então caracterizados e otimizados a compostos líderes.

Fonte: SCHULZ, M. N., 2009, p. 616.¹²¹

Quando fragmentos bioativos são identificados, seja pelos métodos biofísicos ou virtuais, torna-se necessário validá-los e caracterizá-los para que estes sejam evoluídos a compostos líderes com propriedades fármaco-similares. É nesta etapa, que informações estruturais do modo de interação (MOB, do inglês *Mode of Binding*) do complexo enzima-ligante se torna essencial para o crescimento e evolução deste fragmento a um inibidor com potência e propriedades otimizadas.¹³⁶ A cristalografia de raios X tem sido a melhor ferramenta para caracterização do MOB descrevendo de maneira detalhada informações como a posição, estereoquímica, conformação, estado tautomérico e de protonação do ligante, além do completo padrão de interações com o alvo biológico e o mecanismo de interação.¹³¹ Embora docagem e dinâmica molecular tentem prever o MOB, é a integração de informação estrutural proveniente da cristalografia na cascata de triagem que permite, de maneira

eficiente, validar o *hit* e otimizá-lo.¹⁴⁰ Esta otimização a compostos líderes pode ocorrer de três maneiras: (i) podem ser mesclados (*merging*) quando dois ou mais fragmentos ocupam regiões no sítio que se sobrepõem, (ii) podem ser ligados (*linking*) quando ocupam regiões diferentes do sítio, ou (iii) ainda crescidos (*growing*) para ocupar regiões vizinhas diferentes daquelas em que foram encontrado.^{139,141,142}

Alternativamente aos tradicionais ensaios primários em SPR ou RMN, tem crescido uma nova abordagem que utiliza diretamente a cristalografia de raios X para triagem primária e identificação de compostos bioativos, que é conhecido como *High Throughput Crystallography* (HTC).¹³¹ Este método consiste em realizar o *soaking* de coquetéis de fragmentos estruturalmente distintos em cristais do alvo macromolecular na sua forma apo. Assim, caso o alvo tenha afinidade por algum destes fragmentos, será possível identificá-lo em mapas de densidade eletrônica, minimizando assim o esforço experimental de se realizar a triagem da completa biblioteca de fragmentos. Este método tem sido utilizado extensivamente pelas indústrias farmacêuticas Astex, Structural GenomiX, Vertex, Abbott.^{129,130,143} Outra abordagem adotada pelo time de químicos estruturais da Johnson e Johnson tem sido a utilização de cinco fragmentos similares em cada coquetel para a triagem de uma biblioteca de cerca de 900 compostos. Embora esqueletos divergentes sejam importantes para a identificação de cada fragmento dos mapas de densidade eletrônica durante o refinamento, os químicos desta companhia argumentam que os mapas de fragmentos similares podem dar diretamente uma boa representação dos *hot spots* para interações com a proteína.¹²² A ocupação de cada fragmento no sítio serve para medir sua relativa afinidade. Frequentemente é comum em HTC encontrar fragmentos ligados em regiões diferentes do sítio catalítico, sendo esta uma boa maneira de se identificar sítios alostéricos do alvo macromolecular.^{127,130}

Apesar da riqueza de detalhes estruturais do MOB fornecido pelos ensaios cristalográficos, estes não são capazes de descrever dois aspectos essenciais do processo de reconhecimento molecular: a cinética e a termodinâmica das interações.^{144,145} Os parâmetros cinéticos IC_{50} e K_i usados como indicadores de potência são primariamente obtidos, na maioria dos projetos de P&D, por ensaios enzimáticos (HTS). Estes ensaios também fornecem o modo de ação dos ligantes que podem ser do tipo incompetitivo, não competitivo ou competitivo. O IC_{50} , definido como a concentração do composto requerida para inibir 50% da atividade enzimática, não mede a afinidade dos compostos por uma proteína alvo como a constante K_i o faz.

Além disto, como o valor de K_i independe da concentração do substrato, este se torna o mais adequado para comparar diferentes compostos. Outros parâmetros cinéticos derivados do processo de *binding* entre a enzima (E) e o ligante (L) para a formação do complexo E^*L podem ser compreendidos com base no esquema da Figura 1.15. O processo mais simples de *binding* (de primeira ordem) envolve a formação e a dissociação do complexo E^*L descrito pelas respectivas taxas k_{on} e k_{off} . No equilíbrio, ocorre a concomitante formação e dissociação do complexo E^*L sendo caracterizado pela constante aparente de primeira ordem, k_{obs} , dada pela Eq. (1.3).

$$k_{obs} = k_{on}[L] + k_{off} \quad \text{Eq. (1.3)}$$

Figura 1.15 - Esquema representando o mecanismo de *binding* de primeira ordem em termos cinéticos (box) e termodinâmicos. Diagrama de energia nos quais diferentes passos do *binding* entre o ligante L e a enzima E são representados ao longo da coordenada de reação. E e L interagem formando um complexo inicial EL . Para que o complexo final E^*L seja formado, é preciso passar pelo estado de transição $(E^*L)^\ddagger$. A diferença de energia livre de Gibbs entre o estado ligado e não-ligado é dado por ΔG . $\Delta G^\ddagger$ é a energia livre de ativação que mede a barreira energética entre EL e $(E^*L)^\ddagger$.

Fonte: Autoria própria.

Propriedades cinéticas da interação têm sido correlacionadas ao desempenho farmacológico *in vivo* de ligantes.¹⁴⁶ Assim, uma diminuição do valor de k_{off} obtida em geral pelo aumento do número ou da intensidade das interações entre E e L , implica no aumento da estabilidade do complexo proteína-ligante (E^*L). O tempo de meia vida $t_{1/2}$ dado pela Eq. (1.4), proporciona uma estimativa da estabilidade deste complexo. Compostos com lenta taxa de dissociação k_{off} apresentam um longo $t_{1/2}$ sendo proposto que o desenho de inibidores com estas propriedades seja benéfico em termos da duração do efeito farmacológico e seletividade.¹⁴⁴ A média do tempo em que um fármaco permanece ligado ao seu receptor é conhecido como tempo de residência, t_R , e é dada pela Eq. (1.5).¹⁴⁷ A constante de dissociação, K_d , mede a

força da interação entre o ligante e seu alvo. No equilíbrio, a concentração do ligante em que metade dos sítios de ligação do receptor estão ocupados é equivalente a K_d , diretamente relacionada com a diferença de energia entre estados ligados e não-ligados. A termodinâmica e cinética de ligação se relacionam pela Eq. (1.6).¹⁴⁵

$$t_{1/2} = \frac{0,693}{k_{off}} \quad \text{Eq. (1.4)}$$

$$t_R = \frac{1}{k_{off}} \quad \text{Eq. (1.5)}$$

$$K_d = \frac{k_{off}}{k_{on}} \quad \text{Eq. (1.6)}$$

A assinatura termodinâmica do processo de reconhecimento molecular entre E e L é usualmente obtida pela técnica ITC que fornece informações da entalpia (H), entropia (S), ΔG , estequiometria (n) e K_d .¹³⁶ Dois compostos poderiam se ligar a um alvo com distintos ΔH e ΔS , exibindo porém, a mesma ΔG final. Não seria possível distinguir tais compostos sem os ensaios no ITC. Estas variações similares de $T\Delta S$ e ΔH em direções opostas são chamadas de compensações entálpica-entrópica e podem ser entendidas com base na Eq. (1.7) da seguinte forma: o aumento da eficiência ou do número de contatos entre ligante e enzima, que em geral leva a efeitos entálpicos favoráveis, podem causar uma contribuição entrópica desfavorável devido a redução da mobilidade do ligante quando “ligado ao alvo”.¹⁴⁸ O *binding* do ligante também afeta a mobilidade residual da proteína podendo causar um aumento na “ordem interna” que é entropicamente desfavorável. A assinatura termodinâmica, portanto, possibilita que interações sejam otimizadas via modificações estruturais que visem reduzir a penalidade entrópica durante o *binding*.¹⁴⁹ Adicionalmente, estudos termodinâmicos em diferentes temperaturas podem estimar a variação da capacidade calorífica (ΔC_p) dado pela Eq. (1.8), cujas variações estão relacionadas a modificações em regiões de solvatações envolvidas em interações.¹⁴⁸

$$\Delta G = RT \ln K_d = \Delta H - T\Delta S \quad \text{Eq. (1.7)}$$

$$\Delta C_p = \frac{\Delta H}{\Delta T} \quad \text{Eq. (1.8)}$$

Desta forma, os métodos ITC, DSC, HTS e SPR fornecem uma gama de dados termodinâmicos e cinéticos, que quando correlacionadas com informações estruturais provenientes da cristalografia, possibilitam uma otimização racional de interações e propriedades de compostos líderes para candidatos a fármacos.^{150,151}

1.5 Inibidores da cruzaina utilizados nos ensaios cristalográficos iniciais

Dentro do contexto da necessidade urgente de novas quimioterapias para o tratamento da doença de Chagas e a possibilidade de desenvolvê-las utilizando a cruzaina como alvo terapêutico, os métodos de planejamento descritos nas Seções 1.4.1 e 1.4.2 têm sido utilizados pelo grupo de química medicinal (NEQUIMED, IQSC/USP para a identificação de dezenas de potentes inibidores não-peptídicos e reversíveis desta enzima.^{58,59} Conforme enfatizado na Seção 1.4, a caracterização estrutural do MOB entre inibidor e enzima via cristalografia de raios X, traz valiosas informações para a validação ortogonal e a evolução destes inibidores a compostos líderes com propriedades necessárias para seguirem como potenciais candidatos a fármacos. De modo a validar ortogonalmente o MOB de alguns inibidores reversíveis identificados no NEQUIMED e propor racionalmente novos compostos para serem sintetizados e testados, doze inibidores que representavam os mais interessantes de suas séries (Figuras 1.16 e 1.17) foram selecionados para ensaios de cristalização. Estes ensaios foram realizados inicialmente no laboratório de Patologia da Universidade da Califórnia São Francisco (UCSF) durante um estágio sanduíche de um ano (2012) sob a supervisão do professor Dr. James McKerrow. O professor McKerrow, atualmente reitor da Universidade da Califórnia em San Diego (UCSD) e diretor do *Center for Discovery and Innovation in Parasitic Diseases* (CDIPD), foi o pioneiro dos trabalhos envolvendo a enzima cruzaina sendo co-autor de todas as estruturas cristalográficas disponíveis para a cruzaina no PDB.

Dos inibidores reversíveis exibidos na Figura 1.16, três (o Neq0038, Neq0042 e Neq0165) foram selecionados durante o doutorado do Dr. Helton Wiggers utilizando métodos virtuais (VS) consensuais baseados nas estruturas do alvo e do ligante, sendo adquiridos comercialmente para os ensaios bioquímicos.^{58,59} Apesar da maior afinidade do Neq0038, foi o Neq0042 que mais se destacou durante os ensaios celulares com a forma amastigota do parasito tendo um IC_{50} de $10,6 \pm 0,1$ μ M contra cepas Tulahuen lacZ o que é seis vezes mais potente do que o Benzonidazol (BZ) utilizado como controle ($64,3 \pm 12,3$ μ M). Em testes contra a catepsina L, o Neq0038 também se destacou como o de maior seletividade.⁵⁸ Os inibidores Neq0171 e Neq0184 foram selecionados pelo Dr. Josmar Rocha utilizando o esqueleto base de uma série de estruturas de n-acil-hidrazidas originalmente testadas contra a enzima rodesaina do *T. brucei*, o causador da doença do sono.¹⁵²

Estes são inibidores menores que se ligam mais eficientemente a cruzaina (maiores LE) que a partir da informação do MOB, poderão ser crescidos e otimizados.

Figura 1.16 - Inibidores reversíveis da cruzaina ensaiados cristalograficamente na UCSF com seus respectivos parâmetros de caracterização K_i , pK_i ($= -\log K_i$ em molar), LE e MW.

Fonte: Autoria própria.

Os inibidores Neq0409, Neq0410 e Neq0412 foram planejados pelo Dr. Peter Kenny baseado na hipótese de que dipeptidil-nitrilas, que são conhecidos inibidores covalente reversíveis de cisteíno proteases, seriam um bom ponto de partida para a síntese de inibidores da cruzaina onde os substituintes P2 e P3 do esqueleto básico poderiam ser facilmente variados via rotas sintéticas bem estabelecidas para gerar uma SAR visando incremento de potência e seletividade.^{153,154} O Neq0409 é um conhecido inibidor das catepsinas B e L de humanos e serviu de protótipo para teste destas hipóteses e para a síntese de uma série de outras dipeptidil-nitrilas inibidoras da cruzaina.¹⁵⁴ O Neq0409 apresentou ser um potente inibidor reversível da cruzaina com pK_i de 6,3 exibindo também atividade tripanosomicida com valor de IC_{50} igual a $2,7 \pm 0,3 \mu\text{M}$ ($pIC_{50} = 5,6$) contra a forma amastigota da cepa Tulahuen lacZ, sendo mais potente que BZ usado como controle ($pIC_{50} = 4,6$) e com uma citotoxicidade comparável a do BZ que é de $500 \mu\text{M}$. O Neq0410 e o Neq0412 foram planejados e sintetizados com base na estrutura do protótipo, sondando diferentes interações no

sítio S1 e S2 da cruzaina. Estas dipeptidil-nitrilas são excelentes pontos de partida para otimização de propriedades devido ao seu pequeno tamanho e alta LE.

Um segundo conjunto de inibidores (Figura 1.17) do tipo fragmentos (MW < 300 Da), foi selecionado pela Ms. Emanuella Fonseca durante o seu mestrado no NEQUIMED utilizando métodos virtuais de triagem baseados em fragmentos (Seção 1.4.2). Estes fragmentos têm em comum, anéis heterocíclicos e exibem menores afinidades, porém maiores LE em relação aos inibidores *lead-like* convencionais. O Neq0176 foi obtido por uma simplificação molecular do Neq0042 (o mais ativo do ensaio celular) que levou a um inibidor potente com maior LE e LLE_{AT}.

Figura 1.17 - Inibidores da cruzaina identificados por FBDD que foram cristalograficamente ensaiados e seus respectivos parâmetros de caracterização K_i , pK_i , LE, LLE_{AT} e MW.

Neq0147 K_i cruz.: $910,9 \pm 52,6 \mu\text{M}$ pK_i : 3,0 LE: $0,27 \text{ kcal mol}^{-1} \text{ atom}^{-1}$ LLE _{AT} : $0,31 \text{ kcal mol}^{-1} \text{ atom}^{-1}$ MW: 238 g mol^{-1}	Neq0176 K_i cruz.: $68,5 \pm 3,0 \mu\text{M}$ pK_i : 4,2 LE: $0,33 \text{ kcal mol}^{-1} \text{ atom}^{-1}$ LLE _{AT} : $0,75 \text{ kcal mol}^{-1} \text{ atom}^{-1}$ MW: 287 g mol^{-1}	Neq0220 K_i cruz.: $1250,0 \pm 88,0 \mu\text{M}$ pK_i : 2,9 LE: $0,24 \text{ kcal mol}^{-1} \text{ atom}^{-1}$ LLE _{AT} : $0,48 \text{ kcal mol}^{-1} \text{ atom}^{-1}$ MW: 234 g mol^{-1}	Neq0221 K_i cruz.: $607,4 \pm 45,7 \mu\text{M}$ pK_i : 3,2 LE: $0,28 \text{ kcal mol}^{-1} \text{ atom}^{-1}$ LLE _{AT} : $0,16 \text{ kcal mol}^{-1} \text{ atom}^{-1}$ MW: 209 g mol^{-1}

Fonte: Autoria própria.

Com o avançar do projeto e surgimento dos primeiros resultados, outros inibidores foram também propostos e ensaiados via cristalografia de raios X no NEQUIMED conforme detalhado no Capítulo 5.

1.6 Referências bibliográficas

- 1 Coura, J. R.; Vinas, P. A. Chagas disease: a new worldwide challenge. **Nature**, v. 465, n. 7301, p. S6-S7, 2010.
- 2 Rassi, A.; Neto, J. A. Chagas Disease. **The Lancet Journal**, v. 375, p. 1388–1402, 2010.
- 3 Clayton, J. Chagas disease 101. **Nature**, v. 465, n. 7301, p. S4-S5, 2010.
- 4 Bern, C. Antitrypanosomal Therapy for Chronic Chagas' Disease. **N Eng J Med**, v. 364, n. 26, p. 2527-2534, 2011.
- 5 Muñoz, S. S.; Wallukat, G.; Schimke, I. Chronic Chagas' heart disease: a disease on its way to becoming a worldwide health problem: epidemiology, etiopathology, treatment, pathogenesis and laboratory medicine. **Heart Failure Reviews**, v. 17, n. 1, p. 45-64, 2012.
- 6 Urbina, J. A. Specific chemotherapy of Chagas disease: Relevance, current limitations and new approaches. **Acta Tropica**, v. 115, n. 1-2, p. 55-68, 2010.

- 7 Clayton, J. Chagas disease: pushing through the pipeline. **Nature**, v. 465, n. 1, p. S12–S15, 2010.
- 8 Castro, J. A.; Demecca, M. M.; Bartel, L. C. Toxic Side Effects of Drugs Used to Treat Chagas' Disease. **Human & Experimental Toxicology**, v. 25, n. 8, p. 471-479, 2006.
- 9 Masana, M.; Toranzo, E. G. D.; Castro, J. A. Reductive metabolism and activation of Benznidazole. **Biochem Pharmacol**, v. 33, p. 1041-1045, 1984.
- 10 Diaz, E. G.; Castro, J. Reactions of Nifurtimox with critical thiol-containing biomolecules. Their potential toxicological relevance. **J Appl Toxicol**, v. 24, p. 189-195, 2004.
- 11 Duschak, V. G.; An Insight on Targets and Patented Drugs for Chemotherapy of Chagas Disease. **Recent Patents on Anti-Infective Drug Discovery**, v. 2, n. 1, p. 19-51, 2007.
- 12 Chen, C. K.; Guilbert, C.; Jacobson, M. P.; Mckerrow, J. H.; Podust, L. M. Structural Characterization of CYP51 from *Trypanosoma cruzi* and *Trypanosoma brucei* Bound to the Antifungal Drugs Posaconazole and Fluconazole. **PlosNTD**, v. 4, n. 4, p. e651, 2010.
- 13 Gunatilleke, S. S.; Calvet, C. M.; Johnston, J., B.; Chen, C. K.; Erenburg, G.; Gut, J.; Engel, J. C.; Chen, S.; Arkin, M. A.; Mckerrow, J. H.; Podust, L. M. Diverse Inhibitor Chemotypes Targeting *Trypanosoma cruzi* CYP51. **PLoS NTD**, v. 6, n. 7, p. e1736, 2012.
- 14 Urbina, J. A.; Docampo, R. Specific chemotherapy of Chagas disease: controversies and advances. **TRENDS in Parasitology**, v. 19, n. 11, p. 495-501, 2003.
- 15 Souza, D. F.; Garratt, R. C.; Araújo, A. U.; Guimaraes, B. G.; Jesus, W. P.; Michels, P. M.; Hannaert, V.; Oliva, G. *Trypanosoma cruzi* glycosomal glyceraldehyde-3-phosphate dehydrogenase: structure, catalytic mechanism and targeted inhibitor design. **FEBS Letters**, v. 424, n. 1, p. 131-135, 1998.
- 16 Kim, H.; Feil, I. K.; Verlinde, C. L. M. J.; Petra, P. H.; Hol, W. G. Crystal Structure of Glycosomal Glyceraldehyde-3-Phosphate Dehydrogenase from *Leishmania mexicana*: Implications for Structure-Based Drug Design and a New Position for the Inorganic Phosphate Binding Site. **Biochemistry**, v. 34, p. 14975–14986, 1995.
- 17 Guido, R. V. C.; Oliva, G.; Montanari, C. M.; Andricopulo, A. D. Structural Basis for Selective Inhibition of Trypanosomatid Glyceraldehyde-3-Phosphate Dehydrogenase: Molecular Docking and 3D QSAR Studies. **J Chem Inf Model**, v. 48, p. 918–929, 2008.
- 18 Cheleski, J.; Freitas, R. F.; Wiggers, H. J.; Rocha, J. R.; Montanari, C. A. Expression, purification and kinetic characterization of His-tagged glyceraldehyde-3-phosphate dehydrogenase from *Trypanosoma cruzi*. **Prot Exp Pur**, v. 76, p. 190–196, 2011.
- 19 Munier-Lehmann, H.; Vidalain, P. O.; Janin, Y. L. On Dihydroorotate Dehydrogenases and Their Inhibitors and Uses. **J Med Chem**, v. 56, n. 1, p. 3148–3167, 2013.
- 20 Annoura, T.; Nara, T.; Makiuchi, T.; Hashimoto, T.; Aoki, T. The origin of dihydroorotate dehydrogenase genes of kinetoplastids, with special reference to their biological significance and adaptation to anaerobic, parasitic conditions. **J Mol Evol**, v. 60, p. 113–127, 2005.
- 21 Cheleski, J.; Wiggers, H. J.; Nonato, M. C.; Montanari, C. A. Kinetic mechanism and catalysis of *Trypanosoma cruzi* dihydroorotate dehydrogenase enzyme evaluated by isothermal titration calorimetry. **Analytical Biochemistry**, v. 399, p. 13–22, 2010.
- 22 Cazzulo, J. J. Proteinases of *Trypanosoma cruzi*: potential targets for the chemotherapy of Chagas disease. **Curr Top Med Chem**, v. 2, n. 11, p. 1261-1271, 2002.
- 23 Cazzulo, J. J.; Turk, V. The Major Cysteine Proteinase of *T. cruzi*: A Valid Target for Chemotherapy of Chagas disease. **Cur Pharm Design**, v. 7, n. 1, p. 1143-1156, 2001.

- 24 Engel, J. C.; Doyle, P. S.; Hsieh, I.; Mckerrrow, J. H. Cysteine Protease Inhibitors Cure an Experimental *Trypanosoma cruzi* Infection. **J Exp Med**, v. 188, n. 1, p. 725–734, 1998.
- 25 Doyle, P. S.; Zhou, Y. M.; Mckerrrow, J. H.; Engel, J. C. The *Trypanosoma cruzi* Protease Cruzain Mediates Immune Evasion. **PLoS Pathogens**, v. 7, n. 9, p. e1002139, 2011.
- 26 Engel, J. C.; Doyle, P. S.; Palmer, J.; Hsieh, I.; Bainton, D. F.; Mckerrrow, J. H. Cysteine protease inhibitors alter Golgi complex ultrastructure and function in *Trypanosoma cruzi*. **Journal of Cell Science**, v. 111, p. 597-606, 1998.
- 27 Eakin, A. E.; Mills, A. A.; Harth, G.; Mckerrrow, J. H.; Craik, C. S. The sequence, organization, and expression of the major cysteine protease (cruzain) from *Trypanosoma cruzi*. **J Bio Chem**, v. 267, n. 11, p. 7411-7420, 1992.
- 28 Engel, J. C.; García, C. T.; Hsieh, I.; Doyle, P. S.; Mckerrrow, J. H. Upregulation of the secretory pathway in cysteine protease inhibitor-resistant *Trypanosoma cruzi*. **J Cell Sci**, v. 113, n. 1, p. 1345-1354, 2000.
- 29 Cazzulo, J. J.; Hellman, U.; Cousos, R.; Parodi, A. J. Aminoacid and carbohydrate composition of a lysosomal cysteine proteinase from *Trypanosoma cruzi*. Absence of phosphorylated mannose residues. **Mol Biochem Parasitology**, v. 38, p. 41-48, 1999.
- 30 Hellman, U.; Wernstedt, C.; Cazzulo, J. J. Self-proteolysis of the cysteine proteinase, cruzipain, from *Trypanosoma cruzi* gives a major fragment corresponding to its carboxy-terminal domain. **Mol Biochem Parasitology**, v. 44, p. 15-22, 1991.
- 31 Aslund, L.; Henriksson, J.; Campetella, O.; Frasch, A. C. C.; Pettersson, U.; Cazzulo, J. J. The C-terminal extension of the major cysteine proteinase (cruzipain) from *Trypanosoma cruzi*. **Mol Biochem Parasitology**, v. 45, p. 345-348, 1991.
- 32 Pamer, E. G.; Davis, C. E.; Magdalene, S. Expression and Deletion Analysis of the *Trypanosoma brucei rhodesiense* Cysteine Protease in *Escherichia coli*. **Infec Immunity**, v. 59, n. 3, p. 1074-1078, 1991.
- 33 Mcgrath, M. E. The Lysosomal Cysteine Proteases. **Annu Rev Biophys Biomol Struct**, v. 28, p. 181–204, 1999.
- 34 Turk, V.; Stoka, V.; Vasiljeva, O.; Renko, M.; Sun, T.; Turk, B.; Turk, D. Cysteine cathepsins: From structure, function and regulation to new frontiers. **Biochimica and Biophysica Acta**, v. 1824, n. 1, p. 68-88, 2012.
- 35 McIntyre, G. F.; Erickson, A. H. Procathepsins L and D are membrane-bound in acidic microsomal vesicles. **J Biol Chem**, v. 266, p. 15438-15445., 1991.
- 36 Farady, C. J.; Craik, C. S. Mechanisms of Macromolecular Protease Inhibitors. **ChemBioChem**, v. 11, n. 1, p. 2341-2346, 2010.
- 37 Durrant, J. D.; Keranen, H.; Wilson, B. A.; Mccammon, J. A. Computational Identification of Uncharacterized Cruzain Binding Sites. **PLoS NTD**, v. 4, n. 1, p. e676, 2010.
- 38 Polticelli, F.; Zaini, G.; Bolli, A.; Antonini, G.; Gradoni, L.; Ascenzi, P. Probing the Cruzain S2 Recognition Subsite: A Kinetic and Binding Energy Calculation Study. **Biochemistry**, v. 44, n. 8, p. 2781-2789, 2005.
- 39 Serveau, C.; Juliano, M. A.; Scharfstein, J.; Juliano, L. Investigation of the substrate specificity of cruzipain, the major cysteine proteinase of *T. cruzi*, through the use of cystatin-derived substrates and inhibitors. **Biochem J**, v. 313, n. 1, p. 951-956, 1996.
- 40 Gillmor, S. A.; Craik, C. S.; Fletterick, R. J. Structural determinants of specificity in the cysteine protease cruzain. **Protein Science**, v. 6, n. 8, p. 1603-1611, 1997.

- 41 Mckerrow, J. H. Development of cysteine protease inhibitors as chemotherapy for parasitic diseases: insights on safety, target validation, and mechanism of action. **International Journal for Parasitology**, v. 29, n. 1, p. 833 - 837, 1999.
- 42 Roush, W. R.; Reed, B. Potent second generation vinyl sulfonamide inhibitors of the trypanosomal cysteine protease cruzain. **Bioorg Med Chem Lett**, v. 11, p. 2759-2762, 2001.
- 43 Götz, M. G.; Caffrey, C. R.; Hansell, E.; Mckerrow, J. H.; Powers, J. C. Peptidyl allyl sulfones: a new class of inhibitors for clan CA cysteine proteases. **Bioorg Med Chem Lett**, v. 12, n. 19, p. 5203-5211, 2004.
- 44 González, F. V.; Izquierdo, J.; Rodríguez, S.; Mckerrow, J. H.; Hansell, E. Dipeptidyl-a,b-epoxyesters as potent irreversible inhibitors of the cysteine proteases cruzain and rhodesain. **Bioorg Med Chem Lett**, v. 17, p. 6697-6700, 2007.
- 45 Roush, W. R.; Gwaltney, S. L.; Cheng, J.; Scheidt, K. A.; Mckerrow, J. H.; Hansell, E. Vinyl Sulfonate Esters and Vinyl Sulfonamides: Potent, Irreversible Inhibitors of Cysteine Proteases. **J Am Chem Soc**, v. 120, p. 10994-10995, 1998.
- 46 Kerr, I. D.; Lee, J. H.; Farady, C. J.; Marion, R.; Rickert, M.; Sajid, M.; Pandey, K. C.; Caffrey, C. R.; Legac, J.; Hansell, E.; Mckerrow, J. H.; Craik, C. S.; Rosenthal, P. J.; Brinen, L. S. Vinyl Sulfones as Antiparasitic Agents and a Structural Basis for Drug Design. **J Bio Chem**, v. 284, n. 38, p. 25697-25703, 2009.
- 47 Jaishankar, P.; Hansell, E.; Zhao, D. M.; Doyle, P. S.; Mckerrow, J. H.; Renslo, A. R. Potency and selectivity of P2/P3-modified inhibitors of cysteine proteases from trypanosomes. **Bioorg Med Chem Lett**, v. 18, p. 624-628, 2008.
- 48 Huang, L.; Brinen, L. S.; Ellman, J. A. Crystal structures of reversible ketone-based inhibitors of the cysteine protease cruzain. **Bioorg Med Chem**, v. 11, n. 1, p. 21-29, 2003.
- 49 Choe, Y.; Brinen, L. S.; Price, M. S. Development of alpha-ketobased inhibitors of cruzain, a cysteine protease implicated in Chagas disease. **Bioorg Med Chem**, v. 13, n. 6, p. 2141-2156, 2005.
- 50 Du, X.; Hansell, E.; Engel, J. C.; Caffrey, C. R.; Cohen, F. E.; Mckerrow, J. H. Aryl ureas represent a new class of anti-trypanosomal agents. **Chem Bio**, v. 7, p. 733-742, 2000.
- 51 Beaulieu, C.; Isabel, E.; Fortier, A.; Massé, F.; Mellon, C.; Nicoll-Griffith, D.; Lee, D.; Park, Y.; Black, W. C. Identification of potent and reversible cruzipain inhibitors for the treatment of Chagas disease. **Bioorg Med Chem Lett**, v. 20, p. 7444-7449, 2010.
- 52 Malvezzi, A.; Rezende, L.; Amaral, A. T. Pharmacophore Model of Cruzain Inhibitors. **QSAR Comb Sci**, v. 28, n. 8, p. 781 - 784, 2009.
- 53 Blau, L.; Menegon, R. F.; Trossini, G. H. G.; Molino, J. V. D.; Passerini, G. D.; Bosquesi, P. L.; Chin, C. M. Design, synthesis and biological evaluation of new aryl thiosemicarbazone as antichagasic candidates. **Eur J Med Chem**, v. 67, n., p. 142e151, 2013.
- 54 Moreira, D. R. M.; Oliveira, A. D. T.; Simone, C. A.; Ferreira, R. S.; Pereira, V. R. A.; Leite, A. C. L. Conformational restriction of aryl thiosemicarbazones produces potent and selective anti-Trypanosoma cruzi compounds which induce apoptotic parasite death. **Eur J Med Chem**, v. 75, p. 467e478, 2014.
- 55 Fujii, N.; Mallari, J. P.; Hansell, E. J.; Doyle, P. S.; Gut, J.; Rosenthal, P. J.; Mckerrow, J. H.; Guya, K. Discovery of potent thiosemicarbazone inhibitors of rhodesain and cruzain. **Bioorg Med Chem Lett**, v. 15, p. 121-123, 2005.
- 56 Chiyanzu, I.; Hansell, E.; Gut, J.; Rosenthal, P. J.; Mckerrow, J. H.; Chibalea, K. Synthesis and Evaluation of Isatins and Thiosemicarbazone Derivatives against Cruzain, Falcipain-2 and Rhodesain. **Bioorg Med Chem Lett**, v. 13, p. 3527-3530, 2003.

- 57 Du, X.; Guo, C.; Hansell, E.; Doyle, P. S.; Caffrey, C. R.; Mckerrrow, J. H.; Cohen, F. E. Synthesis and Structure-Activity Relationship Study of Potent Trypanocidal Thio Semicarbazone Inhibitors of the Trypanosomal Cysteine Protease Cruzain. **J Med Chem**, v. 45, p. 2695-2707, 2002.
- 58 Wiggers, H. J.; Rocha, J. R.; Fernandes, W. B.; Sesti-Costa, R.; Carneiro, Z. A.; Cheleski, J.; Da Silva, A. B. F.; Juliano, L.; Cezari, M. H. S.; Silva, J. S.; Mckerrrow, J. H.; Montanari, C. A. Non-peptidic Cruzain Inhibitors with Trypanocidal Activity Discovered by Virtual Screening and In Vitro Assay. **PLoS NTD**, v. 7, n. 8, p. e2370, 2013.
- 59 Wiggers, H. J.; Rocha, J. R.; Cheleski, J.; Montanari, C. A. Integration of ligand- and target-based virtual screening for the discovery of cruzain inhibitors. **Molecular Informatics**, v. 30, n. 6-7, p. 365-378, 2011.
- 60 Engel, J. C.; Doyle, P. S.; Palmer, J.; Hsieh, I.; Bainton, D. F.; Mckerrrow, J. H. Growth arrest of *T. cruzi* by cysteine protease inhibitors is accompanied by alterations in Golgi complex and ER ultrastructure. **J Cell Sci**, v. 111, p. 597-606, 1998.
- 61 Choy, J. W.; Calvet, C. M.; Doyle, P. S.; Gut, J.; Arkin, M. R.; Mckerrrow, J. M.; Podust, L. M.; Renslo, A. R. Chemical-biological characterization of a cruzain inhibitor reveals a second target and a mammalian off-target. **Beilstein J Org Chem**, v. 9, n. 1, p. 15-25, 2013.
- 62 Engel, J. C.; Doyle, P. S.; Hsieh, Y.; Mckerrrow, J. H. Cysteine Protease Inhibitors Cure an Experimental *Trypanosoma cruzi* Infection. **J Exp Med**, v. 188, p. 725-734, 1998.
- 63 Barr, S. C.; Warner, K. L.; Piscitelli, J.; Wolfe, A.; Benet, L.; Mckerrrow, J. H. A Cysteine Protease Inhibitor Protects Dogs from Cardiac Damage during Infection by *Trypanosoma cruzi*. **Antimicrobial Agents And Chemotherapy**, v. 49, n. 12, p. 5160-5161, 2005.
- 64 Jacobsen, W.; Christians, U.; Benet, L. Z. In vitro evaluation of the disposition of a novel cysteine protease inhibitor. **Drug Metab Dispos**, v. 28, n. 11, p. 1343-1351, 2000.
- 65 Lombardino, J. G.; Lowe, J. A. The role of the medicinal chemist in drug discovery - Then and now **Nature Drug Discovery**, v. 3, p. 853 - 862, 2004.
- 66 Harvey, A. L. Natural products in drug discovery. **DDT**, v. 13, n. 19/20, p. 894-901, 2008.
- 67 Montanari, C. A. Química Medicinal: métodos e fundamentos em planejamento de fármacos. 1 ed. São Carlos: Edusp, 2011. v. 1. 720 p.
- 68 Burger. Burger's Medicinal Chemistry and Drug Discovery. 6th: Wiley Interscience 2003. v. 2. 817 p.
- 69 Walters, W. P.; Green, J.; Weiss, J. R.; Murck, M. A. What Do Medicinal Chemists Actually Make? A 50-Year Retrospective. **J Med Chem**, v. 54, p. 6405-6416, 2011.
- 70 Dimasi, J. A.; Hansen, R. W.; Grabowski, H. G. The price of innovation: new estimates of drug development costs. **Journal of Health Economics**, v. 22, p. 151-185, 2003.
- 71 Herper, M. The Truly Staggering Cost Of Inventing New Drugs. Forbes Journal - Pharma & Healthcare; 2013.
- 72 Ashburn, T. T.; Thor, K. B. Drug Repositioning: Identifying And Developing New Uses For Existing Drugs. **Nature Drug Discovery**, v. 3, p. 673-683, 2004.
- 73 Ekins, S.; Williams, A. J.; Krasowski, M. D.; Freundlich, J. S. In silico repositioning of approved drugs for rare and neglected diseases. **DDT**, v. 16, n. 7/8, p. 298-310, 2011.
- 74 Stephens, T. D.; Bunde, C. J.; Fillmore, B. J. Mechanism of action in thalidomide teratogenesis. **Biochemical Pharmacology**, v. 59, n. 12, p. 1489-1499, 2000.
- 75 Shetty, K. Current role of thalidomide in HIV-positive patients with recurrent aphthous ulcerations. **Gen Dent**, v. 55, n. 6, p. 537-542, 2007.

- 76 Amato, R.; Loughnan, M. S.; Flynn, E.; Folkman, J. Thalidomide is an inhibitor of angiogenesis. **PNAS**, v. 91, p. 4082–4085, 1994.
- 77 Mehta, S. The emerging role of academia in commercializing innovation. **Nature Biotechnology**, v. 22, n. 1, p. 21-24, 2004.
- 78 Bleicher, K. H.; Bohm, H. J.; Muller, K.; Alanine, A. I. Hit and lead generation: Beyond high-throughput screening. **Nature Drug Discovery**, v. 2, p. 369 - 379, 2003.
- 79 Keserü, G. M.; Makara, G. M. The influence of lead discovery strategies on the properties of drug candidates. **Nature Drug Discovery**, v. 8, p. 203-112, 2009.
- 80 Ferreira, R. S.; Eidam, O.; Mott, B. T.; Mckerrow, J. H.; Maloney, D. J.; Irwin, J. J.; Shoichet, B. K. Complementarity Between a Docking and a High-Throughput Screen in Discovering New Cruzain Inhibitors. **J Med Chem**, v. 53, n. 13, p. 4891-4905, 2010.
- 81 Oprea, T. I.; Matter, H. Integrating virtual screening in lead discovery. **Curr Opin Chem Biol**, v. 8, n., p. 349–358, 2004.
- 82 Oprea, T. I. Chemical space navigation in lead discovery. **Curr Opin Chem Biol**, v. 6, p. 384–389 2002.
- 83 Lipinski, C.; Hopkins, A. Navigating chemical space for biology and medicine. **Nature**, v. 432, p. 855–861, 2004.
- 84 Lipinski, C. A. Lead- and drug-like compounds: The rule-of-five revolution. **DDT Technologies**, v. 1, p. 337, 2004.
- 85 Leeson, P. D.; Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. **Nature Drug Discovery**, v. 6, p. 881-890, 2007.
- 86 Muster, W.; Breidenbach, A.; Fischer, H.; Kirchner, S. Computational toxicology in drug development. **DDT**, v. 13, n. 7/8, p. 303-310, 2008.
- 87 Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings **Adv Drug Deliv Rev**, v. 46, p. 3–26, 2001.
- 88 Yu, H.; Adedoyin, A. ADME–Tox in drug discovery: integration of experimental and computational technologies. **DDT**, v. 8, n. 18, p. 852-861, 2003.
- 89 Kolb, H. C.; Sharpless, K. B. The growing impact of click chemistry on drug discovery. **DDT**, v. 8, n. 24, p. 1128-1137, 2003.
- 90 Stockwell, B. R. Exploring biology with small organic molecules. **Nature**, v. 432, p. 846–854, 2004.
- 91 Ferreira, R. S.; Bryant, C.; Ang, K. K. H.; Mckerrow, J. H.; Shoichet, B. K.; Renslo, A. R. Divergent modes of enzyme inhibition in a homologous structure-activity series. **J Med Chem**, v. 52, n. 16, p. 5005-5008, 2009.
- 92 Jadhav, A.; Ferreira, R. S.; Klumpp, C.; Thomas, C. J.; Maloney, D. J.; Shoichet, B. K.; Simeonov, A. Quantitative Analyses of Aggregation, Autofluorescence, and Reactivity Artifacts in a Screen for Inhibitors of a Thiol Protease. **J Med Chem**, v. 53, p. 37–51, 2010.
- 93 Guido, R. V. C.; Oliva, G.; Andricopulo, A. D. Virtual Screening and Its Integration with Modern Drug Design Technologies. **Curr Med Chem**, v. 15, p. 37-46, 2008.
- 94 Kumar, A.; Zhang, K. Y. J. Hierarchical virtual screening approaches in small molecule drug discovery. **Methods**, v. 71, p. 26-37, 2015.
- 95 Jhoti, H.; Williams, G.; Rees, D. C.; Murray, C. W. The 'rule of three' for fragment-based drug discovery: where are we now? **Nature Drug Discovery**, v. 12, n. 1, p. 644, 2013.

- 96 Freitas, R. F.; Oprea, T. I.; Montanari, C. A. 2D QSAR and similarity studies on cruzain inhibitors aimed at improving selectivity over cathepsin L. **Bio Med Chem**, v. 16, n. 2, p. 838-853, 2008.
- 97 Mcinnes, C. Virtual screening strategies in drug discovery. **Cur Opin Chem Bio**, v. 11, p. 494–502, 2007.
- 98 Sheridan, R. P. Chemical similarity searches: When is complexity justified? **Expert Opinion on Drug Discovery**, v. 2, p. 423, 2007.
- 99 Kortagere, S.; Krasowski, M. D.; Ekins, S. The importance of discerning shape in molecular pharmacology. **Cell Press**, v. 30, n. 3, p. 138-147, 2009.
- 100 Willett, P. Similarity-based approaches to virtual screening. **Biochem Soc Trans**, v. 31, p. 603–606, 2003.
- 101 Yang, S. Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. **DDT**, v. 15 n. 11 / 12, p. 444-450, 2010.
- 102 Wolber, G. LigandScout: 3-D pharmacophores derived from protein bound ligands and their use as virtual screening filters. **J Chem Inf Model**, v. 45, p. 160–169, 2005.
- 103 Rocha, J. R.; Freitas, R.; Montanari, C. A. The GRID/CPCA approach in drug discovery. **Expert Opin Drug Discov**, v. 5, n. 4, p. 333-346, 2010.
- 104 Moda, T. L.; Montanari, C. A.; Andricopulo, A. D. Hologram QSAR model for the prediction of human oral bioavailability. **Bio Med Chem**, v. 15, p. 7738–7745, 2007.
- 105 Montanari, M. L. C.; Beezer, A. E.; Montanari, C. A.; Veloso, D. P. QSAR Based on Biological Microcalorimetry. **J Med Chem**, v. 43, p. 3448-3452, 2000.
- 106 Mendez-Lucio, O.; Castillo, R. 3D-QSAR studies on purine-carbonitriles as cruzain inhibitors: comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). **Med Chem Commun**, v. 2, n. 1058-1065, p., 2011.
- 107 Cramer, R. D. Comparative molecular field analysis (CoMFA).1. Effect of shape on binding of steroids to carrier proteins. **J Am Chem Soc**, v. 110, p. 5959–5967, 1998.
- 108 Berman, H. M.; Bhat, T. N.; Gilliland, G.; Weissig, H.; Westbrook, J. The Protein Data Bank and the challenge of structural genomics **Nat Struct Biol**, v. 7, p. 957-959, 2000.
- 109 Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. **Nucleic Acids Research**, v. 31, p. 3784-3788, 2003.
- 110 Shoichet, B. K.; MCGovern, S. L.; Wei, B.; Irwin, J. J. Lead discovery using molecular docking. **Cur Opin Chem Bio**, v. 6, p. 439–446, 2002.
- 111 Abagyan, R.; Totrov, M. High-throughput docking for lead generation. **Cur Opin Chem Bio**, v. 5, p. 375–382, 2001.
- 112 Kolb, P.; Ferreira, R. S.; Irwin, J. J.; Shoichet, B. K. Docking and chemoinformatic screens for new ligands and targets. **Cur Opin Biotechnology**, v. 20, p. 429–436, 2009.
- 113 Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. T. Improved Protein–Ligand Docking Using GOLD. **Proteins**, v. 52, p. 609–623, 2003.
- 114 Kitchen, D. B.; Furr, J. R.; Bajorath, J. Docking And Scoring In Virtual Screening For Drug Discovery: Methods And Applications. **Nature Drug Discovery**, v. 3, p. 935-949, 2004.
- 115 Mobley, D. L.; Graves, A. P.; McCreynolds, A. C.; Shoichet, B. Predicting Absolute Ligand Binding Free Energies to a Simple Model Site. **J Mol Biol**, v. 371, p. 1118–1134, 2007.

- 116 Mobley, D. L.; Dill, K. A. Binding of Small-Molecule Ligands to Proteins: "What You See" Is Not Always "What You Get". **Structure**, v. 17, p. 489-498, 2009.
- 117 Sams, D. F. Research & market strategy: how choice of drug discovery approach can affect market position. **DDT**, v. 12, n. 7/8, p. 314-318, 2007.
- 118 Salsbury, F. R. Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. **Cur Opin Pharmacology**, v. 10, p. 738-744, 2010.
- 119 Karplus, M.; Kuriyan, J. Molecular dynamics and protein function. **PNAS**, v. 102, n. 19, p. 6679-6685, 2005.
- 120 Jacob D Durrant, J. A. M. Molecular dynamics simulations and drug discovery. **BMC Biology**, v. 9, p. 71, 2011.
- 121 Schulz, M. N.; Hubbard, R. E. Recent progress in fragment-based lead discovery. **Current Opinion in Pharmacology**, v. 9, p. 615-621, 2009.
- 122 Hajduk, P. J.; Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. **Nature Drug Discovery**, v. 6, n. 1, p. 211-219, 2007.
- 123 Sun, C.; Petros, A. M.; Hajduk, P. J. Fragment-based lead discovery: challenges and opportunities. **J Comput Aided Mol Des**, v. 25, p. 607-610, 2011.
- 124 Carr, R. A. E.; Congreve, M.; Murray, C. W.; Rees, D. C. Fragment-based lead discovery: leads by design. **Drug Discovery Today**, v. 10, n. 14, p. 987-992, 2005.
- 125 Leach, A. R.; Hann, M. M. Molecular complexity and fragment-based drug discovery: ten years on. **Cur Opin Chem Bio**, v. 15, p. 489-496, 2011.
- 126 Hubbard, R. Fragment approaches in structure-based drug discovery. **Journal of Synchrotron Radiation**, v. 15, n. 3, p. 227-230, 2008.
- 127 Erlanson, D. A.; Hansen, S. K. Making drugs on proteins: site-directed ligand discovery for fragment-based lead assembly. **Cur Opin Chem Bio**, v. 8, p. 399-406, 2004.
- 128 Seidler, J. Identification and prediction of promiscuous aggregating inhibitors among known drugs. **J Med Chem**, v. 46, n., p. 4477-4486, 2003.
- 129 Blundell, T. L.; Patel, S. High-throughput X-ray crystallography for drug discovery. **Current Opinion in Pharmacology**, v. 4, n. 1, p. 490-496, 2004.
- 130 Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Tickle, I. J.; Jhoti, H. Fragment-Based Lead Discovery Using X-ray Crystallography. **JMedChem**, v. 48, n. 2, p. 403-413, 2005.
- 131 Jhoti, H.; Cleasby, A.; Verdonk, M.; Williams, G. Fragment-based screening using X-ray crystallography and NMR spectroscopy. **Cur Opin Chem Bio**, v. 11, n. 1, p. 485-493, 2007.
- 132 Perspicace, S.; Banner, B.; Benz, J.; Huber, W. Fragment-Based Screening Using Surface Plasmon Resonance Technology. **J Biomol Screen**, v. 14, p. 337-349, 2009.
- 133 Hubbard, R. E.; Murray, J. B. Experiences in fragment-based lead discovery. **Methods Enzymol**, v. 493, p. 509-531, 2011.
- 134 Abad-Zapatero, C.; Metz, J. T. Ligand efficiency indices as guideposts for drug discovery. **DDT**, v. 10, p. 464-469, 2005.
- 135 Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: A useful metric for lead selection. **DDT**, v. 9, n. 10, p. 430-431, 2004.
- 136 Ferenczy, G. G.; Keseru, G. M. How are Fragments optimized? A retrospective analysis of 145 fragments optimizations. **J Med Chem**, v. 56, n. 1, p. 2478 - 2486, 2013.

- 137 Congreve, M. A 'rule of three' for fragment-based lead discovery? **DDT**, v. 8, p. 876–877, 2003.
- 138 Baurin, N. Design and characterization of libraries of molecular fragments for use in NMR screening against protein targets. **J Chem Inf Comput Sci**, v. 44, p. 2157–2166, 2004.
- 139 Hajduk, P. J. Fragment-Based Drug Design: How Big Is Too Big? **J Med Chem**, v. 49, p. 6972-6976, 2006.
- 140 Eidam, O.; Dalmasso, G.; Barelier, S.; Shoichet, B. K.; Prati, F. Fragment-guided design of subnanomolar β -lactamase inhibitors active in vivo. **PNAS**, v. 11, p. 1-6, 2012.
- 141 Hudson, S. A.; Mclean, K. J.; Surade, S.; Ciulli, A.; Munro, A. W.; Abell, C. Application of Fragment Screening and Merging to the Discovery of Inhibitors of the Mycobacterium tuberculosis Cytochrome P450 CYP121. **Angew Chem Int Ed**, v. 51, p. 9311 –9316, 2012.
- 142 Kloe, G. E.; Bailey, D.; Leurs, R.; Esch, I. J. P. Transforming fragments into candidates: small becomes big in medicinal chemistry. **DDT**, v. 14, n. 13, p. 630-647, 2009.
- 143 Patel, D.; Bauman, J. D.; Arnold, E. Advantages of crystallographic fragment screening: Functional and mechanistic insights from a powerful platform for efficient drug discovery. **Progress in Biophysics and Molecular Biology**, v. 116, p. 92-100, 2014.
- 144 Pan, A. C.; Borhani, D. W.; Dror, R. O.; Shaw, D. E. Molecular determinants of drug–receptor binding kinetics. **Drug Discovery Today**, v. 18, n. 13/14, p. 667-673, 2013.
- 145 Martin, S. F.; Clements, J. H. Correlating Structure and Energetics in Protein-Ligand Interactions: Paradigms and Paradoxes. **Annu Rev Biochem**, v. 82, p. 267–293, 2013.
- 146 Chene, P. Can biochemistry drive drug discovery beyond simple potency measurements? **Drug Discovery Today**, v. 17, n. 7 / 8, p. 388-395, 2012.
- 147 Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug–target residence time and its implications for lead optimization. **Nature**, v. 5, p. 730-739, 2006.
- 148 Ferenczy, G. G.; Keseru, G. M. Thermodynamics of Fragment Binding. **J Chem Inf Model**, v. 52, n., p. 1039–1045, 2012.
- 149 Ferenczy, G. G.; Keseru, G. M. Thermodynamics guided lead discovery and optimization. **Drug Discovery Today**, v. 15, n. 21 / 22, p. 919 - 932, 2010.
- 150 Drinkwater, N.; Lovell, K. M.; Criscione, K. R.; Collins, B. M.; Mcleish, M. J.; Grunewald, G. L.; Martin, J. L. Fragment-based screening by X-ray crystallography, MS and ITC to identify PNMT inhibitors. **Biochem J**, v. 431, p. 51–61, 2010.
- 151 Larsson, A.; Berg, A. A. Efficiency of hit generation and structural characterization in fragment-based ligand discovery. **Cur Opin Chem Bio**, v. 15, p. 482–488, 2011.
- 152 Caffrey, C. R.; Brush, M.; Hansell, E.; Cohen, F. E.; Mckerrow, J. H. Screening of acyl hydrazide proteinase inhibitors for antiparasitic activity against *Trypanosoma brucei*. **International Journal of Antimicrobial Agents**, v. 19, n. 1, p. 227-231, 2002.
- 153 Löser, R.; Schilling, K.; Dimmig, E.; Guetschow, M. Interaction of papain-like cysteine proteases with dipeptide-derived nitriles. **J Med Chem**, v. 48, p. 7688-7707, 2005.
- 154 Bethel, A. N.; Coulson, M. D.; Kenny, P. W.; Simader, H.; Steinbacher, S. Dipeptidyl nitrile inhibitors of Cathepsin L. **BioorgMedChemLett**, v. 19, n. 1, p. 4280-4283, 2009.

Capítulo 2

Objetivos

Os principais objetivos deste trabalho são apresentados neste capítulo.

2 OBJETIVOS

Este trabalho de doutorado teve como principal objetivo a determinação e a validação ortogonal do MOB, por cristalografia de raios X, de importantes inibidores da cruzaina selecionados e desenvolvidos no grupo NEQUIMED (Figuras 1.16 e 1.17). A descrição completa do MOB de alguns inibidores da cruzaina representa valiosa informação estrutural para o cumprimento de um dos principais objetivos do grupo NEQUIMED que é a otimização da potência e seletividade de inibidores da cruzaina, sendo esta uma etapa importante para o desenvolvimento de moléculas bioativas que possam avançar na gênese planejada de fármacos contra a doença de Chagas. Os objetivos específicos deste trabalho foram:

- ✓ Realização de ensaios de cristalização e estudos cristalográficos para os complexos entre a enzima cruzaina de *T. cruzi* e os inibidores de interesse;
- ✓ Proposição do mecanismo de inibição dos inibidores e caracterização das interações intemoleculares envolvidas na formação do complexo enzima-inibidor, com base nas estruturas cristalográficas;
- ✓ Mapeamento e investigação de aspectos estruturais fundamentais do sítio catalítico da cruzaina para incremento de potência e seletividade dos inibidores, utilizando informações estruturais disponíveis no PDB e ferramentas *in silico* para a visualização e análise estrutural;
- ✓ Planejamento de novos inibidores da cruzaina;
- ✓ Desenvolvimento e otimização de estratégias para obtenção de grande quantidade de estruturas cristalográficas de qualidade para os futuros inibidores de interesse;

Capítulo 3

Expressão, Purificação e Cristalização da Cruzaína Nativa e das Mutantes C25S e C25A para os Ensaios Cristalográficos

Este capítulo aborda conceitos da biologia molecular e da cristalização de macromoléculas essenciais para a obtenção de cristais de qualidade da amostra protéica. Os métodos e resultados da expressão, purificação e ensaios de cristalização para a cruzaína nativa são apresentados na sequência, seguidos do desenvolvimento e utilização de duas formas mutantes inativas da cruzaína que foram imprescindíveis para a cristalização com alguns dos inibidores reversíveis de interesse.

3 EXPRESSÃO, PURIFICAÇÃO E CRISTALIZAÇÃO DA CRUZAÍNA NATIVA E DAS MUTANTES C25A E C25S PARA ENSAIOS CRISTALOGRAFICOS

3.1 Da obtenção da proteína pura até os ensaios de cristalização

3.1.1 Métodos em biologia molecular para expressão e purificação de proteínas

O sucesso do experimento cristalográfico é dependente da qualidade dos cristais da amostra, cujo pré-requisito crítico para a cristalização é a pureza e a solubilidade da proteína. A obtenção de quantidades suficientes destas substâncias é outro problema pertinente em bioquímica. Com o desenvolvimento da tecnologia do DNA recombinante, uma proteína de interesse pode hoje ser expressa em abundância em um organismo hospedeiro e facilmente purificada devido à inserção de caudas de afinidade em proteínas de fusão.^{1,2} Outras vantagens incluem a possibilidade de se produzir apenas domínios isolados de interesse, formas mutantes inativas ou projetadas para cristalizarem mais facilmente, ou ainda marcar aminoácidos na enzima com selênio metionina (SeMet) para a elucidação de estruturas inéditas.²⁻⁴ A principal idéia da clonagem molecular é a inserção do segmento de DNA contendo um gene de interesse em um sistema replicativo de DNA autônomo ou vetor de clonagem. Estes podem ser plasmídeos geneticamente modificados, cosmídeos, cromossomos artificiais de leveduras ou bacterianos.^{5,6}

Plasmídeos são usados para clonagem de pequenos fragmentos de DNA (25 kpb) e possuem uma origem de replicação que permite a replicação autônoma do vetor de clonagem independente do cromossomo hospedeiro. Plasmídeos também possuem um gene marcador de seletividade que confere resistência a antibióticos, permitindo a identificação de células positivamente transformadas. Exibem sítios de clonagem com pontos de reconhecimento para diversas enzimas de restrição permitindo a inserção de variadas sequencias de genes em um mesmo vetor de clonagem. Estas enzimas de restrição são utilizadas para clivar o DNA de interesse em sequencias específicas que serão ligadas ao correspondente sítio de clonagem do vetor. A fusão do DNA recombinante deve ser inserida em uma região que está sobre controle de um promotor particular necessário para a expressão do gene de interesse no hospedeiro escolhido. Desta forma, a expressão do gene pode ser controlada pela indução do promotor e, portanto, a proteína será produzida somente quando requerida.^{5,6} O plasmídeo bacteriano clássico *pET-21a* por exemplo, possui

o gene *lacI* que codifica a proteína repressora *lac*, um promotor que é específico somente para a T7 RNA polimerase, e um operon *lac* que modula a transcrição do gene de interesse conforme representado na Figura 3.1. A proteína repressora *lac* é um tetrâmero que se liga ao DNA especificamente na região do operon *lac*, impedindo a transcrição do DNA realizada pela T7 RNA polimerase, a partir do promotor. Quando a alolactose ou uma molécula similar como o isopropil β-D-1-tiogalactopiranosídeo (IPTG) estão presentes, estas interagem com a enzima repressora *lac* induzindo uma mudança conformacional que diminui a afinidade pelo operon *lac* ativando, portanto, o mecanismo de transcrição realizado pela T7 RNA polimerase que rapidamente transcreve o gene de interesse (Figura 3.1).^{5,6}

Figura 3.1 - Representação do sistema de expressão de proteína heteróloga em plasmídeo bacteriano clássico *pET-21a* induzido por IPTG.

Fonte: Adaptado de STUDIER, W., 2005.⁷

Este mecanismo de regulação gênica via operon *lac* foi um dos primeiros deste tipo elucidado em procariontes. Em seu ambiente natural, o operon *lac* (que possui três genes estruturais *lacZ*, *lacY* e *lacA*) permite às bactérias entéricas como a *E. coli* realizar um catabolismo eficiente da lactose sendo este regulado por vários fatores tais como a disponibilidade de glicose e lactose no meio extracelular. Desta forma, a *E. coli* pode utilizar lactose como fonte de energia ao produzir a enzima β-galactosidase, codificada pelo gene *lacZ*, que digere lactose em glicose e galactose. Além desta enzima, a permease codificada pelo gene *lacY* permite a captação de lactose do meio extracelular. O operon *lac* utiliza um mecanismo de controle duplo para assegurar que a célula produza as enzimas somente quando haja lactose no meio. Isso é conseguido com o repressor *lac* que paralisa a produção de enzimas na

ausência de lactose, e com a proteína ativadora catabólica (CRP), que auxilia na produção no caso de ausência de glicose no meio. Esse controle duplo faz com que a glicose seja metabolizada pelas bactérias antes que a lactose.⁵⁻⁸

Um método atualmente empregado na expressão de grandes quantidades de proteínas de fusão solúvel e na marcação com isótopos raros (^{15}N , ^{13}C e ^2H) para estudos de RMN é o da **autoindução mínima**, que utiliza toda esta maquinaria de regulação gênica naturalmente encontrada na *E. coli*.⁹ Um estudo completo identificando os componentes necessários para a autoindução gênica e a expressão de proteínas heterólogas quando as células de *E. coli* atingissem a saturação, foi realizado por William Studier em 2005. Lactose atua como o agente indutor neste meio, que utiliza como fonte de energia primária a glicose e como fonte de energia secundária o glicerol. A presença de glicose no meio previne a indução por lactose. Quando a glicose é completamente utilizada durante o crescimento bacteriano, a lactose disponível no meio passa a ser convertida em alolactose (um dissacarídeo análogo ao IPTG) que libera o repressor *lac*, e induz a expressão enzimática. Assim, a autoindução pode ser regulada pelo ajuste dos níveis de glicose / lactose (Figura 3.2).^{7,10} O principal benefício sobre a indução por IPTG é a quantidade de proteína solúvel que é produzida, uma vez que na autoindução a expressão da proteína heteróloga ocorre de maneira lenta e gradual levando aos grandes níveis de expressão, evitando assim a formação de corpos de inclusão. Em geral, a densidade optica OD₆₀₀ é até três vezes maior na autoindução do que na induzida por IPTG.¹¹

Figura 3.2 - Representação esquemática do mecanismo de autoindução mínima.

Fonte: Adaptado de STUDIER, W., 2005.⁷

A bactéria *E. coli* e a levedura *Saccharomyces cerevisiae* são os hospedeiros procarioto e eucarioto respectivamente mais estudados e compreendidos para a expressão de proteínas heterólogas. A maquinaria celular da *S. cerevisiae* é mais complexa e elaborada, sendo adequada para a expressão de proteínas codificadas por genes provenientes de organismos eucariotos. A clonagem de plasmídeos em hospedeiros como a *E. coli* e a *S. cerevisiae* levam a superprodução da proteína de interesse.⁶ Contudo é comum que proteínas expressas em *E. coli* formem agregados protéicos insolúveis, designados corpos de inclusão, que dificultam sua purificação.⁷ Uma possível solução para o problema é realizar a expressão como uma proteína de fusão, que em geral são mais estáveis que as proteínas nativas em *E. coli*, e são facilmente purificadas por cromatografia de afinidade. As fusões mais comuns são a Glutathione S Transferase (GST) e o segmento poli-histidínico ou cauda de histidina.^{6,12} A proteína de fusão com cauda GST, pode ser facilmente purificada em coluna contendo glutathione. Grande destaque tem sido dado às fusões de enzima contendo cauda de histidina, que possibilita eficiente purificação por Cromatografia de Afinidade por Metal Imobilizado (IMAC). Este método explora a capacidade do aminoácido histidina se ligar como um quelante ao metal níquel (Ni^{2+}).

Caudas de purificação são regiões terminais muito flexíveis sendo em geral, removidas da enzima pura para fins cristalográficos.^{6,13,14} Em geral, métodos cromatográficos adicionais, tais como gel filtração e troca iônica devem ser utilizados para promover a pureza necessária. Mesmo com todos os avanços tecnológicos da cromatografia, a purificação de complexos proteicos instáveis, proteínas expressas como agregados insolúveis e proteínas de membrana ainda são desafiadores. Técnicas biofísicas e químicas tais como eletroforese em gel de poliacrilamida (SDS-PAGE), *Dynamic Light Scattering* (DLS), dicroísmo circular e espectroscopia de massas podem ser utilizadas para assegurar a pureza, homogeneidade e o correto enovelamento proteico antes de se configurar os ensaios de cristalização.^{15,16}

3.1.2 Cristalização de proteínas

Uma vez obtida a proteína pura e homogênea, na concentração adequada, parte-se para a cristalização que é a etapa chave na cristalografia. Condições ideais para o crescimento de cristais de proteínas são difíceis de prever devido a grande variabilidade química e flexibilidade destas macromoléculas, não existindo ainda uma relação óbvia entre condições de cristalização e estrutura tridimensional de

uma família protéica.^{3,15,17} Cristais de proteínas são frágeis e muito mais sensíveis quando comparados aos monocristais de compostos orgânicos ou inorgânicos. Esta fragilidade é devida as fracas interações entre as moléculas biológicas e ao alto conteúdo de solvente no estado líquido coexistindo nestes cristais (entre 20% e 75%). Assim, cristais de proteínas devem ser mantidos em um ambiente saturado de solvente, caso contrário, a desidratação levaria à quebra destas ligações fracas e consequente destruição dos cristais. Estes cristais são menos ordenados do que cristais clássicos, parte devido aos grupos superficiais da proteína em contato com o solvente que podem apresentar grande mobilidade. Como consequência, os dados de difração não podem ser medidos na resolução normalmente encontrada em cristais de pequenas moléculas. Contudo, o alto conteúdo de solvente também tem consequências úteis, pois os canais de solvente nos cristais de proteínas permitem a difusão de pequenas moléculas, sendo esta propriedade usada no preparo de complexos binários de proteínas e ligantes por *soaking*.¹⁸ Cristais de proteína privilegiam o contato entre seus constituintes, ao contrário dos cristais de compostos orgânicos e inorgânicos que privilegiam o empacotamento cristalino, uma vez que estes são menores e exibem interações de caráter mais iônico e covalente.^{15,19}

Grandes avanços na cristalização de proteínas foram alcançados nos últimos anos devido à miniaturização e a automação dos ensaios. Outros avanços incluem o controle cinético e dinâmico da cristalização, que pode ainda ser realizada sobre condições especiais incluindo campos elétricos e magnéticos, separação das etapas de nucleação e crescimento, e novas estratégias de triagem.^{15,19-22} Apesar destes avanços, o problema da cristalização ainda continua devido à complexidade do comportamento de solubilidade de proteínas, que é influenciado por fatores tais como pureza da amostra, homogeneidade, temperatura, pH, força iônica, tempo, difusão, pressão, vibração, densidade e viscosidade, campos elétricos e magnéticos, e outros.^{19,23} A cristalização é um fenômeno de transição de fases em que cristais crescem de uma solução aquosa de proteína levada a supersaturação, um estado que é termodinamicamente instável, no qual pode se desenvolver uma fase amorfa ou cristalina quando esta retornar ao equilíbrio. Supersaturação pode ser obtida pela evaporação lenta do solvente, variação da temperatura, pH e outros parâmetros físico-químicos intrínsecos da amostra. O processo de cristalização é ilustrado por um diagrama de fases (Figura 3.3) que indica qual estado (líquido, cristalino, ou sólido amorfo) é estável sobre uma variedade de parâmetros de cristalização.^{15,19,24}

Figura 3.3 - Diagrama de fases de cristalização de proteínas. Os métodos de cristalização (i) *batch*, (ii) difusão de vapor, (iii) diálise (iv) FID, estão representados mostrando suas diferentes rotas de nucleação e zonas metaestáveis, assumindo como parâmetro variável, a concentração de precipitante. Condições iniciais são representadas pelos círculos pretos.

Fonte: Adaptado de CHAYEN, N. E., 2004, p. 579.¹⁹

Neste diagrama de fases, quatro zonas representam os distintos graus de supersaturação: uma zona de alta supersaturação, onde a proteína irá precipitar; uma zona de moderada saturação, onde a nucleação ocorre espontaneamente; uma zona meta-estável (abaixo da curva de supersaturação), onde os cristais são estáveis e podem crescer sem a ocorrência de nucleações adicionais; e uma zona de instauração, onde a proteína está totalmente solubilizada e nunca cristalizará. A curva de solubilidade delimita a zona metaestável e é definida como a região limite em que o soluto está em equilíbrio com cristais. A curva de supersaturação é uma linha que separa a região onde ocorre a precipitação ou nucleação espontânea, da região onde a solução permanece clara se deixada sem perturbações.^{19,24,25}

A cristalização ocorre em duas fases principais: a nucleação e o crescimento. A nucleação espontânea, fase onde ocorre a formação dos primeiros agregados ordenados, requer elevados níveis de supersaturação. Por outro lado, o crescimento é favorecido por um estado de menor supersaturação que possibilita o crescimento lento e ordenado dos cristais.²⁵ Este é um dos paradoxos da cristalização onde condições ótimas para a nucleação de cristais não são ideais para suportar o crescimento subsequente. Assim, os procedimentos de semeadura (*seeding*) com um material cristalino pré-formado são ferramentas poderosas para a separação da

etapa de nucleação e crescimento.²⁶ Nesta técnica, cristais previamente nucleados são usados como “gérmenes de cristalização” em gotas equilibradas para menores níveis de supersaturação. Gérmenes podem ainda ser usados como nucleantes heterogêneos para facilitar a cristalização de proteínas similares relacionadas em um processo conhecido sementeadura cruzada ou *crosseeding*. As técnicas de *feeding* e *weeding* são outras maneiras de se dividir a etapa da nucleação do crescimento. No *feeding* mais solução contendo a proteína é adicionada a gota o que diminui a supersaturação enquanto que ao mesmo tempo mais proteína se torna disponível para o crescimento do cristal. Na técnica de *weeding*, excessos de núcleos são removidos da gota diminuindo assim a concentração da proteína homogênea.^{15,26} Diversos motivos interrompem o crescimento dos cristais tais como defeitos na rede, envenenamento das faces e a remoção dos cristais do meio de cristalização.^{15,19}

Todas as técnicas de cristalização atuais têm como objetivo trazer a solução de proteína à supersaturação. Dentre as várias técnicas de cristalização existentes, a técnica de difusão de vapor ainda é a mais utilizada.^{19,25} Neste método, uma gota da solução de proteína é misturada com uma gota da solução de cristalização. A mesma solução de cristalização é pipetada para dentro do reservatório em um arranjo tal como o representado na Figura 3.4. Como a concentração da solução de cristalização na gota suspensa é menor do que no reservatório, a água difunde na forma de vapor da gota para o reservatório com o tempo, até que as concentrações da solução de cristalização da gota e do reservatório sejam iguais (mesma pressão de vapor). Desta forma, ocorre uma diminuição no volume da gota e o consequente aumento da concentração de todos os seus constituintes, levando à nucleação.²⁴

Figura 3.4 - Representação do experimento de difusão de vapor com gota suspensa, onde [ppt] é a concentração de precipitante.

Fonte: DRENTH, J., 2007.¹⁷

A primeira etapa para encontrar uma condição de cristalização inicial (*hit*) é expor a proteína a uma variedade de agentes químicos que poderão conduzir a

cristalização. Isto inclui variar a concentração da proteína e do precipitante, pH e outros. A função do precipitante (ppt) no experimento é diminuir a solubilidade da proteína, sendo esta proporcional ao aumento da concentração do polímero. Sais, solventes orgânicos, e polímeros como os polietilenoglicóis, são os precipitantes mais comuns. A mesma relação de proporcionalidade observada em polímeros, não é comum para sais, que estão sujeitos a outros efeitos além da força iônica como o *salt in* e *out* por exemplo. Desta forma, pode-se dizer que é difícil encontrar zonas metaestáveis apenas variando-se a concentração salina.^{15,22,24} Dentre as estratégias de triagem que são usadas para cobrir o gigantesco espaço químico de parâmetros de cristalização, duas se destacam: (i) matriz esparsa (ou fatorial incompleto) que mapeia um amplo setor do espaço químico de maneira aleatória usando condições de cristalização bem sucedidas disponíveis em bancos de dados como o *Biological Macromolecule Crystallization Database* (BMCD), sendo ideal para triagens iniciais; (ii) a estratégia *grid screens*, onde dois fatores de cristalização são sistematicamente avaliados ao mesmo tempo, sendo em geral utilizados para refinamento dos *hits*.^{19,24}

Tratamentos pós-cristalização tais como a crioproteção, podem ser realizados para melhorar a qualidade dos cristais. Neste tratamento, o cristal é congelado e mantido em nitrogênio líquido, o que os preserva intactos durante o transporte e armazenamento, reduzindo a mobilidade de regiões flexíveis e as chances de dano causado pela radiação ionizante durante a coleta de dados, o que é importante especialmente quando se utiliza fontes de raios X de alta intensidade como a luz síncroton. Cristais de gelo produzem um sinal muito forte e indesejável no padrão de difração. Portanto, é preciso evitar a cristalização da água interna da amostra durante a crioproteção usando soluções crioprotetoras. Etilenoglicol e glicerol são os crioprotetores mais comuns.^{15,27} Na Figura 3.5 está representado o que geralmente ocorre quando um cristal de proteína é induzido a congelamento rápido sem a presença de um crioprotetor. Cristais tendem a congelar de fora para dentro. À medida que este congelamento se processa, a água ainda líquida migra para outras regiões no canal de solvente até também ser congelada. Esta reorganização da água do canal de solvente, juntamente com a contração da proteína, pode causar o reempacotamento da rede cristalina com consequente aumento da mosaicidade e diminuição do poder de difração do cristal.^{28,29} No entanto, esta característica tem sido ultimamente explorada em um processo de desidratação controlada de cristais para aumentar a resistência e o seu poder de difração.²⁷

Figura 3.5 - Modelo do dano causado por resfriamento rápido induzido sem crioproteção que pode levar a uma reformatação do canal de solvente. Setas curvas pretas mostram o fluxo hipotético do líquido. Neste exemplo, um cristal perfeito é quebrado em três domínios menores o que levará ao baixo poder de difração e a patologias cristalográficas.

Fonte: Adaptado de ALCORN, T., 2010, p. 367.²⁸

3.2 Produção da cruzaina nativa e ensaios de cristalização

Em geral, a determinação do MOB de inibidores por cristalografia de raios X em campanhas de P&D de fármacos requer grandes quantidades da amostra proteica pura. Este requerimento tem sido alcançado para a cruzaina em virtude do enorme trabalho de biologia molecular que vem sendo realizado há alguns anos pelos grupos de pesquisa dos professores James Mckerrow e Charles Craik nos departamentos de patologia da UCSF. A cruzipaina foi isolada pela primeira vez a partir da forma epimastigota do parasita por Rangel e colaboradores em 1981.³⁰ Porém a clonagem e a expressão em *E. coli* da forma recombinante denominada cruzaina, só foi possível em 1992 por Eakin *et al.*, sendo obtida como uma proteína de fusão insolúvel que era facilmente isolada do lisado bacteriano, reenovelada e processada a cruzaina ativa.³¹ A expressão em células DH5α rendia cerca de 20 mg de proteína por litro de cultura bacteriana que após reenovelamento, purificação e ativação findava em cerca de 1,5 mg da protease ativa.³¹ O gene usado nesta clonagem havia sido extraído da forma epimastigota do *T. cruzi* (cepas Tulahuen, clone tctc2), sequenciado e depositada no GeneBank com o código M84342-1.³¹

O sequenciamento deste gene indicou um domínio C-terminal de 130 aminoácidos incomum a outros membros da família da papaína de proteases, porém similar ao que havia sido reportado na sequência de sua homóloga encontrada no parasita *T. brucei rhodesiense*, a rodesaina.^{31,32} A partir deste sequenciamento, também ficou constatada grande identidade na sequência primária do domínio

catalítico entre a cruzaina e a catepsina L de humanos (42,2%), e em relação a rodesaina (59,3%).^{31,33} O domínio catalítico da cruzaina compreendia apenas 37,9 % da fusão original da proteína (22,7 kDa / 60 kDa), cuja ativação era feita com incubação à 37°C em tampão 0,1 M acetato de sódio pH 5,5, 5 mM DTT e temperatura de 37°C, com máxima atividade enzimática observada entre 4 e 7 horas de incubação. Durante este período, ocorria o processamento auto-proteolítico que removia os primeiros 40 aminoácidos da fusão CheY, o prodomínio e o domínio C-terminal da protease conforme mapa apresentado na Figura 3.6.^{31,34}

Figura 3.6 - Diagrama da construção do plasmídeo para expressão da cruzaina em *E. coli*. O gene que codifica a proforma da cruzaina (Cys-104 até Leu342) foi inserido em um vetor pUC derivado do plasmídeo de expressão pCheY15LOX. A sequência de aminoácidos em negrito representa o sítio de clivagem para a remoção do prodomínio N-terminal e da extensão C-terminal, e em itálico o sítio de clivagem da junção CheY / protease. Setas acima da sequência indicam posições de clivagem autoproteolítica. As cadeias laterais da tríade catalítica estão enumeradas de acordo com o padrão de numerações da papaína.

Fonte: EAKIN, A. E., 1992, p. 7417.³¹

No trabalho publicado por Eakin no ano seguinte, os eventos requeridos para o processamento enzimático e obtenção da enzima ativa foram caracterizados, juntamente com as primeiras tentativas de cristalização e dados de difração.³⁴ Eles demonstraram a partir de uma variante inativa da cruzaina (C25A mutante) que a clivagem proteolítica do domínio C-terminal a partir da proenzima recombinante era independente da atividade enzimática.³⁴ Também demonstraram através de uma forma truncada no aminoácido Gly212 (cruzaina-Δc), que a extensão C-terminal tinha pouco efeito no nível de atividade da protease.³⁴ Estes dados concordavam com as observações feitas para a rodesaina no *T. brucei* que não requeria a extensão C-terminal para ser expressa no correto enovelamento e manter sua atividade.³⁵ Eakin *et al.* ainda testaram uma variante cujo prodomínio havia sido removido (cruzaina-Δp). Esta fusão de proteína, porém, se mostrou inativa e instável após o reenovelamento, precipitando da solução durante a purificação.³⁴ Este fato foi igualmente observado no sistema de expressão da catepsina L recombinante onde a

deleção do prodomínio também havia resultado na enzima inativa.³⁶ Notavelmente, ambos os sítios de clivagem para a remoção autoproteolítica do prodomínio e da extensão C-terminal, tem a mesma sequência de reconhecimento VVG (Figura 3.6).

Dos primeiros ensaios de cristalização realizados por Eakin em 1993 surgiram apenas cristais altamente geminados e birrefringentes da cruzaína em complexo com o inibidor Z-phe-ala-fluorometil-cetona em condições contendo citrato de sódio e sulfato de amônio em cerca de seis meses a temperaturas de 21°C. Cristais obtidos nestas duas condições eram isomórficos e cristalizavam no grupo espacial $P2_1$ com dimensões de cela unitária a , b , c , de 45,4, 51,0, 45,7 Å e β de 116,1°, com uma molécula na unidade assimétrica e 41% de conteúdo de solvente. Os cristais do mesmo inibidor em complexo com a cruzaína- Δc , planejada para ter uma porção C-terminal mais homogênea, exibia maior poder de difração e eram menos geminados em relação aos cristais da cruzaína contendo a extensão C-terminal. Logo, todos os estudos posteriores foram feitos nesta versão truncada cuja massa molecular exata determinada por espectroscopia de massas era de $22,700 \pm 5,6$ kDa.³⁴ Uma vez que a nucleação pareceu ser problemática, um protocolo de semeadura foi desenvolvido para produção de cristais de melhor qualidade. Com excessão das cristalizações iniciais que requeriram meses, não era possível crescerem cristais na ausência de gérmenes para semeadura. A primeira estrutura cristalográfica deste complexo foi resolvida em 1995 por McGrath e as primeiras estruturas disponíveis no PDB foram obtidas para complexos contendo inibidores análogos desta fluorometilcetona, sendo estas estruturas essenciais para a compreensão da especificidade do sítio S2.^{37,38}

Os protocolos de expressão, purificação e cristalização da cruzaína- Δc de Eakin e McGrath, foram empregados para a obtenção da maioria das estruturas atualmente disponíveis no PDB.^{34,37} No entanto, a quantidade de proteína obtida nestes protocolos que envolviam a expressão em corpos de inclusão eram muito limitadas, impossibilitando a realização de triagens de condições químicas de cristalização para os inibidores de interesse que tinham esqueletos e propriedades cada vez mais diversificados. Lee *et al.* desenvolveram em 2012 uma construção do plasmídeo da cruzaína- Δc em fusão com cauda de histidina (6xHis tag) que proporcionou a obtenção de quantidades suficientes da cruzaína pura para estudos de RMN e para as triagens de cristalização.⁹ Esta versão recombinante truncada da cruzaína em fusão com cauda de histidina, denominada daqui em diante de cruzaína nativa, foi utilizada em todos os ensaios de cristalização deste projeto.

3.2.1 Material e métodos

A cruzaina nativa para os ensaios de cristalização foi preparada a partir de uma versão modificada do protocolo empregado na obtenção das estruturas recém-publicadas por Ferreira *et al.*, usando o plasmídeo pET21a que contém o gene da procruzaína com cauda de histidina cuja sequência está exibida na Figura 3.7.^{9,39}

Figura 3.7 - Sequência protéica codificada pelo gene da procruzaína. O domínio catalítico (1 a 215) está em negrito. Os resíduos da cauda de histidina e o prodomínio N-terminal estão numerados de -112 até -1.

His ₆ + Pro-Region		Cruzain	
-112		HH	HHHHSCLVP
-100	AATASLHAE	TLTSQFAEFK	QKHGRVYESA AEEAFRLSVF RENLFLARLH
-50	AAANPHATFG	VTPFSDLTRE	EFRSRYHNGA AHFAAQERA RVPKVEVVG
1	APA	AVDWRAR	GAVTAVKDQG QCGSCWAFSA IGNVECWFL AGHPLTNLSE
51	QMLVSCDKTD	SGCSGGLMNN	AFEWIVQENN GAVYTEDSYP YASGEGISPP
101	CTTSGHTVGA	TITGHVELPQ	DEAQIAAWLA VNGPVAVAVD ASSWMTYTGG
151	VMTSCVSEQL	DHGVLLVGYN	DSAAVPYWII KNSWTTQWGE EGYIRIAKGS
201	NQCLVKEEAS	SAVVG	

Autoria: LEE, G., 2012, p. 10091.⁹

Esse plasmídeo foi transformado em células de *E. coli* competentes do tipo *Arctic Express* (DE3)RIL da Agilent Technologies, e plaqueado em meio *Luria Broth* (LB) com antibióticos carbanicilina (100 µg/mL) e gentamicina (20 µg/mL). As colônias transformadas cresceram a 37 °C por 16 horas no incubador. Estas foram inoculadas em 150 mL de meio LB contendo os mesmos antibióticos que foram levados ao shaker por 20 horas a 37°C. Este pré-inóculo foi adicionado em 1L de meio de autoindução até a mínima densidade óptica (OD₆₀₀) de 0,16, sendo levado ao shaker por 72 horas a 18 °C. Para a preparação de 1L do meio, foram adicionados 10g de N-Z amine, 5g de *yeast extract*, 5 mL de glicerol em 900 mL de água destilada. Este foi autoclavado, sendo posteriormente adicionadas soluções salinas (1M de estoque) filtradas nas seguintes quantidades: 25 mL de Na₂HPO₄, 25 mL de KH₂PO₄, 50 mL de NH₄Cl, 5 mL de Na₂SO₄ e 2 mL MgSO₄. Os antibióticos e as fontes de carbono na concentração final de 0,05% (w/v) D-glicose e 0,2% (w/v) α-Lactose foram adicionados. Após 72 horas de expressão, as culturas foram centrifugadas a 5000 rpm durante 30 minutos na temperatura de 4°C e os *pellets* contendo as células foram armazenados em -80°C até a etapa de lise celular.

Todos os reagentes utilizados na expressão e purificação foram fornecidos pela Sigma e todas as etapas da purificação foram realizadas a 4°C. Os *pellets* obtidos foram ressuspensos em 200 mL de tampão 50 mM Tris (pH 10), 300 mM

NaCl, 10 mM imidazol, 1 mM CaCl_2 , 1 mM MgSO_4 , 1 tablete do *complete protease inhibitor cocktail* (Roche). Após duas horas neste tampão, as células ressuspensas foram submetidas a quatro ciclos de cisalhamento no microfluidizador M110 da Microfluidics. O lisado foi centrifugado a 20000 rpm por 30 minutos e o sobrenadante resultante, foi incubado com 5 mL de solução de níquel (Ni-NTA agarose) para purificação em cromatografia IMAC. As frações puras da proenzima foram coletadas pela eluição com a solução 500 mM imidazol, 50 mM Tris e 300 mM NaCl no pH 10. Estas frações foram dializadas por 12 horas no tampão de ativação constituído por 100 mM NaAc, 300 mM NaCl e 5 mM EDTA no pH 5,5. Após a diálise, a solução da proteína foi diluída para $0,5 \text{ mg mL}^{-1}$ no tampão de ativação.

A auto-proteólise do prodomínio a partir do zimógeno foi iniciada com 5 mM de DTT em banho maria a 37°C em cerca de duas horas. Após este período, a solução de aspecto leitoso ficou límpida novamente, indicando a transformação do zimógeno inativo em domínio catalítico ativo e a completa proteólise do prodomínio. A etapa de ativação foi monitorada por eletroforese com gel SDS-PAGE 12% de Bis-Tris sobre condições desnaturantes para a amostra. O gel foi corado em 40% (v/v) metanol, ácido acético a 10% e 0,5% de Coomassie Blue R-250 e descorado na mesma solução sem Coomassie. Após a ativação, foi adicionado 1 mM de MMTS para prevenir a degradação da enzima. Em seguida foi realizada nova diálise a 4°C durante 12 horas no tampão 100 mM NaCl, 50 mM Tris (pH 7,5) para a estabilização da enzima. Gel filtração utilizando a coluna Superdex 200 (10/300 GL) e tampão 100 mM NaCl, 50 mM Tris (pH 7,5) foi usado como etapa final da purificação.

A cruzaína inibida com MMTS foi concentrada até 1 mg mL^{-1} em tampão 2 mM Bis Tris pH 5,8, e 5 mM de DTT foi acrescido para reverter a inibição por MMTS, seguida da adição dos inibidores de interesse. Vários lotes da cruzaína foram preparados em diferentes concentrações (entre $7,5$ e 15 mg mL^{-1}) nos tampões Tris (pH 7,5) e Bis Tris (pH 5,8) para vários experimentos de cristalização configurados. Outra estratégia de preparação incluía a inibição direta da enzima com os inibidores na etapa onde se utiliza o MMTS. Para estes ensaios, foram utilizados *kits* de matriz esparsa da Qiagen desenvolvidos com base nas condições de cristalização mais frequentes observadas pelo *Joint Center for Structural Genomics* (JCSG), JCSG I até IV, e também os kits Peg I e II, pH Clear I e II, *Anions suite*, *Cations suite* e *AmSO₄ suite*, com 96 condições químicas únicas em cada *kit*. Foi utilizado o dispositivo *Mosquito Nano liter Dropsetter* (TTP Labtech) para estes ensaios, sendo

também testados diferentes volumes da gota suspensa (150 a 300 nL) e diferentes proporções entre volume da amostra e da solução de cristalização (0,5:1 até 1:1,5).

3.2.2 Resultados e discussão

Os primeiros ensaios de cristalização da cruzaina nativa com os inibidores de interesse (Figuras 1.15 e 1.16) foram realizados durante um estágio inicial de dois meses (Agosto e Setembro de 2011) no laboratório de patologia da UCSF. Nesta ocasião, foi utilizado o plasmídeo desenvolvido por Lee *et al.* e o protocolo descrito na Seção 3.2.1 para expressão e purificação da cruzaina nativa. Utilizando o método de autoindução mínima e a fusão com cauda de histidina, Lee e colaboradores foram capazes de aumentar o rendimento global de cruzaina madura em até 40 vezes (cerca de 60 mg) em relação às preparações originais de Eakin *et al.*, cuja expressão levava a formação de corpos de inclusão.⁹ Tal como descrito na Seção 3.1.1, o meio de autoindução mínima contém os nutrientes necessários para que o crescimento celular e a expressão gênica ocorram de maneira lenta e gradual na *E. coli*.⁷ Este fator também é favorecido por menores temperaturas de expressão. Para tal, foram utilizadas células especiais competentes do tipo *Arctic express* DE3-RIL, cuja melhor performance de expressão para a cruzaina ocorria à 18 °C.⁹ Desta forma, elevados níveis de procruzaina eram obtidos na fração solúvel evitando assim, a necessidade de se reenvolver a proteína extraída dos corpos de inclusão o que diminuía a homogeneidade conformacional e dificultava a cristalização. Outro importante avanço desta construção foi à fusão com cauda de histidina que permitia a eficiente purificação da procruzaina por cromatografia IMAC em coluna de níquel. Neste caso, como a procruzaina já era expressa e purificada no enovelamento correto e como a cauda de histidina era muito menor do que a fusão CheY do plasmídeo antigo, o processo de ativação (clivagem e digestão do prodomínio N-terminal exibido no gel de eletroforese da Figura 3.8) utilizando as mesmas condições da construção antiga, era cerca de duas vezes mais veloz sendo observada máxima atividade enzimática após uma hora e meia do início da ativação.

As primeiras caixinhas de cristalização foram configuradas usando condições de cristalização disponíveis para a cruzaina no PDB e no banco de dados de cristalização de macromoléculas biológicas (BMCD). A maioria destas condições de cristalização possuía como agente precipitante o citrato de sódio e o sulfato de amônio, cujo pH variava de 5,0 à 7,0, conforme dados apresentados na Tabela 3.1.

Assim, diferentes soluções químicas para os ensaios de cristalização pelo método *grid screen* foram preparadas variando-se as concentrações dos precipitantes da Tabela 3.1 em 0,5 M para mais e para menos, e ainda a faixa de pH entre 4,5 e 8,0 em incrementos de 0,1 para 0,1 M dos tampões acetato de sódio, hepes e tris.

Figura 3.8 - Ativação da cruzaína nativa a partir da autoproteólise do prodomínio N-terminal.

Autoria: LEE, G., 2012, p. 10091.⁹

Poucos cristais surgiram nesta etapa do projeto e a grande maioria destes eram frágeis e exibiam baixo poder de difração, segundo imagens coletadas na linha de luz BL8.3.1 do síncroton ALS em Berkeley em 28 de setembro de 2011. A análise destes resultados preliminares levantou a hipótese de que a baixa qualidade dos cristais obtidos era consequência da variedade de esqueletos químicos e do tipo de inibidores reversíveis que havia o interesse em se co-cristalizar. É bem conhecido na literatura que ligantes com esqueletos químicos muito diferentes induzem mudanças conformacionais distintas no sítio catalítico durante o processo de reconhecimento bimolecular. Estas podem ser refletidas em mudanças expressivas na superfície da proteína especialmente em *loops* e alças que em uma co-cristalização, poderão levar a um diferente padrão de interações entre cadeias vizinhas e por consequência distintos empacotamentos e condições de cristalização.^{18,40-43} Segundo dados da Tabela 3.1, existia uma grande semelhança nas condições de cristalização dos complexos cruzaína-inibidores reportadas até 2008 no PDB, possivelmente devido à similaridade estrutural destes inibidores do tipo covalente irreversíveis baseados em fluorometilcetonas e vinilsulfonas peptídicas. Como estes potentes inibidores eram análogos entre si, às mudanças conformacionais induzidas na proteína seriam insignificantes o que levaria a um empacotamento cristalino idêntico (justificado pelo isomorfismo nos parâmetros de cela unitária e mesmo grupo espacial) e condições de cristalização similares. Estruturas cristalográficas depositadas a partir de 2009 foram obtidas para inibidores com estruturas químicas bem diferentes das estruturas

anteriores. Neste caso, triagens de um grande número de condições de cristalização tiveram que ser ensaiadas para se obter cristais destes complexos, que só foram possíveis devido aos avanços da biologia molecular da cruzaína realizado por Lee *et al.* e a coleta de dados em fontes de raios X mais intensas nos laboratórios de luz síncrotron que permitiram a obtenção de estruturas com maiores resoluções.

Tabela 3.1 - Estruturas cristalográficas da cruzaína disponíveis no PDB no início deste projeto mostrando o precipitante e o pH de cristalização, fonte de coleta de dados cristalográficos, mecanismo de inibição, resolução, número de moléculas na ASU e grupo espacial. Todas co-cristalizaram pelo método de difusão de vapor com gotas suspensas.

Código PDB (Ano)	Condições de cristalização (pH)	Fonte de raios X (λ em Å)	Tipo de Inibidor	Resolução (ASU)	Grupo espacial
1AIM (1997)	0,8 M citrato de sódio (5,5)*	Ânodo rotativo	C.I.	2,00 (1)	P12 ₁ 1
2AIM (1997)	0,9 M citrato de sódio (5,0)*	Ânodo rotativo	C.I.	2,20 (1)	P12 ₁ 1
1EWP (2000)	0,8 M citrato de sódio (5,5)*	Ânodo rotativo	C.I.	1,75 (1)	P12 ₁ 1
1F29 (2000)	0,8 M citrato de sódio (6,0)*	Ânodo rotativo	C.I.	2,15 (3)	P12 ₁ 1
1F2A (2000)	1,0 M citrato de sódio (6,5)*	Ânodo rotativo	C.I.	1,60 (1)	P12 ₁ 1
1F2B (2000)	1,0 M citrato de sódio (6,5)*	Ânodo rotativo	C.I.	1,80 (1)	P12 ₁ 1
1F2C (2000)	1,0 M citrato de sódio (6,0)*	Ânodo rotativo	C.I.	2,00 (1)	P12 ₁ 1
1ME3 (2002)	1,0 M citrato de sódio (6,8)*	SSRL BL9-1 (0,98) #	N.C.R.	1,20 (1)	P12 ₁ 1
1ME4 (2002)	1,0 M citrato de sódio (6,8)*	SSRL BL9-1 (0,98) #	N.C.R.	1,20 (1)	P12 ₁ 1
1EWL (2003)	0,9 M citrato de sódio (6,2)	Ânodo rotativo	C.I.	2,00 (1)	P12 ₁ 1
1EWM (2003)	0,9 M citrato de sódio (6,5)	Ânodo rotativo	C.I.	2,00 (1)	P12 ₁ 1
1EWO (2003)	0,8 M citrato de sódio (7,0)	Ânodo rotativo	C.I.	2,10 (1)	P12 ₁ 1
1U9Q (2005)	1,0 M citrato de sódio (6,0)	Ânodo rotativo	C.I.	2,30 (1)	P12 ₁ 1
2OZ2 (2008)	1,2 M sulfato de amônio (7,5)	SSRL BL9-1 (0,98) #	C.I.	1,95 (2)	C121
3HD3 (2009)	2,0 M sulfato de amônio, 2% PEG 400, 30% Etanol (7,5)	SSRL BL7-1 (1,00) #	C.I.	1,75 (1)	P12 ₁ 1
3I06 (2009)	2,0 M fosfato de amônio	SSRL BL7-1 (1,00) #	C.I.	1,10 (1)	P6 ₅ 22
3IUT (2010)	0,1 M acetato de sódio, 20 % PEG 3000 (4,5)	SSRL BL7-1 (0,97) #	C.I.	1,20 (1)	P12 ₁ 1
3KKU (2010)	2,0 M fosfato de amônio (8,5)*	ALS BL8.3.1 (1,00) #	N.C.R.	1,28 (1)	P6 ₅ 22
3LXS (2010)	1,2 M sulfato de amônio, 0,2 M sulfato de lítio (6,0)	SSRL BL7-1 (0,97) #	C.I.	1,50 (2)	C121

*Método *microseeding* utilizado, #coleta de dados em condições criogênicas (100 K), C.I.: Covalente Irreversível, N.C.R.: Não Covalente Reversível, Ânodo rotativo modelo RIGAKU RU200, SSRL: *Stanford Synchrotron Radiation Lightsource*, ALS: *Advanced Light Source*, ASU: número de moléculas na unidade assimétrica do cristal.

Fonte: Autoria própria.

Com a consolidação da bolsa de pesquisa PDSE da CAPES durante o ano de 2012, foram iniciados os ensaios de cristalização utilizando os *kits* de matriz esparsa da QIAGEN para uma ampla varredura do espaço químico em busca de condições de cristalização. Esta etapa demandou grande quantidade de proteína para as co-cristalizações e, assim, 14 lotes da proteína tiveram que ser preparados. Resultados

préveos haviam evidenciado que amostras da cruzaína quando armazenadas e posteriormente utilizadas, dificilmente cristalizavam. A cruzaína também não poderia ser armazenada na geladeira devido ao efeito auto-proteolítico que a degradava rapidamente. Logo, todos os ensaios cristalográficos deveriam ser configurados imediatamente após a purificação, dada a dificuldade em armazenar e estabilizar a enzima.¹⁸ Após configurar numerosos experimentos de cristalização da cruzaína nativa com os inibidores de interesse (Seção 1.5), foram obtidos novamente cristais com alto conteúdo de solvente e baixo poder de difração. Tais informações foram obtidas após coletas de dados realizadas na linha de luz BL8.3.1 do ALS em Berkely com o apoio do cientista da linha James Holton, cujos trabalhos publicados são referência no estudo do dano causado pela radiação ionizante em amostras cristalinas e em estratégias de coletas para evitá-los. Logo, pode-se considerar que os dados foram cuidadosamente coletados e que, portanto, outros fatores anteriores à coleta seriam os culpados pela baixa qualidade das informações obtidas.

Assim, dentre as hipóteses levantadas para o problema foi considerada como a mais relevante, o fato de que o efeito auto-proteolítico da enzima estaria de alguma forma influenciando negativamente a maneira como esta se cristalizava. Para investigar e comprovar esta hipótese, foram realizados alguns experimentos cujo resultado mais interessante está exibido no gel de eletroforese da Figura 3.9, onde amostras da cruzaína nativa coletadas dois dias após a inibição com três classes distintas de inibidores, evidenciam a degradação da enzima decorrente da proteólise especialmente para os inibidores reversíveis e fragmentos. Este gel foi preparado com a cruzaína nativa na concentração de 11 mg ml^{-1} (ou $488 \text{ } \mu\text{M}$) inibida na razão estequiométrica molar (entre proteína e inibidor) de 1:1,5 para os potentes inibidores covalentes irreversíveis K777 e WRR483, 1:3 para os reversíveis Neq0042 e Neq0171, e 1:6 para os fragmentos Neq0147 e Neq0220.

A cruzaína, devido a sua capacidade de reconhecer um número muito maior de sequências de aminoácidos em relação a outras endoproteases homólogas, ostenta de um extraordinário efeito proteolítico sendo capaz inclusive de clivar outras moléculas de si própria (auto-proteólise). Em geral, esta enzima perde metade de sua atividade catalítica em cerca de duas horas.⁴⁴⁻⁴⁶ Conforme enfatizado na Seção 3.1.2, a cristalização requer uma elevada pureza química e conformacional. Portanto, se a cruzaína não estiver totalmente inibida, haverá um efeito proteolítico residual que clivará a enzima madura pura em sequências de aminoácidos menores

diminuindo a homogeneidade e a concentração da enzima durante o processo de cristalização, que em geral se processa em algumas semanas.

Figura 3.9 - Efeito proteolítico da cruzaína em relação ao tipo de inibidores utilizados. Gel SDS Page preparado após dois dias de inibição.

Fonte: Autoria própria.

Das 19 estruturas cristalográficas da cruzaína depositadas no PDB até o início deste projeto (Tabela 3.1), 16 eram para complexos com inibidores covalentes irreversíveis. Estes se ligam a cisteína catalítica cessando o efeito proteolítico e, portanto, a enzima permanece estável durante o processo de cristalização. O mesmo não ocorre para ligantes reversíveis onde uma proteólise residual ainda remanesce. As três estruturas da cruzaína em complexos com inibidores não covalentes reversíveis disponíveis no PDB até então, só foram obtidas devido à elevada potência destes inibidores. No caso das estruturas 1ME3 e 1ME4, os inibidores do tipo hidroximetil cetonas do complexo exibiram constantes de inibição (K_i) de 64 e 146 nM respectivamente sendo portanto inibidores reversíveis de alta afinidade que interagem de maneira não covalente, conforme a maioria dos inibidores que o Nequimed estava interessado em desenvolver.⁴⁷ Além disto, estes inibidores também cristalizaram em um pH elevado (6,8) em relação a maioria das outras estruturas obtidas conforme dados da Tabela 3.1. Neste pH, a enzima perde quase metade de sua atividade proteolítica cuja faixa ideal de pH para esta atividade é entre 4,8 e 6,4 com ápice em 5,3 conforme gráfico exibido na Figura 3.10.⁴⁴

Um caso interessante é observado na estrutura de código PDB 3KKU onde o inibidor não covalente do tipo benzimidazol bromofenoxi-acetamida (código PDB B95), realiza poucas interações intermoleculares no sítio exibindo um K_i de 2 μ M.³⁹ Portanto, este inibidor possui uma afinidade pelo menos 13 vezes menor do que os inibidores das estruturas 1ME3 e 1ME4, tendo afinidade comparável aos inibidores Nequimed reversíveis (Figura 1.16) que foram ensaiados contra a cruzaína nativa.

Da análise desta estrutura, observa-se a presença de uma pequena molécula de metil metano tiosulfonato (MMTS) ligada covalentemente a cisteína catalítica com 30% de ocupação no sítio. Os outros 70% de ocupação são destinados ao ligante B95. O MMTS é geralmente utilizado nas etapas finais da purificação para evitar a degradação causada pelo efeito proteolítico da enzima, sendo posteriormente removido com DTT e incubado com o inibidor de interesse antes da co-cristalização. Possivelmente, nem todo o MMTS foi removido antes da incubação com o B95. Desta forma, a cristalização levou a formação de um cristal com as duas populações da cruzaina com topologias similares coexistentes: a primeira covalentemente ligada ao MMTS (30%) e outra em complexo com o inibidor B95 (70%). Este complexo também co-cristalizou em um pH elevado (8,5) sendo o maior registrado na Tabela 3.1, que segundo o gráfico da Figura 3.10, é uma região onde quase não se observa atividade proteolítica da cruzaina. Portanto, o pH elevado e a presença do MMTS foram os responsáveis pela cristalização deste inibidor.

Figura 3.10 - Perfil de atividade proteolítica da cruzaina em relação ao pH de operação em hidrólise monitorada para o substrato fluorogênico peptídico Abz-KLRFSKQ-EDDnp.

Fonte: JUDICE, W. A., 2001, p. 6583.⁴⁴

Este efeito proteolítico da cruzaina, também dificulta a realização de ensaios em técnicas tais como SPR, RMN, ITC, DSC. Além disto, este é um dos motivos para a não existência de estruturas cristalográficas da cruzaina na forma apo, que além de limitar a nossa compreensão do processo de reconhecimento molecular para esta enzima também impossibilita a realização de ensaios por *soaking*. Neste método, cristais da proteína apo são imersos na solução de cristalização contendo um excesso molar do ligante. Tal como descrito na Seção 3.1.1, a vantagem deste método é que bastaria uma única condição otimizada de cristalização da enzima

apo para a obtenção do MOB de diversos inibidores. Tanto no *soaking* quanto na co-cristalização, a concentração do ligante que tipicamente se usa é de 5 a 1000 vezes maior do que a constante de dissociação (K_d) entre a enzima e o inibidor.¹⁸ Esta faixa de concentração geralmente assegura que o sítio catalítico da enzima esteja povoado com moléculas do ligante, garantindo assim, sua presença e fácil descrição nos mapas de densidade eletrônica. Dependendo da potência e da solubilidade de um inibidor, este valor de concentração pode não ser alcançado. Isto porque, para inibidores menos potentes como os fragmentos, a mesma faixa de inibição da enzima é obtida usando uma quantidade maior do inibidor. Isto implica em maiores concentrações do solvente em solução que pode causar problemas na estabilidade e enovelamento proteico, como é o caso do DMSO. Este é considerado solvente universal no mundo orgânico, e é geralmente usado na solubilização dos compostos para os ensaios com a cruzaina, cuja tolerância máxima em solução é de 4%.

Analizando-se o resultado exibido na Figura 3.9 tendo como base todas estas argumentações, observa-se que a razão estequiométrica usada para a inibição com os fragmentos Neq0147 e Neq0220 correspondiam a uma quantidade de DMSO de quase 4% estando próximo ao limite tolerado pela cruzaina. Porém, a concentração dos inibidores na gota de cristalização correspondia apenas duas vezes a afinidade destes. Aliado ao fato de serem reversíveis, estes fragmentos não eliminaria o efeito proteolítico e mesmo se cristalizassem, talvez não fosse possível ver densidade eletrônica para estes inibidores no sítio. Novamente na Tabela 3.1, observa-se que várias estruturas cristalográficas foram obtidas pelo uso da técnica de *seeding*, que conforme discutido na Seção 3.2, evitavam a geminação em cristais da cruzaina. Porém, a obtenção do primeiro gérmen de cristalização caracterizou a maior dificuldade destes ensaios. Estes quando obtidos, rapidamente se solubilizavam quando em contato com a solução contendo a proteína a ser cristalizada.

Dos inibidores testados na enzima nativa, apenas os cristais obtidos por co-cristalização com o Neq0176 (o de maior LE e LLE_{AT}) e com o Neq0409 tiveram um padrão de difração adequado exibindo uma máxima resolução de 2,60 Å. A concentração final do Neq0176 na solução de co-cristalização foi de 4,2 mM o que correspondia à razão molar entre proteína e inibidor de 1:8 e 1,5% de DMSO. Os cristais foram equilibrados por difusão de vapor com gotas de 0,5:0,5 µL entre a solução da proteína inibida e a solução de cristalização composta do tampão 0,1 M Bicine pH 9,0 e 1,6 M de sulfato de amônio. O cristal do complexo com o Neq0409 foi obtido por co-cristalização com 2,5 mM do inibidor (equivalente a uma razão

molar entre proteína e inibidor de 1:5 e 3,5% de DMSO) na condição 0,1M Hepes, 1,2M K/Na tartrato e pH 7,5. Estes cristais cresceram entre seis e oito semanas à 20°C. De maneira a contornar o problema da proteólise, obter a informação do MOB para inibidores reversíveis pouco potentes como os fragmentos e tentar obter a estrutura da enzima na forma APO para experimentos de cristalização por *soaking*, novas construções mutantes da enzima foram idealizadas para serem inativas.

3.3 Desenvolvimento da cruzaina mutante C25S e C25A

Devido aos desafios encontrados na cristalização de inibidores reversíveis com a cruzaina nativa, partiu-se para novas construções da enzima com mutação sítio dirigida. As mutantes C25A e C25S foram projetadas para serem inativas mantendo, porém, as mesmas características do sítio catalítico da cruzaina nativa. Construções com enzimas mutantes e formas inativas já foram estudadas em outras proteases como a cathepsina S e L.⁴⁸ A mutação, e os protocolos de expressão e purificação da C25A foram desenvolvidos pelo grupo do professor Brian Schoichet da UCSF. As etapas de transformação, expressão e purificação foram realizadas seguindo o protocolo da cruzaina nativa detalhado na Seção 3.2.1, tendo como principal diferença, a etapa de ativação. Uma vez que a enzima seria inativa, esta não seria capaz de se autoativar. O prodomínio N-terminal é importante para o correto enovelamento e estabilização da cruzaina, contribuindo para a expressão na fração solúvel do lisado.⁴⁹ No entanto, a remoção do grande prodomínio (14 kDa) é necessária principalmente para os estudos cristalográficos com inibidores, uma vez que este cobre o sítio catalítico da enzima conforme mostrado na Figura 3.11. A C25A foi desenhada para conter um sítio de clivagem por trombina, com inserção dos aminoácidos LVPA/GS entre o domínio catalítico e o prodomínio na sequência da Figura 3.7. Porém, após inúmeras tentativas de remoção do prodomínio foi obtida uma estrutura cristalográfica mostrando que mesmo após a clivagem por trombina, o prodomínio permanecia interagindo com o domínio catalítico da enzima por fortes interações intermoleculares no *Prosegment Binding Loop* (PBL) existente entre os resíduos Met145 e Met152 da cruzaina, conforme demonstrado na Figura 3.11B.⁴⁹

A falta de densidade eletrônica entre os resíduos Tyr-34 e Gly-1 (numeração da sequência exibida na Figura 3.7) do segmento C-terminal do prodomínio, sugere que este foi removido durante clivagem com trombina e purificação no FPLC. O resíduo Arg-35 interage com o subsítio S2' da cruzaina, impedindo a aproximação

dos inibidores em ensaios cristalográficos. A porção C-terminal do domínio catalítico inicia-se com os aminoácidos GS antes do resíduo Ala1, na estrutura. O projeto da C25A foi cancelado devido aos desafios na remoção do prodomínio e crescimento de cristais com inibidores de interesse. Paralelamente, foi iniciada a engenharia da enzima C25S cuja única diferença em relação à cruzaina nativa, era a substituição da cisteína catalítica pela serina. Ao invés de utilizar sítio de clivagem por trombina, pretendia-se utilizar pequenas quantidades da cruzaina nativa ativa para a clivagem e digestão do prodomínio em um mecanismo conhecido como **trans-ativação**.

Figura 3.11- Estrutura cristalográfica da procatepsina L (código PDB 1CJL) em **A**, com o prodomínio em coral e região catalítica em ciano. Em **B**, sobreposição da estrutura da procatepsina L em ciano, com o domínio catalítico da cruzaina C25A em dourado, e parte N-terminal do prodomínio após clivagem por trombina em vermelho.

Fonte: Autoria própria.

3.3.1 Materiais e métodos: expressão, purificação e cristalização da C25S

A mutação para criação da cruzaina C25S foi realizada pela Dra. Ana Maria do departamento de patologia da UCSF usando como template, o gene da cruzaina nativa apresentado na Seção 3.2.1.⁹ O vetor de clonagem utilizado foi o pET21a da Novagem, contendo genes para cauda de histidina e para resistência a carbanicilina. O protocolo de expressão e purificação desenvolvido é similar ao da cruzaina nativa (descrito na Seção 3.2.1) tendo como diferenças básicas, a temperatura em que é realizada a auto-indução (18°C desde a inoculação do meio), e o processo de trans-

ativação que ocorre a temperatura de 20°C no pH 5,0 utilizando de 10% (razão molar) de cruzaina nativa ativa durante três horas. Após a trans-ativação é realizada uma diálise por 16 horas à 4°C em tampão 50 mM Tris (pH 7,5), 100 mM NaCl, e posterior purificação em FPLC utilizando coluna Superdex 200 no mesmo tampão. Vários experimentos de co-cristalização com os fragmentos foram configurados, variando a concentração da C25S entre 15 e 25 mg mL⁻¹, e o pH entre 5,8 e 7,5.

Centenas de condições de cristalização provenientes dos *kits* de cristalização da Qiagen foram ensaiadas com auxílio do equipamento MOSQUITO (LabTech) usando o método de difusão de vapor com gotas suspensas. Após a identificação dos *hits* iniciais, parâmetros como concentração da proteína e do precipitante, e pH, foram otimizados para a produção de cristais mais resistentes. Cristais do complexo **C25S-Neq0147** foram equilibrados no tampão 0,1 M Hepes pH 7,5 e 1,2 M malonato de sódio, sendo utilizado proteína a 13 mg mL⁻¹ em tampão Tris pH 7,5 com 11 mM do inibidor (concentração final) e 2% (v/v) de DMSO em solução. Cristais para o complexo **C25S-Neq0220-Neq0176** foram equilibrados em 0,1 M Tris pH 8,5 e 1,2 M de tartrato de Na/K, com 5 mM do inibidor Neq220 e 12 mM do inibidor Neq176 adicionados na proteína a 18 mg mL⁻¹ (3% de DMSO). Para o complexo **C25S-K777**, cristais cresceram em condições similares às obtidas na enzima nativa com gotas contendo 1 µL de solução de proteína a 23 mg mL⁻¹ incubada com 3 mM de K777, e 0,5 µL de solução de cristalização 0,1 M Hepes pH 7,5 com 2% PEG 400 e 2,0 M sulfato de amônio. Os cristais cresceram entre quatro e doze semanas a 18°C, sendo crioprotetidos em soluções de etilenoglicol variando entre 20 e 25% (v/v).

3.3.2 Resultados e discussão

A superexpressão de proteínas heterólogas em *E. coli* pode resultar na produção de proteínas no enovelamento incorreto formando corpos de inclusão, cuja purificação é difícil conforme discutido na Seção 3.1.1. As células de *E. coli Arctic Express* DE3-RIL fornecem uma abordagem *in vivo* para aumentar o rendimento da proteína heteróloga na fração solúvel, utilizando para isto baixas temperaturas de expressão. Estas células expressão as chaperoninas Cpn10 e Cpn 60 (originárias da bactéria *Oleispira antártica*) adaptadas ao frio, que facilitam o correto enovelamento da proteína expressa. Estas células de *E. coli* são resistentes à gentamicina, enquanto que o vetor pET21a, possui o gene que confere resistência a carbanicilina.

Inicialmente é adicionada uma quantia de pré-inóculo (contendo *E. coli* crescidas a 37°C) para se atingir um DO₆₀₀ mínima de 0,16 em 1L do meio de auto-indução. A proteína heteróloga é então, expressa a 18°C pelo método de auto-indução mínima usando lactose como agente indutor.⁷ Tal como detalhado na Seção 3.1.1, ocorre uma produção gradual da proteína até a saturação das células de *E. coli* e consumo dos nutrientes do meio. Partindo de 1L de meio, geralmente se obtém cerca de 20 g de *pellet* seco. Após lise celular, se inicia a purificação em cromatografia IMAC onde a proteína ligada a resina Ni-NTA pela cauda 6xHisTag é transferida para a coluna (fração LO), sendo lavada com o tampão A (5 mM de imidazol) e eluída em frações de 5 mL com o tampão B (500 mM de imidazol), conforme exibido na Figura 3.12.

Figura 3.12 - Cruzaína C25S mutante após purificação IMAC. As frações são lavadas com o tampão A e eluídas com o tampão B.

Fonte: Autoria própria.

As frações eluídas com o tampão B, foram unidas e dializadas por 16 horas em 20 mM Tris (pH 7,5) 100 mM NaCl, seguida de uma segunda diálise por 4 horas em 100 mM NaAc, 300 mM NaCl (pH 5,5) para a redução da concentração de imidazol e ajuste do pH para a trans-ativação, cuja identificação das condições ideais foi a parte mais complexa no desenvolvimento da C25S. Testes iniciais da trans-ativação, com 10% (razão molar) de cruzaína nativa ativa, 37°C e 5 mM de DTT (mesma temperatura e concentração de DTT usados na ativação da cruzaína nativa), sempre resultavam na formação de um precipitado branco logo nos primeiros minutos da trans-ativação, cujas quantidades aumentavam-se rapidamente indicando a degradação da proteína. Outras tentativas sem sucesso incluíram a ativação sem DTT, segunda purificação em coluna de níquel e em FPLC.

Após vários testes envolvendo diferentes formas de preparo e proporções da cruzaína nativa em relação a C25S, pH, tempo e temperatura, ficou constatado alguns pontos chave para o processo. A 37°C, a cruzaína nativa está no ápice de sua atividade catalítica podendo clivar o prodomínio e também partes do domínio

catalítico da C25S, além de si mesma (auto-proteólise), causando frequentemente a precipitação da proteína durante o processo de trans-ativação. Por outro lado, a baixas temperaturas (4°C), a cruzaina não tem atividade suficiente para clivar o prodomínio da C25S. Em pH maior que 6, a atividade da cruzaina nativa é menor e a digestão do prodomínio ocorre lentamente. O zimógeno é totalmente processado após 80 minutos quando se utiliza 20% de cruzaina nativa (Figura 3.13A). Porém é aconselhável utilizar a menor quantidade possível da cruzaina nativa para evitar possível proteólise residual, sendo 10% da cruzaina ativa (K_m igual a dois μM), portanto, a melhor razão para o processamento da C25S o que aumenta o tempo de clivagem e digestão do prodomínio para 3 horas conforme Figura 3.13B.

Figura 3.13 - Trans-ativação da cruzaina C25S monitorado em gel SDS, utilizando 20% de cruzaina nativa em **A** e 10% em **B**. Em **C**, a estabilidade da C25S foi monitorada a 20 °C dia após dia. A coluna WT representa a cruzaina nativa após um dia a 20 °C.

Fonte: Autoria própria.

Durante o preparo da cruzaina nativa, esta deve ser congelada em nitrogênio líquido após a ativação e descongelada a 37 °C antes do uso. O zimógeno da C25S deve estar em uma concentração menor que $0,5 \text{ mg mL}^{-1}$ para evitar precipitação. A banda a 26 kDa nos géis da Figura 3.13, correspondem ao domínio catalítico das cruzainas C25S e nativa sendo praticamente impossível separá-las. Durante a trans-ativação a cruzaina nativa digere partes de si mesma, diminuindo gradativamente sua atividade proteolítica. O prodomínio (14 kDa) é clivado e digerido em cadeias polipeptídicas menores que são removidas da solução durante as etapas de diálise, concentração e purificação por cromatografia de exclusão de tamanho ou de troca iônica. A C25S processada é inativa e estável podendo ser guardada por semanas sem degradação em pH 7,5 e temperatura de 20°C, conforme monitorado no gel da Figura 3.13C, onde amostras da proteína na concentração 6 mg mL^{-1} e 20°C foram coletadas todos os dias, desnaturadas, e guardadas a 4°C.

A C25S foi preparada para ensaios de cristalização nos tampões Tris pH 7,5 e Bis-Tris pH 5,8, sendo estes correspondentes ao pH dos ensaios de estabilidade, e

ao pH onde foram obtidos a maior parte dos cristais da cruzaína nativa. Os primeiros cristais (Figura 3.14A) cresceram em condições contendo os ânions malonato, succinato e tartarato como precipitantes. Estas condições foram então otimizadas para diferentes pH, concentração da proteína e do precipitante. Os melhores cristais cresceram nas condições: 0,1 M Hepes (pH 7,5) 1,2 M tartrato de Na/K; 0,1 M Hepes (pH 7,5) e 1,6 M tartrato de K/Na (Figuras 3.14B, C e D). Ensaios de cristalização para a forma Apo também foram feitos, tendo como resultados, cristais difratando até 2,45 Å de resolução principalmente na condição 100 mM Tris (pH 7,5) e 300 mM NaCl. No entanto, estes cristais se demonstraram frágeis para *soaking* com os inibidores, e difíceis de produzir em grande quantidade. Por este motivo, todos os esforços foram dirigidos para as experiências de co-cristalização com fragmentos, que em geral levavam a condições ligeiramente diferentes para cada inibidor.

Figura 3.14 - Em **A**, cristais da C25S obtidos nos ensaios iniciais. Em **B** e **C** estão os cristais otimizados após sucessivos ensaios até a produção de cristais maiores em **D**.

Fonte: Autoria própria

Os melhores co-cristais difrataram a 2.4 Å no grupo espacial $P4_32_12$ com 5 ou 6 moléculas de cruzaína dentro da ASU. Vários conjuntos de dados foram coletados para estes cristais que exibiam, entretanto, pouca densidade eletrônica para a construção do inibidor no sítio. A causa desta resolução mediana e pouca densidade no sítio é também devido ao tipo de inibidores utilizados (reversíveis e de baixa afinidade) e ao possível efeito proteolítico residual da cruzaína nativa usada na trans ativação. Estes cristais eram frágeis e sensíveis à radiação, possivelmente devido ao alto conteúdo de solvente. A solução crioprotetora adequada foi determinada após a obtenção de alguns padrões de difração contendo anéis concêntricos característicos de cristais de gelo. O cristal da C25S em complexo com o Neq0147 e o padrão de difração correspondente, são apresentados nas Figuras 3.15A e C. O cristal produzido para a forma Apo incubado em solução contendo o inibidor Neq0220 e o padrão de difração correspondente danificado pelo *soaking* (sobretudo nas regiões de alta resolução) é exibido nas Figuras 3.15B e D.

Figura 3.15 - Cristal da C25S co-cristalizada com o Neq0147 em **A** e o padrão de difração correspondente em **C**. O cristal da C25S na forma Apo com o inibidor Neq0220 adicionado por *soaking* é mostrado em **B** com padrão de difração em **D**.

Fonte: Autoria própria.

Embora o *soaking* de ligantes dentro de cristais possa ser o método de escolha de uma proteína em particular, é preferível validar o sistema de *soaking* com experimentos de co-cristalização quando possível. O MOB de ligantes em estruturas obtidas por *soaking* muitas vezes não permitem enxergar o grande espectro de mudanças conformacionais que naturalmente ocorreriam durante o reconhecimento molecular entre a proteína e o ligante. No *soaking*, o ligante força o caminho até o sítio catalítico (podendo causar defeitos na rede cristalina e consequente perda de poder de difração e qualidade do cristal), ou muitas vezes é forçado a assumir uma conformação específica que nem sempre é a de menor energia para o complexo, uma vez que o sítio já está definido devido ao empacotamento cristalino. No entanto o *soaking* ainda possibilita a obtenção de estruturas cristalográficas para diferentes tipos de ligantes, sem a necessidade das dispendiosas, demoradas e trabalhosas triagens de condições de cristalização para diferentes complexos. A otimização de condições de cristalização da forma apo da C25S possibilitará a obtenção de grande quantidade de modos de interação para inibidores no futuro próximo.^{22,50}

3.4 Referências bibliográficas

- 1 Hefti, M. H.; Vervoort, J. A Novel Purification Method for Histidine-Tagged Proteins Containing a Thrombin Cleavage Site. **Analytical Biochemistry**, v. 295, p. 180–185, 2001.
- 2 Goulding, C. W.; Perry, L. J. Protein production in Escherichia coli for structural studies by X-ray crystallography. **Journal of Structural Biology**, v. 142, p. 133–143, 2003.
- 3 Derewenda, Z. S. The use of recombinant methods and molecular engineering in protein crystallization. **Methods**, v. 34, n. 1, p. 354–363, 2004.
- 4 Messerschmidt, A. **X-ray Crystallography of Biomacromolecules** 1ed. USA: WILEY-VCH, 2007. v. 1. 318 p.

- 5 Lehninger, A. L.; Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*. 5 ed. USA: Artmed, 2011. v. 1. 839 p.
- 6 Voet, D.; Voet, J. G. **Biochemistry**. 4 ed. USA: Artmed, 2013. v. 1.
- 7 Studier, W. Protein production by auto-induction in high-density shaking cultures. **Protein Expression and Purification**, v. 41, n. 1, p. 207–234, 2005.
- 8 Oehler, S.; Eismann, E. R.; Krämer, H.; Müller-Hill, B. The three operators of the lac operon cooperate in repression. **The EMBO Journal**, v. 9, n. 4, p. 973–979, 1997.
- 9 Lee, G. M.; Balouch, E.; Goetz, D. H.; Lazic, A.; Mckerrow, J. H.; Craik, C. S. Mapping Inhibitor Binding Modes on an Active Cysteine Protease via Nuclear Magnetic Resonance Spectroscopy. **Biochemistry**, v. 51, n. 50, p. 10087–10098, 2012.
- 10 Grabski, A.; Mehler, M.; Drott, D. The Overnight Express Autoinduction System: High-density cell growth and protein expression. **NATURE METHODS**, v. 2, p. 233 - 235, 2005.
- 11 Villaverde, A.; Carrió, M. M. Protein aggregation in recombinant bacteria: biological role of inclusion bodies. **Biotechnology Letters**, v. 25, p. 1385–1395, 2003.
- 12 Shen, A.; Lupardus, P. J.; Morell, M.; Ponder, E. L.; Sadaghiani, A. M.; Garcia, C.; Bogoy, M. Simplified, Enhanced Protein Purification Using an Inducible, Autoprocessing Enzyme Tag. **PLoS ONE**, v. 4, n. 12, p. e8119, 2009.
- 13 Hefti, M. H.; A Novel Purification Method for Histidine-Tagged Proteins Containing a Thrombin Cleavage Site. **Analytical Biochemistry**, v. 295, n. 1, p. 180–185, 2001.
- 14 Graslund, S.; Sagemark, J.; Berglund, H.; Kotenyova, T.; Nilsson, M.; Nordlund, P.; Weigelt, J. The use of systematic N- and C-terminal deletions to promote production and structural studies of recombinant proteins. **Prot Exp Pur**, v. 58, p. 210–221, 2008.
- 15 Chayen, N. E.; Saridakis, E. Protein crystallization: from purified protein to diffraction-quality crystal. **NATURE METHODS**, v. 5, n. 2, p. 147–153, 2008.
- 16 Wilson, W. W. Light scattering as a diagnostic for protein crystal growth—A practical approach. **J Struc Bio**, v. 142, p. 56–65, 2003.
- 17 Drenth, J. **Principles of Proteins X-Ray Crystallography**. 3 ed. USA: Springer Verlag, 2007. v. 1. 332 p.
- 18 Hassell, A. M.; An, G.; Bynum, J. M.; Carter, L.; Gampe, R. T.; Grisard, T. E.; Madauss, K. P.; Wang, L.; Weaver, K. L.; Williams, S. P.; Wisely, G. B.; Shewchuka, L. M. Crystallization of protein–ligand complexes. **Acta Cryst D**, v. 63, p. 72–79, 2007.
- 19 Chayen, N. E. Turning protein crystallisation from an art into a science. **Current Opinion in Structural Biology**, v. 14, n. 1, p. 577–583, 2004.
- 20 Bodenstaff, E. R.; Hoedemaeker, F. J.; Kuil, M. E.; Vrinda, H. P.; Abrahamsa, J. P. The prospects of protein nanocrystallography. **Acta Cryst D**, v. 58, p. 1901–1906, 2002.
- 21 Ochi, T.; Bolanos, V. M.; Stojanoff, V.; Moreno, A. Perspectives on protein crystallisation. **Progress in Biophysics and Molecular Biology**, v. 101, n.1, p. 56–63, 2009.
- 22 Hui, R.; High-throughput protein crystallization. **J Struc Bio**, v. 142, p. 154–161, 2003.
- 23 Dale, G. E.; Oefner, C.; Arcy, A. D. The protein as a variable in protein crystallization. **J Struc Bio**, v. 142, p. 88–97, 2003.
- 24 Asherie, N. Protein crystallization and phase diagrams. **Methods**, v.34, p. 266–272, 2004.
- 25 McPherson, A. **Introduction to Macromolecular Crystallography**. 2° ed. New Jersey, United States: John Wiley & Sons, Inc., 2009. 277 p.
- 26 Bergfors, T. Seeds to crystals. **J Struc Bio**, v. 142, p. 66–76, 2003.
- 27 Krauss, I. R.; Sica, F.; Mattia, C. A.; Merlino, A. Increasing the X-ray Diffraction Power of Protein Crystals by Dehydration: The Case of Bovine Serum Albumin and a Survey of Literature Data. **Int J Mol Sci**, v. 13, n. 1, p. 3782–3800, 2012.
- 28 Alcorn, T.; Juersa, D. H. Progress in rational methods of cryoprotection in macromolecular crystallography. **Acta Cryst D**, v. 66, p. 366–373, 2010.
- 29 Garman, E. F.; Owen, R. L. Cryocooling and radiation damage in macromolecular crystallography. **Acta Cryst D**, v. 62, p. 32–47, 2006.
- 30 Rangel, H. A.; Araújo, P. M. F.; Costa, M. G. Trypanosoma cruzi: Isolation and characterization of a proteinase. **Experimental Parasitology**, v. 52, n. 2, p. 199–209, 1981.

- 31 Eakin, A. E.; Mills, A. A.; Harth, G.; Mckerrow, J. H.; Craik, C. S. The sequence, organization, and expression of the major cysteine protease (cruzain) from *Trypanosoma cruzi*. **J Bio Chem**, v. 267, n. 11, p. 7411-7420, 1992.
- 32 Pamer, E. G.; Eakin, A. Cloning and sequencing of the cysteine protease cDNA from *Trypanosoma brucei rhodesiense*. **Nucleic Acids Research**, v. 18, n. 20, p. 6141, 1990.
- 33 Pamer, E. G.; Davis, C. E.; Magdalene, S. Expression and Deletion Analysis of the *Trypanosoma brucei rhodesiense* Cysteine Protease in *Escherichia coli*. **Infec Immunity**, v. 59, n. 3, p. 1074-1078, 1991.
- 34 Eakin, A. E.; Mcgrath, M. E.; Mckerrow, J. H.; Fletterick, R. J.; Craik, C. S. Production of Crystallizable Cruzain, the Major Cysteine Protease from *Trypanosoma cruzi*. **The journal of biological chemistry**, v. 268, n. 9, p. 6115-6118, 1993.
- 35 Pamer, E. G.; Davis, C. E.; Magdalene, S. Expression and Deletion Analysis of the *Trypanosoma brucei rhodesiense* Cysteine Protease in *Escherichia coli*. **INFECTION AND IMMUNITY**, v. 59, n. 3, p. 1074-1078, 1991.
- 36 Smith, S. M.; Gottesman, M. M. Activity and Deletion Analysis of Recombinant Human Cathepsin L Expressed in *Escherichia coli*. **THE JOURNAL OF BIOLOGICAL CHEMISTRY**, v. 264, n. 34, p. 20487-20495, 1989.
- 37 Mcgrath, M. E.; Mckerrow, J. H.; Craik, C. S.; Fletterick, R. J. The Crystal Structure of Cruzain: A Therapeutic Target for Chagas' Disease. **J Mol Biol**, v. 247, p. 251-259, 1995.
- 38 Gillmor, S. A.; Craik, C. S.; Fletterick, R. J. Structural determinants of specificity in the cysteine protease cruzain. **Protein Science**, v. 6, n. 8, p. 1603-1611, 1997.
- 39 Ferreira, R. S.; Eidam, O.; Keiser, M. J.; Mckerrow, J. H.; Maloney, D. J.; Irwin, J. J.; Shoichet, B. K. Complementarity Between a Docking and a High-Throughput Screen in Discovering New Cruzain Inhibitors. **J Med Chem**, v. 53, n. 13, p. 4891-4905, 2010.
- 40 Derewenda, Z. S. Rational Protein Crystallization by Mutational Surface Engineering. **Structure**, v. 12, n. 1, p. 529-535, 2004.
- 41 Nunez, S.; Venhorst, J.; Kruse, C. G. Target-drug interactions: first principles and their application to drug discovery. **DDT**, v. 17, n. 1 / 2, p. 10 - 23, 2012.
- 42 Mobley, D. L.; Dill, K. A. Binding of Small-Molecule Ligands to Proteins: "What You See" Is Not Always "What You Get". **Structure**, v. 17, p. 489-498, 2009.
- 43 Baron, R.; Mccammon, J. A. Molecular Recognition and Ligand Association. **Annu Rev Phys Chem**, v. 64, p. 151-175, 2013.
- 44 Judice, W. A. S.; Cezari, M. H. S.; Scharfstein, J.; Chagas, J. R.; Juliano, M. A.; Juliano, L. Comparison of the specificity, stability and individual rate constants with respective activation parameters for the peptidase activity of cruzipain and its recombinant form, cruzain, from *Trypanosoma cruzi*. **Eur J Biochem**, v. 268, p. 6578-6586, 2001.
- 45 Alves, L. C.; Melo, R. L.; Cezari, M. L.; Juliano, L.; Juliano, M. A. Analysis of the S2 subsite specificities of the recombinant cysteine proteinases CPB of *Leishmania mexicana*, and cruzain of *Trypanosoma cruzi*, using fluorescent substrates containing non-natural basic amino acids. **Mol Biochem Parasitology**, v. 117, p. 137-143, 2001.
- 46 E D Nery, M. A. J., M Meldal, I Svendsen, J Scharfstein, A Walmsley, L Juliano. Characterization of the substrate specificity of the major cysteine protease (cruzipain) from *Trypanosoma cruzi* using fluorogenic peptides. **Biochem J**, v. 323, p. 427-433, 1997.
- 47 Huang, L.; Brinen, L. S.; Ellman, J. A. Crystal structures of reversible ketone-based inhibitors of the cysteine protease cruzain. **Bioorg Med Chem**, v. 11, n. 1, p. 21-29, 2003.
- 48 Turkenburg, J. P.; Lamers, M. B. A.; Sturtb, S. L.; Williams, D. H. Structure of a C25S mutant of human cathepsin S. **Acta Cryst D**, v. 58, n.1, p. 451-455, 2002.
- 49 Reif, M. M.; Lukas Mach, A. C. O. Molecular Insight into Propeptide-Protein Interactions in Cathepsins L and O. **Biochemistry**, v. 51, p. 8636-8653, 2012.
- 50 Derewenda, Z. S.; Vekilov, P. G. Entropy and surface engineering in protein crystallization. **Acta Cryst D**, v. 62, p. 116-124, 2006.

Capítulo 4

Coleta de Dados e Elucidação Estrutural do MOB de Inibidores na Cruzaína Nativa e C25S Mutante

Uma vez obtida a amostra proteica cristalina, parte-se para a coleta de dados de difração de raios X, sendo esta, a última etapa experimental do método cristalográfico. Neste capítulo são apresentadas todas as etapas deste método desde a coleta de dados até a obtenção do modelo cristalográfico refinado e validado. Este capítulo encerra-se com a apresentação das estruturas cristalográficas obtidas e a descrição do modo de interação de inibidores reversíveis na cruzaína nativa e C25S mutante.

4 COLETA DE DADOS E ELUCIDAÇÃO ESTRUTURAL DO MOB DE INIBIDORES NA CRUZAÍNA NATIVA E C25S MUTANTE

4.1 Fundamentos da cristalografia de raios X em macromoléculas

4.1.1 Cristais e a simetria cristalina

Cristais são sólidos formados pelo arranjo ordenado e periódico de seus constituintes, exibindo uma geometria definida e um padrão regular de repetição em todas as direções.^{1,2} Esta periodicidade nos materiais cristalinos é descrita por um modelo matemático associado ao conceito de simetria. Desta maneira é possível caracterizar um monocristal a partir do conteúdo de uma unidade fundamental de repetição estrutural (a **unidade assimétrica** ou ASU) e a maneira como esta se repete no espaço tridimensional por operações de simetria.^{1,3} A Figura 4.1 representa o modelo de um sólido cristalino constituído por um conjunto de pontos idênticos em um arranjo periódico, que definem a **rede cristalina** $L(\mathbf{r})$.¹ Assim, cada ponto da rede é equivalente com seus vizinhos por simetria translacional, exibindo um padrão de vizinhança similar. A **simetria translacional** descreve uma repetição periódica em intervalos regulares da rede cristalina através de um comprimento, área ou volume e está presente em todos os cristais. A menor unidade que se repete por simetria translacional dentro de um cristal é a **cela unitária** que é caracterizada por três vetores unitários não coplanares **a**, **b**, **c**, e pelos ângulos entre eles α , β e γ . Cristais de macromoléculas exibem células unitárias de grandes dimensões.^{1,2}

Figura 4.1 - Representação esquemática da rede cristalina e dos seis parâmetros de cela unitária correspondentes: **a**, **b**, **c**, α , β e γ .

Fonte: Adaptado de CHATTERJEE, S. K., 2008.¹

Desta forma, a rede de difração $L(\mathbf{r})$ é descrita pelas magnitudes u , v , w nas direções dos três vetores **a**, **b**, **c** da cela e pela simetria translacional, conforme a

Eq. (4.1). O termo δ é a função delta de Dirac que pode assumir somente os valores zero ou um, garantindo que a rede seja tratada matematicamente como um conjunto de pontos discretos, levando a enormes simplificações nas equações.^{1,4}

$$L(\mathbf{r}) = \sum \delta[\mathbf{r} - (u\mathbf{a} + v\mathbf{b} + w\mathbf{c})] \quad \text{Eq. (4.1)}$$

Para uma rede qualquer, muitas escolhas diferentes de celas unitárias são possíveis podendo conter um ou mais pontos de rede por cela. Entretanto, existe sempre uma em que os comprimentos da cela são os três menores vetores não coplanares, sendo esta preferida por convenção. Por valores especiais dos parâmetros de rede, podem ser obtidas somente sete celas unitárias de formas diferentes com os pontos de rede localizados em seus vértices. Estes correspondem aos sete **sistemas cristalinos** em que todos os cristais são classificados (Tabela 4.1). O cúbico é o mais simples e simétrico. Os sistemas restantes ordenam-se de acordo com a seguinte sequência decrescente de simetria: hexagonal, trigonal, tetragonal, ortorrômbico, monoclinico e triclínico.^{1,3} Para algumas estruturas cristalinas com simetria de rotação e reflexão, é por convenção escolhida a cela unitária contendo mais de um ponto de rede. Celas primitivas *P* (ou *R* para sistemas romboédricos) têm pontos de rede somente em seus vértices, enquanto que as celas denominadas centradas contêm pontos de rede adicionais podendo ser em suas faces ou em seu interior. Os símbolos *F* e *I* referem-se a face centrada e a corpo centrado respectivamente, enquanto que *A*, *B*, *C* referem-se a celas unitárias centradas em bases opostas. Das combinações possíveis da simetria de rede primitiva com os diferentes tipos de centragens, resultam as 14 **redes de Bravais**.^{1,5,6}

Tabela 4.1 - Os sete sistemas cristalinos e suas respectivas simetrias pontuais essenciais (grupos de Laue), restrições dos parâmetros de rede e possíveis centragens.

Sistema	Grupos de Laue	Restrições de cela	Tipos de cela
Triclínico	1	Nenhuma	<i>P</i>
Monoclinico	2	$\alpha = \gamma = 90^\circ$	<i>P</i> , <i>C</i>
Ortorrômbico	222	$\alpha = \beta = \gamma = 90^\circ$	<i>P</i> , <i>C</i> , <i>I</i> , <i>F</i>
Tetragonal	4, 422	$a = b$; $\alpha = \beta = \gamma = 90^\circ$	<i>P</i> , <i>I</i>
Trigonal	3, 322	$a = b$; $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$	<i>P</i> (<i>R</i>)
Hexagonal	6, 622	$a = b$; $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$	<i>P</i>
Cúbico	332, 432	$a = b = c$; $\alpha = \beta = \gamma = 90^\circ$	<i>P</i> , <i>I</i> , <i>F</i>

Fonte: Autoria própria.

Quando alguma simetria de inversão, rotação ou reflexão relaciona moléculas entre si em uma cela unitária, a menor unidade estrutural de repetição passa a ser

uma fração desta cela unitária (a unidade assimétrica) que pode consistir de uma molécula simples ou ainda um grupo de mais de uma molécula. Não é a métrica entre os parâmetros de rede que define o sistema cristalino e sim os elementos de simetria rotacional (**simetria de Laue**) que impõem restrições e valores especiais a estes parâmetros (Tabela 4.1). Inversões em relação a um ponto, rotações em torno de um eixo, e reflexões sobre um plano, caracterizam as operações de **simetria pontuais** que podem ser classificadas em rotações próprias (rotações de frações de 360° em torno de um eixo), e rotações impróprias (combinação simultânea de uma rotação e um centro de inversão).^{1,7} A simetria translacional impõem restrições fazendo com que algumas ordens de rotação sejam incompatíveis com a repetição natural da rede. Deste modo, as únicas ordens de rotação possíveis em sólidos cristalinos são 1, 2, 3, 4 e 6, para ambas as rotações próprias e impróprias.^{1,7} As operações de reflexão e inversão são casos especiais das operações de rotação impróprias. Todas as operações de simetria impróprias envolvem a mudança de quiralidade da molécula (presente em estruturas de moléculas quirais), enquanto que as operações de rotação próprias mantêm a mesma configuração.^{1,2,7} A coleção de todas as operações de simetria possíveis envolvendo as rotações impróprias e próprias excluindo-se componentes translacionais para uma molécula, compõem os **32 grupos pontuais** descritos da *International Tables for Crystallography* (ITC) volume A.^{1,2,5} Estes grupos pontuais estão arranjados segundo os sete sistemas cristalinos sendo subdivididos em três categorias. Os 11 grupos pontuais centrosimétricos ou grupos de Laue, permitem determinar qual parte da esfera de difração de dados do espaço recíproco é única e qual é redundante. Dos 21 grupos restantes não centrosimétricos, 11 são enantiomorfos, e contém uma mistura não racêmica de espécies quirais. Os outros 10 são não enantiomorfos que apesar de não terem centro de simetria, apresentam eixos de rotação impróprios sem inversão.^{1,2,7}

Combinando operações de simetria pontuais com elementos da simetria translacional, surgem os eixos helicoidais (*screw axis*) e os planos de deslizamentos (*glide planes*). Os eixos helicoidais resultam da combinação das operações de rotação com translação de quantidade igual a um múltiplo de $1/n$ de repetição da rede ao longo da direção do eixo de rotação de ordem n . Como exemplo, se o eixo helicoidal é ao longo da direção b , a operação de simetria 2_1 irá mover um ponto inicial da posição x, y, z para $-x, \frac{1}{2} + y, -z$. As operações de simetria com eixos helicoidais possíveis são $2_1, 3_1, 3_2, 4_1, 4_2, 4_3, 6_1, 6_2, 6_3$ e 6_4 . Os planos de deslizamentos

resultam das operações de reflexão e translação podendo ocorrer paralelamente aos eixos de cela unitária cujos símbolos são a , b , c , ou n , quando paralelos a quaisquer diagonais.¹⁻³ Eixos helicoidais e planos de deslizamentos são mais comuns do que rotações e reflexões puras em estruturas cristalinas, ocorrendo em moléculas com formas e distribuição de cargas irregulares que em geral, se empacotam de maneira mais compacta e eficiente.^{1,5} Da combinação dos 32 grupos pontuais com elementos da simetria translacional, surgem os 230 **grupos espaciais** convencionais (Figura 4.2) detalhados no volume A da ITC, que representam as diferentes maneiras em que a matéria se empacota no estado sólido cristalino.² Grupos espaciais de maior simetria são usualmente encontrados em sólidos iônicos, inorgânicos e minerais.⁸ Proteínas são formadas por L-aminoácidos que são quirais e, portanto elementos de simetria de inversão não são permitidos. Desta forma, o número de grupos espaciais possíveis para proteínas é reduzido de 230 para 65 grupos espaciais quirais.^{2,8}

Figura 4.2 - Combinação dos elementos de simetria pontuais e translacionais para a formação dos 230 grupos espaciais.

Fonte: Autoria própria

4.1.2 Difração de raios X por monocristais

A simetria dos materiais cristalinos possibilita o fenômeno da difração de raios X. Qualquer arranjo regular de objetos pode atuar como “rede de difração” de ondas que tenham comprimento de onda (λ) da mesma ordem de grandeza que a distância entre estes objetos, que no caso de moléculas orgânicas ou de macromoléculas, é da ordem dos raios X.^{5,6} Esta radiação eletromagnética é produzida quando um feixe

de partículas eletricamente carregadas e com grande energia cinética é rapidamente desacelerado pela colisão com um alvo metálico produzindo radiação com λ bem definido variando de 0,4 à 2,5 Å dependendo do metal, sendo este o intervalo de interesse da cristalografia estrutural.^{4,6} Quando um feixe de raios X atinge um material cristalino, os elétrons de seus átomos são forçados a oscilar devido ao campo elétrico da radiação incidente tornando, segundo a teoria clássica das ondas radiação eletromagnética, uma nova fonte espalhadora em todas as direções.^{5,9} O fenômeno da difração surge da interferência destas ondas espalhadas pelos vários espalhadores discretos, que quando arranjados ordenadamente, possibilitam a obtenção de um padrão regular de ondas difratadas, recíproco ao padrão de distribuição dos átomos.^{5,6,9} A condição de interferência construtiva em amostras cristalinas foi caracterizada por Bragg em 1912 que propôs uma equação simples, a Eq. (4.6) representada graficamente pela Figura 4.3, informando a direção dos feixes difratados.^{5,6} Bragg considerou a difração como o resultado da reflexão do feixe de raios X por conjuntos de planos reticulares (que passam por pontos da rede).⁶ Estes planos são caracterizados por três números inteiros hkl , denominados **índices de Miller**, e por suas respectivas distâncias interplanares (d_{hkl}).^{5,6}

Figura 4.3 - Representação esquemática fora de escala da Lei de Bragg e dependência do vetor espalhamento $\mathbf{h}$ em relação à $\mathbf{s}$ e $\mathbf{s}_0$.

Fonte: Adaptado de GIACOVAZZO, C., 2002.⁵

No cristal concebido pela Figura 4.3 destacam-se dois centros espalhadores discretos (A e C) distantes um do outro por d_{hkl} , cuja descrição é dada pelo vetor posição $\mathbf{r}$ da Eq. (4.2). O vetor unitário $\mathbf{s}_0$, descreve a direção da radiação incidente enquanto que o outro vetor unitário $\mathbf{s}$, representa uma direção particular do feixe espalhado. A diferença de caminho óptico (δ) e a diferença de fase (φ) entre ondas espalhadas pelo conteúdo eletrônico dos átomos A e C são dadas pelas Eq. (4.3) e Eq. (4.4) respectivamente, considerando a radiação incidente monocromática.^{4,5}

$$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c} \quad \text{Eq. (4.2)}$$

$$\delta = \mathbf{r} \cdot \mathbf{s} - \mathbf{r} \cdot \mathbf{s}_0 = \mathbf{r}(\mathbf{s} - \mathbf{s}_0) \quad \text{Eq. (4.3)}$$

$$\varphi = \frac{2\pi}{\lambda} \delta = 2\pi \mathbf{r} \frac{(\mathbf{s} - \mathbf{s}_0)}{\lambda} = 2\pi \mathbf{r} \cdot \mathbf{h} \quad \text{Eq. (4.4)}$$

Para que as duas ondas espalhadas estejam em fase, é necessário que a diferença de caminho (δ) entre elas seja zero ou um número inteiro do λ . Pela Eq. (4.4), se observa que o cálculo da interferência entre ondas espalhadas depende da combinação dos três parâmetros $\mathbf{s}$, $\mathbf{s}_0$ e λ , que resultam no **vetor espalhamento atômico** ($\mathbf{h}$) descrito na Eq. (4.5) cujo módulo, é função do ângulo de espalhamento θ . Por definição, o vetor $\mathbf{h}$ indica a direção que satisfaz a condição de interferência construtiva para o feixe espalhado e seu módulo pode variar de 0 a $2/\lambda$. A Eq. (4.5) estabelece ainda que $\mathbf{h}$ seja perpendicular aos planos hkl e que o seu módulo $|\mathbf{h}|$ seja igual ao inverso da distância d_{hkl} . Assim, o vetor $\mathbf{h}$ está descrito num espaço onde cada eixo do seu sistema de coordenadas tem dimensão do recíproco da distância.⁴ O rearranjo da Eq. (4.5) fornece a **equação de Bragg**, Eq. (4.6), sendo d_{hkl} a distância entre os planos hkl e θ o ângulo de incidência.^{5,6}

$$|\mathbf{h}| = \left| \frac{\mathbf{s} - \mathbf{s}_0}{\lambda} \right| = \frac{2|\sin\theta|}{\lambda} = \frac{1}{d_{hkl}} \quad \text{Eq. (4.5)}$$

$$2d_{hkl}\sin\theta = n\lambda \quad \text{Eq. (4.6)}$$

A equação de Bragg é uma consequência da periodicidade da rede cristalina sendo a condição necessária para que ocorra interferência construtiva em um feixe espalhado por duas camadas adjacentes de planos reticulares. Esta lei estabelece que se a radiação espalhada estiver em fase, a diferença de caminho óptico será um múltiplo inteiro do λ do feixe incidente. Nesta condição são determinadas as direções de interferência construtivas em um cristal formado por uma distribuição discreta de densidades eletrônicas.^{5,10} Para o espalhamento elástico, cada feixe de raios X espalhado na direção $\mathbf{h}$ possui o mesmo λ do feixe incidente, tendo porém, fases e amplitudes próprias. Conforme descrito na Eq. (4.4), a fase depende da mudança de direção $\mathbf{h}$ em relação ao feixe incidente, e da posição $\mathbf{r}$ do espalhador. A amplitude do feixe espalhado depende da densidade eletrônica $\rho(\mathbf{r})$ dos espalhadores, ilustrado pelos átomos A e C na Figura 4.3.^{5,10,11} Para cada rede cristalina é possível construir uma **rede recíproca** que possui propriedades recíprocas a esta. Considerando uma rede cristalina que possua uma cela unitária definida pelos

vetores **a**, **b**, **c** definimos uma cela unitária da rede recíproca pelos vetores, **a***, **b***, **c*** dados pela Eq. (4.7), onde *V* é o volume da cela unitária.^{5,6,10}

$$\mathbf{a}^* = \frac{1}{V}(\mathbf{b} \times \mathbf{c}) \quad \mathbf{b}^* = \frac{1}{V}(\mathbf{c} \times \mathbf{a}) \quad \mathbf{c}^* = \frac{1}{V}(\mathbf{a} \times \mathbf{b}) \quad \text{Eq. (4.7)}$$

O espaço recíproco é um conceito essencial para o estudo do retículo cristalino e suas propriedades de difração. O vetor espalhamento **h** e a diferença de fase φ podem agora ser descritos neste espaço, conforme as Eq. (4.8) e Eq. (4.9).

$$\mathbf{h} = \frac{\mathbf{s} - \mathbf{s}_0}{\lambda} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \quad \text{Eq. (4.8)}$$

$$\varphi = 2\pi(\mathbf{x}\mathbf{a} + \mathbf{y}\mathbf{b} + \mathbf{z}\mathbf{c}) \cdot (h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*) \quad \text{Eq. (4.9)}$$

Para que difração ocorra na perspectiva do espaço recíproco, é necessário que o vetor **h** esteja sobre um ponto do retículo recíproco satisfazendo assim as **equações de Laue** Eq. (4.10) que são obtidas a partir do produto escalar da Eq. (4.8) pelos vetores **a**, **b** e **c** do retículo cristalino.^{5,10,11} Uma família de planos do espaço recíproco perpendiculares à direção **b**, com espaçamento interplanar igual a $1/\mathbf{b}$ e que satisfaz as Eq. (4.10), estão representados na Figura 4.4. As direções representadas para o vetor **h** satisfazem a relação $\mathbf{h} \cdot \mathbf{b} = k$. Os máximos de difração ocorrerão apenas para algumas direções particulares definidas pelo vetor **h**, sendo o produto $\mathbf{r} \cdot \mathbf{h}$ nulo ou inteiro, de forma a satisfazer às condições de Laue.^{5,11}

$$\mathbf{h} \cdot \mathbf{a} = h \quad \mathbf{h} \cdot \mathbf{b} = k \quad \mathbf{h} \cdot \mathbf{c} = l \quad \text{Eq. (4.10)}$$

Figura 4.4 - Planos cristalográficos em **A** e representação geométrica das condições de Laue em **B**.

Fonte: Adaptado de GIACOVAZZO, C., 2002.⁵

A construção de Ewald, ilustrada na Figura 4.5, permite entender a difração na perspectiva de monocristais. Uma esfera de raio $1/\lambda$ é desenhada em torno do cristal tendo o vetor $\mathbf{s}_0/\lambda$ paralelo ao feixe incidente e o ponto (000) tomado como origem do retículo recíproco. Assim a condição para ocorrer difração a partir dos

planos (hkl) é que o ponto $P(hkl)$ toque a superfície da **esfera de Ewald**. A direção do feixe difratado ($\mathbf{s}/\lambda$) é encontrada unindo-se a origem com a reflexão (101).^{10,12}

Figura 4.5 - A esfera de Ewald. Definição dos parâmetros e do vetor espalhamento $\mathbf{h}$ em **A** e intersecção da reflexão (101) com a superfície da esfera de Ewald em **B**, satisfazendo assim as condições de Bragg e de Laue para a difração.

Fonte: Adaptado de CLEGG, W., 2009.¹²

4.1.3 Função densidade eletrônica e o problema das fases

Considerando que um elétron isolado espalhe raios X com intensidade I , é de se esperar que em um átomo de Z elétrons o espalhamento tenha, portanto, uma intensidade ZI . Porém, como as distâncias entre os elétrons em um átomo são da ordem do λ dos raios X, as ondas espalhadas interferem umas com as outras, de forma a existir uma intensidade ZI somente na direção de incidência ($2\theta = 0$). Para o espalhamento em outras direções temos interferência parcialmente destrutiva, e assim, a amplitude total cai com o aumento do ângulo de espalhamento (2θ).⁹ Esta habilidade de um átomo espalhar raios X é definida pelo **fator espalhamento atômico** f que corresponde ao quociente entre a amplitude da onda espalhada por um átomo (E_a) e a espalhada por um elétron (E_e) conforme a Eq. (4.11). O valor de f é dependente do λ da radiação incidente e do fator de temperatura T_{θ} , tendo o número atômico Z do átomo como valor máximo.^{5,10,11}

$$f = \frac{E_a}{E_e} \quad \text{Eq. (4.11)}$$

Uma maneira conveniente de se analisar o espalhamento de raios X por todos os átomos do cristal é a partir da identificação do espalhamento resultante da cela unitária que é descrito pelo **fator estrutura** F_{hkl} conforme a Eq. (4.12). A amplitude total da onda espalhada por uma cela unitária definida pelos vetores $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, será a

soma vetorial das contribuições dos n átomos da cela onde f_j é o fator de espalhamento atômico do j -ésimo átomo. A magnitude deste fator depende unicamente da posição relativa de todos os átomos na cela unitária e de seus respectivos fatores de espalhamento f . F_{hkl} é um número complexo que representa o espalhamento de raios X por todos os componentes da cela unitária sendo uma medida da amplitude de reflexão para um conjunto de planos hkl .^{6,9,11}

$$F_{hkl} = \sum_{j=1}^n f_j e^{2\pi i \cdot (hx_j + ky_j + lz_j)} \quad \text{Eq. (4.12)}$$

Os fatores estruturas dependem somente da estrutura cristalina e dão uma descrição matemática do **padrão de difração** obtido do experimento de raios X. Se as posições de todos os átomos na cela unitária são conhecidas então o correspondente padrão de difração pode ser calculado.¹² Na ausência de um efeito conhecido como espalhamento anômalo, que só é significativo quando elementos pesados estão presentes na estrutura, todo o padrão de difração exibe simetria de inversão, seja a estrutura cristalina centro simétrica ou não. Isto é conhecido como **lei de Friedel** e diz que $|F(hkl)| = |F(\bar{h}\bar{k}\bar{l})|$.⁵ A função matemática que descreve todo o conteúdo da estrutura cristalina é a **função de densidade eletrônica** $\rho(xyz)$ que varia periodicamente de uma cela unitária a outra através do cristal, como resultado do ordenamento regular dos átomos no estado cristalino. Devido a esta periodicidade, $\rho(xyz)$ pode ser descrita por uma série de Fourier tridimensional. Portanto, o padrão de difração $F(\mathbf{h})$ é a transformada de Fourier $\mathbf{T}$ da função de densidade eletrônica $\rho(\mathbf{r})$ conforme Eq. (4.13), e esta por sua vez, é a transformada inversa $\mathbf{T}^{-1}$ do padrão de difração $F(\mathbf{h})$ dado pela Eq. (4.14).^{6,11}

$$F_{hkl} = \mathbf{T}_{\rho(xyz)} \quad \text{Eq. (4.13)}$$

$$\rho(xyz) = \mathbf{T}^{-1}_{F_{hkl}} \quad \text{Eq. (4.14)}$$

A partir do padrão de difração, composto pela amplitude e fase intrínseca dos F_{hkl} difratados em direções de espalhamento discretas e definidas, a função $\rho(xyz)$ pode ser então calculada para cada ponto da cela através de suas coordenadas fracionárias por meio da síntese de Fourier dada pela Eq. (4.15), viabilizando a construção de mapas de densidade eletrônica de uma molécula em um cristal.^{6,11}

$$\rho(xyz) = V^{-1} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i \cdot (hx + ky + lz)} \quad \text{Eq. (4.15)}$$

Quando um mapa de densidade eletrônica com resolução em torno de 1,0 Å é calculado, os picos de densidade eletrônica podem ser atribuídos às posições dos centros espalhadores atômicos. Todavia, o cálculo desses mapas envolve a obtenção das fases intrínsecas φ_{hkl} dos fatores de estrutura F_{hkl} . O **problema da fase** fica evidente através da Eq. (4.16), uma vez que o conhecimento da função $\rho(\mathbf{r})$ para cada posição $\mathbf{r}$ da cela unitária depende do conhecimento das fases $\varphi(\mathbf{h})$ que são sistematicamente perdidas no processo de medida. O problema da fase constitui uma limitação fundamental em medições físicas estando relacionada em última análise a natureza das medições na mecânica quântica. Durante o experimento são obtidas apenas as amplitudes $|F_{hkl}|$ em posições indexadas por medidas diretas das intensidades considerando um número de correções.^{5,10,12}

$$F_{hkl} = |F_{hkl}|e^{i\varphi_{hkl}} \quad \text{Eq. (4.16)}$$

Várias aproximações foram desenvolvidas para contornar este problema. Em cristalografia de pequenas moléculas, fases são obtidas por relações diretas entre as amplitudes de todas as reflexões hkl pelos Métodos Diretos ou pela determinação direta de átomos pesados no Método de Patterson.⁵ Em cristalografia de proteínas, as fases podem ser determinadas por métodos experimentais tais como *Multiple Isomorphous Replacement* (MIR) em que átomos pesados são inseridos na estrutura, *Single-wavelength Anomalous Dispersion* (SAD) que utiliza o efeito da dispersão anômala, ou ainda pelo *Multi-wavelength Anomalous Dispersion* (MAD) onde dados são medidos em diferentes λ .^{4,13} O método de faseamento utilizado neste projeto, o de substituição molecular, necessita que estruturas cristalográficas de proteínas homólogas estejam disponíveis para que suas fases sejam “emprestadas” para o cálculo dos primeiros mapas de densidade eletrônica da estrutura de interesse.⁴⁻⁶

4.2 Método cristalográfico: da coleta de dados até a validação do modelo

4.2.1 Coleta de dados cristalográficos

Quando cristais de qualidade (grande isomorfismo e poder de difração) são obtidos, parte-se para a última etapa experimental considerada fundamental para a precisão do modelo final, que é a coleta de dados cristalográficos. Esta coleta pode ocorrer em difratômetros convencionais ou em fontes de luz síncrotron cujo emprego tem crescido em cristalografia de proteínas.¹⁴⁻¹⁶ Estas máquinas aceleram partículas

carregadas a velocidades relativísticas com energia da ordem de giga elétron-volts que são mantidas no vácuo orbitando em uma trajetória quase circular através de campos magnéticos no anel de armazenamento. Quando as partículas são expostas a um campo magnético de um ímã de flexão, estas sofrem uma aceleração angular e uma mudança de trajetória que causa a emissão de radiação eletromagnética em uma ampla faixa de λ , incluindo raios X. As principais vantagens da utilização de linhas de luz sincrotron são a elevada intensidade dos raios X produzidos e a capacidade de ajuste do λ (tunabilidade) de acordo com as necessidades do experimento.¹⁵ Maior intensidade de luz significa realizar a coleta de maneira rápida, com melhor resolução e preservação das propriedades da amostra, que em geral é frágil e susceptível ao dano causado pela radiação ionizante.¹⁴ Esta radiação pode induzir a mudanças estruturais durante o experimento tais como expansão da cela unitária e perda de isomorfismo do cristal, levando a dados imprecisos e desordens que dificultam a elucidação estrutural.¹⁷⁻¹⁹ Portanto, a estratégia de coleta deve priorizar a obtenção do máximo de informações com a mínima quantidade de dados e tempo de exposição. O ajuste do λ do feixe incidente, permite que este cubra as bordas de absorção dos átomos pesados presentes ou introduzidos na proteína, permitindo a realização dos experimentos MIR, MAD e SAD (Seção 4.1.3) e a solução indireta do problema das fases.^{14,20}

Durante a coleta de dados, são obtidas imagens com o padrão de difração bidimensional contendo pixels de intensidades e incertezas associadas às medidas de cada reflexão (Figura 4.6A). Como os cristais exibem simetria, algumas reflexões serão equivalentes e desta forma terão intensidades idênticas teoricamente. Porém, toda reflexão é medida com certo grau de erro. Logo, quanto maior a **multiplicidade** (número de medidas de cada reflexão equivalente), mais acurada é a estimativa final das intensidades médias das reflexões.²¹ O padrão de difração é fruto do arranjo dos “planos de reflexão” da rede recíproca. Quando estes planos interseccionam a esfera de Ewald (Seção 4.1.2), projetam no detector pontos arranjados em um conjunto de elipses concêntricas originárias da mesma família de planos paralelos do espaço recíproco, conforme exibido na Figura 4.6A.²² Como cristais de macromoléculas têm dimensões de celas unitárias muito maiores do que o λ dos raios X utilizados, muitos pontos da rede recíproca (reflexões) não estarão na superfície da esfera de Ewald. Além disto, a posição de cada um dos átomos na estrutura cristalina influencia as intensidades de todas as reflexões e, por outro lado, a intensidade de cada reflexão

individual depende das posições de todos os átomos da cela unitária. Logo não será possível resolver apenas uma pequena fração da estrutura cristalina sem modelar todo restante. Portanto, é preciso obter as intensidades de todas as reflexões únicas teoricamente possíveis da cela unitária para se resolver a estrutura cristalina, sendo este parâmetro conhecido como **completeza**. Para que as intensidades difratadas de todos os “planos de reflexão” sejam coletadas, considerando uma radiação monocromática (mesmo λ , portanto, uma esfera de Ewald com raio constante), é preciso que o cristal seja rotacionado (método rotativo de coleta) por ângulos de rotação $\Delta\varphi$ específicos em torno de um eixo perpendicular ao feixe incidente, cuja projeção no detector formará agora duas elipses na forma de meia-lua como mostradas na Figura 4.6B. Como as reflexões estão arranjadas em famílias de planos paralelos, existirão famílias de meia-lua concêntricas no detector.²²⁻²⁴

Figura 4.6 – Em **A**, representação do padrão de difração para um cristal estacionário formando um pequeno número de reflexões arranjadas em um conjunto de elipses restritas e em **B**, quando o cristal é rotacionado por um ângulo $\Delta\varphi$, as reflexões dos mesmos conjuntos de planos na rede recíproca formarão uma “meia-lua” limitada por duas elipses correspondendo às posições iniciais e finais.

Fonte: Adaptado de DAUTER, Z., 1999, p. 1708.²²

Para valores menores de $\Delta\varphi$ (Figura 4.7A), as reflexões individuais poderão ser repartidas entre sucessivas imagens de difração (Figura 4.8C), requerindo um número maior de imagens coletadas o que torna a amostra mais susceptível a dano causado pela radiação. Por outro lado, $\Delta\varphi$ muito grande pode levar a sobreposição de luas vizinhas (Figura 4.7C) sendo um fator prejudicial na etapa de processamento dos dados. Valores entre $0,2^\circ$ e $1,5^\circ$ são geralmente utilizados.^{22,23} No método de coleta *fine φ -slicing*, onde os valores de $\Delta\varphi$ são menores que $0,1^\circ$ ou cerca de um terço da mosaicidade do cristal, as reflexões estão presentes em uma série de imagens consecutivas (Figura 4.8C) que quando coletadas utilizando os modernos detectores CCD e Pilatus, permitem a construção de perfins 3D mais acurados para

cada reflexão. Além disto, o ruído de fundo é minimizado o que melhora o parâmetro $\bar{I}/\sigma(\bar{I})$ para as reflexões fracas e a qualidade dos dados para amostras de baixo poder de difração. No entanto, o uso de fatias muito finas podem levar a erros sistemáticos associados sincronização dos movimentos do cristal e do detector.²⁵

Figura 4.7 - Meias-luas concêntricas resultantes da família de planos paralelos do espaço recíproco. Estas apresentam perfins estreitos para ângulos de rotação $\Delta\varphi$ pequenos em **A**, que se alargam com o aumento de $\Delta\varphi$ em **B**, podendo se sobrepor com a meia-lua vizinha se $\Delta\varphi$ for ainda maior em **C**.

Fonte: Adaptado de DAUTER, Z., 1999, p. 1708.²²

A estratégia de coleta deve ser ajustada para garantir a máxima completeza, que é em geral conseguida coletando-se uma “esfera de dados” (φ , ou ângulo total de rotação do cristal) entre 180° e 360° dependendo da simetria do grupo espacial, o que resulta também em múltiplas medidas de intensidades equivalentes, e, portanto, dados mais acurados. Existem regiões mascaradas, sobretudo próximas ao eixo de rotação, onde os pontos da rede recíproca nunca tocarão a esfera de Ewald durante a rotação do cristal. Por isto a completeza nunca será exatamente 100%, sobretudo para a camada de maior resolução.²⁵ A distância do cristal até o detector deve ser ajustada para que se colete intensidades da camada de maior resolução dos dados. Quanto maior esta distância, melhor será a razão entre sinal e ruído $[\bar{I}/\sigma(\bar{I})]$ e maior será a distância entre os pontos ou detalhamento na imagem. Porém para distâncias muito grandes, perdem-se informações da camada de maior resolução. O tempo de exposição por imagem é o fator que mais influencia nas intensidades das reflexões e portanto, na razão $\bar{I}/\sigma(\bar{I})$. A decisão sobre o tempo de exposição dependerá de fatores tais como o número de imagens coletadas, tempo disponível para coleta, fragilidade da amostra e resolução.^{25,26} É melhor ter dados completos de baixa intensidade, do que apenas uma parte incompleta de dados com intensidade e qualidade superiores. A completeza é fundamental para a identificação das fases, enquanto que a qualidade dos dados influenciará no refinamento para um modelo mais realista. Esta qualidade pode ser avaliada pelo parâmetro R_{merge} , que mede a

concordância entre intensidades de múltiplas medidas equivalentes de uma dada reflexão ($I_{h,i}$) em relação a sua média ($\bar{I}_h$) definida pela Eq. (4.17), onde h enumera as reflexões únicas e i os n contribuintes por simetria equivalente.^{27,28}

$$R_{merge} = \frac{\sum_h \sum_{i=1}^n (|\bar{I}_h - I_{h,i}|)}{\sum_h \sum_{i=1}^n I_{h,i}} \quad \text{Eq. (4.17)}$$

O valor de R_{merge} pode ser muito influenciado pela multiplicidade dos dados. Um parâmetro mais elaborado livre do efeito da multiplicidade, R_{meas} , é calculado ajustando-se o valor de R_{merge} por um fator $\sqrt{n}/(n-1)$ que relata de forma confiável a consistência das medidas individuais. Uma variante adicional, R_{pim} , reporta a precisão esperada da intensidade média ($\bar{I}_h$) sendo diminuído por um fator $1/\sqrt{n}$ quando comparado com o R_{meas} . Um bom conjunto de dados de difração é caracterizado por um valor $R_{merge} < 10\%$. Na camada de alta resolução, o R_{merge} pode atingir 30-40 % em cristais de baixa simetria e até 80 % para cristais com grupos espaciais de maior simetria onde a redundância é geralmente maior.^{25,27,28}

4.2.2 Processamento e redução de dados coletados

Uma vez coletado um conjunto de imagens contendo o padrão de difração da amostra, parte-se para o processamento e redução de dados que envolve as etapas de: (i) indexação das reflexões observadas no detector plano; (ii) integração das intensidades de difração; (iii) correção de dados, escalonamento e pós-refinamento; (iv) determinação do grupo espacial e (v) redução dos dados.²⁴ Programas com esta finalidade incluem o XDS, imosflm, HKL2000, XSCALE, Scala, Aimless e Denzo scalepack.^{23,29,30} Duas informações são obtidas a partir do padrão de difração: a disposição geométrica das reflexões que fornecem elementos da rede e da simetria cristalina; e as intensidades das reflexões que fornecem informações sobre o conteúdo da ASU. Na **indexação** ocorre à localização dos picos de difração mais intensos para duas imagens localizadas em geral a 90° uma da outra, sendo obtidas informações sobre cela unitária e sistema cristalino. Como resultado da indexação, para cada pico de intensidade são atribuídos índices de Miller, hkl , para as reflexões observadas de acordo com sua localização. Os parâmetros iniciais de cela definidos a partir das localizações dos picos são limitados em precisão, uma vez que estes estão correlacionados com a distância entre cristal e detector. Todos os parâmetros do cristal e do detector são então refinados de maneira a minimizar as discrepâncias

entre posições de pontos observados e preditos para um conjunto de algumas imagens.^{24,31} Desta forma, valores acurados dos parâmetros de cela unitária e da orientação do cristal são determinados, e as imagens podem então ser integradas.

Os dados obtidos da coleta formam uma imagem 3D do espaço recíproco constituído por uma série bidimensional de imagens de difração, cada uma delas representando uma fatia curva diferente deste espaço conforme representado na Figura 4.8. Assim, na etapa de **integração**, as “fatias” de intensidades individuais para a mesma reflexão presentes em uma ou mais imagens são integradas.³² Na integração bidimensional (2D) implementada no programa imosflm, as intensidades de reflexões parcialmente gravadas (Figura 4.8B e C), são avaliadas de maneira independente pelo método de ajuste de perfil (*profile fitting*) 2D e apenas somadas para compor a intensidade total durante as etapas de escalonamento e redução.^{29,32} Em contradição, no procedimento de integração 3D utilizado pelo software XDS, o método de ajuste de perfil é aplicado a intensidade total obtida pela soma das intensidades parcialmente gravadas sendo esta etapa realizada antes do escalonamento. Em geral, a integração 3D leva a dados processados ligeiramente melhores principalmente quando o método de coleta *fine ϕ -slicing* é utilizado.^{24,29}

Figura 4.8 - Representação esquemática de uma reflexão F_{hkl} qualquer inteiramente gravada em **A**, e parcialmente gravada em diferentes imagens de difração em **B** e **C** de acordo com o ângulo $\Delta\phi$.

Fonte: LESLIE, A. G. W., 2006, p. 52.³²

Após a integração, parte-se para a determinação do correto grupo espacial sendo este, em alguns casos, definido apenas ao final do refinamento. Felizmente, para proteínas o número de grupos espaciais possíveis é reduzido de 230 para 65 possibilidades (Seção 4.1.1). Além disto, toda a etapa de integração e determinação de fatores de correção para as intensidades integradas só depende da simetria do

grupo de Laue (simetria do padrão de difração) que é obtida durante a indexação a partir da posição das intensidades observadas. Quando as intensidades integradas estão disponíveis, análises estatísticas possibilitam determinar a existência de algum centro de inversão na estrutura cristalina. Pela análise das ausências sistemáticas no padrão de difração, é possível determinar a existência de elementos de simetria translacionais do tipo eixos helicoidais e planos de deslizamentos. Esta série de informações associadas a indicadores de qualidade como o R_{meas} , calculado para cada uma das possibilidades, reforçam a decisão sobre o correto grupo espacial.^{24,33}

A imagem de difração 2D tem uma visão distorcida do espaço recíproco. Esta distorção ocorre em função de vários fatores que podem surgir de mudanças na intensidade do feixe incidente, variação do volume iluminado do cristal, dano causado pela radiação, mosaicidade da amostra e outros. Vários fatores de correção e de escala relativa entre as medidas são aplicados durante o **escalonamento** que tenta corrigir erros sistemáticos para que medidas de reflexões equivalentes sejam idênticas. Os parâmetros do cristal são precisamente refinados usando agora, todo o conjunto de dados. Também no escalonamento, o limite de resolução dos dados que será utilizado na solução e no refinamento da estrutura é selecionado baseado em parâmetros para a camada de maior resolução tais como a completeza maior do que 80%, R_{merge} menor que 50%, $I/\sigma(I)$ maior que dois. Truncar os dados em um limite de resolução adequado permite remover ruídos do modelo e aumentar a razão entre dados e parâmetros, o que leva a um maior detalhamento estrutural. No final do processamento dos dados, as intensidades equivalentes são mergidas em um único conjunto de fatores estruturas, $|F_{hkl}|$, de acordo com a simetria do grupo espacial, e as intensidades fora dos padrões de qualidade são removidas do conjunto.^{23,24,27,34}

4.2.3 Faseamento e solução estrutural

Cada reflexão (ou fator estrutura, F_{hkl}) no padrão de difração corresponde a uma onda com amplitude $|F_{hkl}|$ e fase φ_{hkl} próprias, conforme descrito na Eq. (4.16) da Seção 4.1.3. A amplitude é facilmente calculada pela raiz quadrada da intensidade, mas as fases são perdidas durante a coleta dos dados. Muitos métodos têm sido usados para o cálculo das fases iniciais (φ_{calc}), levando a uma aproximação dos mapas de densidade eletrônica, $\rho(xyz)$. Dentre estes métodos, destaca-se o de substituição molecular que requer uma estrutura modelo, em geral obtida a partir de

proteínas homólogas com elevada identidade estrutural ($> 50\%$) e um desvio médio quadrático (rmsd) menor do que $2,0 \text{ \AA}$ entre os átomos de carbono alfa (C_α) do modelo e da estrutura a ser determinada.³⁵ A substituição molecular utiliza a função Patterson, que é calculada empregando a mesma soma de Fourier da função $\rho(xyz)$ (densidade eletrônica, Eq. 4.15) tomando, porém, o quadrado dos fatores estrutura (F_{hkl})² como coeficientes. Logo, em um mapa Patterson todas as fases são definidas como zero resultando em um mapa de densidade onde picos correspondem a vetores distância entre átomos individuais, ao invés de posições atômicas absolutas. Um mapa Patterson é calculado a partir das intensidades experimentais e este é comparado ao mapa Patterson teórico obtido a partir das coordenadas da estrutura homóloga. Pela rotação e posterior translação de um mapa Patterson no outro, a posição dos átomos pode ser encontrada e as fases iniciais φ_{calc} determinadas.³⁵⁻³⁹ Programas como Molrep e Phaser realizam esta tarefa.^{36,40}

O resultado primário do experimento de difração de raios X é o **mapa de densidade eletrônica** do cristal. Este é construído a partir da função $\rho(xyz)$ calculada pela transformada de Fourier do conjunto de amplitudes das reflexões observadas no experimento (F_{obs}) e as fases calculadas (φ_{calc}) obtidas na solução da estrutura. Tais mapas (F_{obs} , φ_{calc}) representam uma aproximação da estrutura verdadeira, dependendo de quão acurada é a φ_{calc} . O **mapa de diferença** é um tipo de mapa de densidade eletrônica calculado utilizando as diferenças entre amplitudes observadas e calculadas, e o conjunto de fases calculadas ($F_{obs} - F_{calc}$, φ_{cal}) que mostra as diferenças entre a estrutura verdadeira e a estrutura modelada. Nestes mapas, as partes existentes na estrutura, porém não incluídas no modelo, são mostradas como contornos positivos enquanto que as partes adicionadas no modelo que estão ausentes na estrutura verdadeira são representadas em contornos negativos. Na prática, utiliza-se mapas ($2F_{obs} - F_{calc}$, φ_{cal}), correspondendo a superposição de ambos os mapas (positivo e negativo), para mostrar o modelo de densidade eletrônica e os aspectos que ainda requerem correções. Como todos os dados utilizados para calcular os dois mapas (amplitudes e fases) contêm certo grau de erro, os mapas também podem conter algum nível de ruído. Em geral, um bom contorno de exibição para o mapa ($F_{obs} - F_{calc}$, φ_{cal}) é 3σ e para o mapa ($2F_{obs} - F_{calc}$, φ_{cal}) em torno de 1σ , onde σ é o rmsd para todos os pontos do mapa a partir do valor médio. A aparência dos mapas de densidade eletrônica depende mais das fases do que das amplitudes. Portanto, fases imprecisas podem introduzir erros e ruídos nos

mapas que por sua vez irão gerar um novo conjunto incerto de fases calculadas, que podem ser difíceis de eliminar no processo de refinamento iterativo.^{28,36,41} Tal problema pode ser minimizado pelo mapa de omissão (**omit map**), uma variação do mapa de diferença, em que os valores são calculados F_{calc} a partir de um modelo sem os fragmentos moleculares suspeitos, o que leva a uma representação imparcial do mapa de densidade eletrônica com regiões duvidosas omitidas.^{35,42}

Em geral, a clareza e a facilidade de interpretação destes mapas, mesmo para os calculados a partir de fases corretas, dependem da **resolução** dos dados de difração. Esta é medida em Å e define o nível de detalhe entre aspectos estruturais que podem ser distinguidos nos mapas, sendo definida no espaço recíproco como a distância mínima d_{hkl} (Seção 4.1.2) entre os planos da rede cristalina que ainda fornecem difração mensurável de raios X.^{27,28} Quanto maior a resolução (menor a distância d_{hkl}), melhor será a estrutura, porque existirão mais reflexões independentes disponíveis para defini-la.⁴³ A Figura 4.9 mostra mapas de densidade eletrônica calculados para dados truncados em vários limites de resolução.

Figura 4.9 - Mapas de densidade eletrônica (F_{obs} , φ_{calc}) calculados com diferentes limites de resolução para o fragmento LVP da lisozima (código PDB: 2VB1). 184676 reflexões foram utilizadas para o mapa de maior resolução (1Å) enquanto que apenas 2415 reflexões foram incluídas no mapa de menor resolução (3Å).

Fonte: Adaptado de WLODAWER, A., 2008, p. 6.²⁸

Estruturas de baixa resolução ($> 2,7$ Å) trazem pouca informação a respeito da posição dos átomos individuais. Para estruturas de resolução mediana ($< 2,7$ Å), cuja dimensão equivale às distâncias de ligações de hidrogênio, existe a chance de ver moléculas de água bem definidas na primeira camada de solvatação. Para resoluções maiores do que 1,5 Å, que corresponde a uma ligação covalente C-C típica em proteínas, um modelo anisotrópico de deslocamentos atômicos pode ser refinado. A localização direta de átomos de hidrogênio se torna possível em mapas com resoluções superiores a 1 Å (resolução atômica). A resolução de 0,77 Å corresponde ao limite físico definido pela radiação K_{α} de cobre (1,542 Å).^{27,28,43}

Proteínas contêm regiões com alto grau de flexibilidade, que podem causar desordens estáticas e dinâmicas em estruturas cristalinas. Desordem estática resulta de diferentes conformações adotadas por um determinado fragmento estrutural em diferentes celas unitárias, enquanto que a desordem dinâmica é fruto do aumento da mobilidade ou vibrações destes fragmentos. Como o mapa de densidade eletrônica equivale à média da distribuição de elétrons em todas as células unitárias do cristal durante o tempo total da coleta de dados, estas desordens poderão aparecer como densidade eletrônica espalhada ao longo de vários estados conformacionais dos fragmentos desordenados. Em mapas de baixa resolução, a densidade eletrônica estará sobreposta a ruídos e, portanto, tais fragmentos não serão interpretáveis.^{44,45} Um caso especial de desordem, está sempre presente na região do solvente em todos os cristais de proteínas. Algumas moléculas de água interagem com átomos da superfície da proteína por ligações de hidrogênio formando a primeira camada de solvatação, cujas posições podem ser modeladas com confiança. As moléculas de água distantes desta camada estão totalmente desordenadas, ocupando posições parcialmente preenchidas que são difíceis de modelar até mesmo em alta resolução. Esta região representa cerca de 50% do volume do cristal e não traz nenhuma informação para o modelo, sendo tratada pelo fator de correção do solvente (*bulk solvent*).^{35,46,47} A quantidade de solvente presente no cristal pode ser estimada a partir do **coeficiente de Matthews**, V_M , dado pela razão entre o volume da cela unitária e tamanho da proteína. Em geral, o valor de V_M varia entre 1,7 e 3,5 Å³/Da para a maioria dos cristais de proteína, sendo 2,15 Å³/Da o valor mais provável.⁴⁸

4.2.4 Refinamento e validação do modelo cristalográfico

Após a solução e construção do modelo inicial, parte-se para o refinamento, cujo objetivo é melhorar as fases φ_{cal} e completar o modelo cristalográfico. Nesta etapa, resíduos de aminoácidos são ajustados na densidade eletrônica observada, moléculas de substratos, ligantes, íons metálicos e água são adicionados em vários ciclos de minimização de energia intercalados com modelagens manuais.²⁸ No refinamento, todos os parâmetros do modelo (posições atômicas xyz, ocupação, fator de deslocamento atômico B , e outros) são variados até alcançar o melhor arranjo entre as amplitudes das reflexões observadas (F_{obs}) e aquelas calculadas (F_{calc}). Esta concordância é julgada pelo fator residual R ou R_{work} , definido como

$\Sigma|F_{obs} - F_{calc}| / \Sigma F_{obs}$, sendo este, um indicativo da similaridade do modelo em relação à realidade. Estruturas proteicas bem refinadas exibem em geral, um R_{work} menor do que 20%. Em cristalografia de pequenas moléculas, onde os modelos contêm menos átomos e os dados podem ser corrigidos para diversos erros sistemáticos, é comum ter valores de R_{work} entre 2 e 3 %. O fator R_{work} pode ser influenciado por fatores que convergem para mínimos. Por esta razão um fator R_{free} , livre de parâmetros que tendem a modelar erros no conjunto de dados, tem sido o mais utilizado para a validação cruzada do refinamento. O R_{free} é calculado de forma análoga ao fator R_{work} , porém utilizando um conjunto menor e exclusivo de reflexões (cerca de 1000) selecionadas aleatoriamente e que nunca entram no refinamento do modelo. A diferença $R_{free} - R_{work}$ não deve ser maior do que 7%, pois indicaria um excesso de ajuste dos dados experimentais e um modelo defeituoso.^{28,49-51}

A função de máxima verossimilhança (ML) é a mais empregada para o refinamento de proteínas. Nesta função, as fases φ_{cal} são ajustadas de forma a minimizar o fator R , embora para proteínas menores também seja usado o método de mínimos quadrados baseado nas amplitudes. A estratégia de refinamento a ser adotada é influenciada pela razão entre a quantidade de dados observados e de parâmetros refinados, cujo valor é muito pequeno para estruturas de proteínas (sobretudo na camada de maior resolução) devido ao baixo poder de difração destes cristais.^{50,52} Cada átomo do modelo é descrito por três parâmetros de posição (xyz) e pelo menos um parâmetro indicando o fator de deslocamento atômico B , quando o refinamento é isotrópico. Modelos mais sofisticados do movimento molecular requerem muito mais parâmetros no refinamento, que são obtidos pela adição de observações na forma de *restraints*, ou redução do número de parâmetros usando *constraints*. *Restraints* típicos incluem restrições estereoquímicas (comprimentos e ângulos de ligação, e outros) inseridas na função de refinamento baseadas em valores ideais encontrados em bancos de dados tais como o *Cambridge Structural Database*, CSD.⁵³ *Constraints* atuam na redução do número total de parâmetros a serem refinados, que podem ser conseguidos refinando apenas os ângulos de torção, em vez de todas as coordenadas individuais xyz , por exemplo. Outro caso muito comum ocorre quando há cópias múltiplas de uma molécula na unidade assimétrica do cristal relacionadas entre si por simetria não cristalográfica (NCS). Neste caso, considerando que existam três cópias do modelo na ASU relacionados por NCS, o número de parâmetros ajustáveis poderia ser reduzido por três.^{28,53-56}

O fator B é refinado inicialmente de maneira isotrópica para todos os átomos, mantendo elevada a razão entre dados e parâmetros, o que aumenta as chances do ajuste correto das posições xyz antes de refinar os deslocamentos. À medida que o refinamento converge, parte-se para a modelagem anisotrópica do fator B sendo este caracterizado por seis parâmetros para cada átomo, descrevendo assim, o seu deslocamento em todas as direções. Porém, esta modelagem é limitada pela resolução estrutural do modelo devido à proporcionalidade com o número de dados (reflexões $|F_{hkl}|$) disponíveis. Outra maneira de modelar o fator B envolve dividir a estrutura em fragmentos rígidos expressando o deslocamentos atômicos de maneira anisotrópica para cada grupo de fragmentos através dos parâmetros TLS (*translation, libration, screw*). O TLS pode dar uma boa aproximação da anisotropia com um número menor de parâmetros em qualquer resolução.^{53,54,57}

Os programas Refmac e Phenix-refine permitem que diferentes protocolos de minimização de energia sejam executados de maneira combinada em múltiplos passos podendo ainda ser aplicados a partes específicas do modelo tanto para o espaço direto quanto recíproco.^{49,50} Tais protocolos incluem: o refinamento de corpo rígido, onde partes do modelo são tratadas como fragmentos rígidos que serão ajustadas à densidade eletrônica disponível por rotação e translação; e o *simulated annealing* onde a estrutura é aquecida para adicionar aleatoriedade e lentamente resfriada reduzindo a probabilidade de convergência em um mínimo local de energia. Estes programas também possuem módulos para refinamento dos fatores B isotrópicos individuais ou agrupados, anisotrópicos individuais ou TLS e ainda o parâmetro de ocupação (utilizado para refinar partes da molécula com múltiplas conformações, devido a desordens estáticas ou dinâmicas). Quando integrados ao Coot e Pymol permitem uma análise visual interativa dos mapas e modelos associados.^{50,53,57-60}

Os átomos de hidrogênio são os piores dispersores de raios X devido ao seu único elétron, sendo geralmente um desafio localizá-los diretamente até mesmo em cristalografia de pequenas moléculas. Estes podem ser localizados diretamente em estruturas de resolução atômica a partir do mapa de diferença, após o cuidadoso refinamento do modelo. A localização da maior parte destes átomos pode ser calculada com boa precisão a partir das posições dos átomos mais pesados pelo método *riding model* onde restrições estereoquímicas baseadas em conhecimento prévio são utilizadas. Uma vez adicionado em posições calculadas, os átomos de

hidrogênio são refinados como corpos rígidos em relação aos seus átomos pesados, e suas coordenadas são recalculadas após cada ciclo de refinamento.^{28,50,53,54}

O processo de refinamento é repetido até que os termos estatísticos R_{work} e R_{free} se estabilizem em valores adequados, e o modelo completo tenha sentido químico e bioquímico. A qualidade do modelo final e a validade prevista para a sua interpretação, é avaliada com base em indicadores na etapa de validação, incluindo os parâmetros de qualidade da coleta de dados discutidos na Seção 4.2.1, tais como $I/\sigma(I)$, R_{merge} , R_{meas} , R_{pim} e completeza. Durante o refinamento, outros parâmetros são avaliados além do R_{work} e do R_{free} . O RMSD de ângulos e distancias, indica o quanto o modelo diverge dos parâmetros geométricos ideais. Estruturas de qualidade com alta resolução têm um RMSD de ligação ideal próximo a 0,02 Å. Valores elevados ($> 0,03$ Å) podem significar erros no modelo, e valores muito pequeno ($< 0,004$ Å) indicam que devido às rígidas restrições, o modelo foi submetido à otimização da geometria ao invés do refinamento impelido pelos dados experimentais. Os planos formados pelas ligações peptídicas estão sobre restrições estereoquímicas muito rígidas e são em geral avaliados pelo gráfico de Ramachandran, onde os ângulos torcionais φ e ψ do polipeptídeo, são plotados contra contornos ideais.^{28,53,58}

A análise do fator B também é considerada no modelo final, cujo valor em geral varia entre 2 e 100 Å² dependendo da resolução. Regiões com fatores B muito maiores em relação à vizinhança podem indicar um movimento vibracional atípico ou uma desordem não refinada. As moléculas de água adicionadas no modelo (conservada no sítio catalítico e na primeira camada de solvatação) também devem ser checadas, considerando suas possíveis interações intermoleculares, distâncias e correlação com a densidade eletrônica. A presença de moléculas de água com fatores B elevados (> 100 Å²), é um indicativo de problemas sendo em muitos casos melhor excluir esta molécula do modelo que implicará em uma mudança sutil nos parâmetros R_{work} e R_{free} . Para íons metálicos que algumas vezes são adicionados em regiões do mapa com os maiores picos de densidade eletrônica deve-se examinar, por exemplo, a geometria da esfera de coordenação.^{53,58} Para uma molécula de substrato ou um inibidor enzimático, além do fator B , também são analisadas a coerência de todas as interações intermoleculares envolvidas tais como ligações hidrogênio fortes e fracas (provenientes de sistemas aromáticos como a C-H... π , por exemplo), ligações de halogênio, interações ortogonais multipolar, entre anéis aromáticos e hidrofóbicas.^{61,62} Toda esta análise de interações compõem parte da

informação do MOB entre enzima e inibidor, que será utilizada na evolução destes inibidores à compostos com melhores propriedades farmacocinéticas e farmacodinâmicas. Uma vez que o modelo está completo e validado, este pode ser depositado em banco de dados de macromoléculas tais como o PDB.

4.3 Material e métodos: coleta de dados e elucidação estrutural

A coleta de dados cristalográficos e as elucidações estruturais que foram realizadas neste projeto se apoiaram nos métodos descritos na seção anterior, e na fundamentação teórica detalhada na Seção 4.1 deste capítulo. Dados de difração de raios X para cristais da cruzaína nativa e da mutante C25S crioprotetidos, foram coletados na linha de luz 8.3.1 do laboratório síncrotron *Advanced Light Source* (ALS) usando o programa ELVES (v.1.2) para seleção da estratégia de coleta pelo método rotativo.⁶³ Estas coletas foram realizadas na temperatura de 100 K com comprimento de onda (λ) de 1,116 Å. As reflexões foram indexadas, integradas e escalonadas nos programas iMosflm (v.7.1.1) e SCALA da interface de programas CCP4 (v.6.5.0) para os dados de melhor qualidade, e nos programas XDS e XSCALE (v.2013) para dados de qualidade e resolução inferiores.^{23,29,30} A estrutura da cruzaína de código PDB 3KKU (determinada a 1,2 Å no grupo espacial P6₅22) foi preparada com a remoção de águas, ligantes e conformações alternadas, sendo usada como modelo para cálculo das fases iniciais (φ_{cal}) por substituição molecular no programa Phaser-MR (v.1.8.4).⁴⁰ O Phenix_Refine do pacote Phenix (v.1.8.4) e o coot (v.0.8.1) foram utilizados em todas etapas do refinamento e construção do modelo tendo entre 4 e 8% das reflexões selecionadas para o cálculo de R_{free} , dependendo da quantidade de reflexões únicas observadas.^{50,59} Os monômeros da cruzaína foram inicialmente posicionados na ASU por refinamento de corpo rígido, seguido de múltiplos ciclos de refinamento de coordenadas e dos *fatores B* isotrópicos individuais, utilizando restrições de NCS para ângulos torcionais entre as subunidades. Moléculas dos ligantes foram manualmente inseridas baseado nos mapas de densidade eletrônica $F_{obs}-F_{calc}$ e $2F_{obs}-F_{calc}$. Moléculas de água foram posicionadas automaticamente para os dados de maior resolução, e posteriormente verificadas através da análise dos mapas de densidade eletrônica. O *fator B* foi refinado anisotropicamente nos últimos estágios de refinamento utilizando TLS para grupos de aminoácidos identificados pelo módulo TLSMD do Phenix_Refine.⁵⁰ A geometria da estrutura foi avaliada

utilizando o programa MolProbity (v.4.1).⁵⁸ O programa CCP4mg (v.2.10.3) e UCSF Chimera (v.1.10.1) foram utilizados para o preparo das figuras.

4.4 Resultados: elucidação estrutural e análise do MOB de inibidores

4.4.1 Análise do MOB do Neq0409 em complexo com a cruzaína nativa

Várias tentativas de co-cristalizar a cruzaína nativa com as dipeptidil-nitrilas de interesse (Neq0409, 0410 e 0412 da Figura 1.16) foram realizadas conforme descrito nas Seções 3.2.1 e 3.2.2 sendo obtido, no entanto, apenas cristais para complexos com o inibidor Neq0409 (composto prototípico). Os melhores cristais obtidos para este complexo eram frágeis e muito sensíveis à radiação, difratando a resoluções entre 2,95 Å e 3,65 Å para $I/\sigma(I) > 2$. Melhores resoluções foram obtidas para tempos maiores de exposição por imagem devido ao aumento na quantidade e intensidade das reflexões. No entanto, um maior dano causado pela radiação foi evidenciado, sobretudo, pelo aumento fora dos padrões aceitáveis do R_{merge} para as últimas imagens que quando excluídas do processamento, comprometia a completeza dos dados. Todos estes fatores associados à presença de “anéis de gelo” (*ice rings*) no padrão de difração para alguns cristais deste complexo limitaram a quantidade de conjuntos de dados úteis para o processamento.

O cristal utilizado para a caracterização do MOB difratou a 3,03 Å para $I/\sigma(I)$ igual a 2 com parâmetros de cela unitária compatíveis a sistema cristalino tetragonal com indícios do grupo espacial $P4_32_12$. Foram coletadas 230 imagens deste cristal com $\Delta\phi$ de 0,45° e dois segundos de exposição por imagem. Este cristal exibiu uma mosaicidade de 0,43°, sem a presença de anéis de gelo. Após indexação e integração inicial no XDS, foi iniciado o escalonamento no XSCALE. Inicialmente para o corte de $I/\sigma(I) = 2$, o que equivaleria a 3,03 Å de resolução, o valor de R_{merge} calculado era igual a 14,7% com 153,4% para a camada de maior resolução. Uma análise gráfica dos resultados do escalonamento indicava um grande aumento no valor de R_{merge} para as vinte últimas imagens coletadas, que foram excluídas do escalonamento implicando em redução da multiplicidade de nove para 8,2. De modo a manter a completeza elevada (99,6%) reduzindo o valor de R_{merge} (77,2% para a camada de maior resolução) o corte da resolução máxima foi então reduzido para 3,13 Å o que correspondia ao valor de $I/\sigma(I)$ igual a 2,92. Intensidades foram então

convertidas em fatores estruturais F_{hkl} e 5% das reflexões foram destinadas ao cálculo do R_{free} . As estatísticas do processamento estão disponíveis na Tabela 4.2.

Tabela 4.2 - Estatísticas relacionadas à coleta de dados e ao refinamento do modelo para a cruzaína nativa em complexo com a dipeptidil-nitrila Neq0409

Coleta e processamento de dados		Refinamento	
Grupo espacial	P4 ₃ 2 ₁ 2 (Nº 96)	Resolução (Å)	34,73 - 3,13 (3,24 - 3,13)*
Nº de mol. na ASU	5	Nº de reflexões (conj. teste)	27306 (1440)*
Cela unitária		R_{work} (%)	20,2 (34,0)*
a; b; c (Å)	137,8; 137,8; 166,5	R_{free} (%)	24,3 (37,6)*
α ; β ; γ (°)	90 90 90	Nº átomos:	
R_{merge} (%)	12,8 (77,2)*	Proteína	7960
R_{meas} (%)	11,7 (69,9)*	Ligante	23
Completeza (%)	99,7 (99,6)*	Água	0
$I/\sigma(I)$	15,9 (2,9)*	Fator B (Å ²):	
Multiplicidade	8,2 (8,5)*	Proteína	74,5
Reflec. obs. (únicas)	237244 (28790)	Ligante	83,7
R_{pim}	4,7 (51,9)*	Água	0
		RMSD: Comp. de ligação (Å)	0,005
		Ângulo de ligação (°)	0,87

* Valores para a camada de maior resolução

Fonte: Autoria própria.

Uma análise no software Xtriage não indicou patologias cristalográficas tais como geminação, anisotropia e pseudo-simetria translacional ou rotacional. O “teste L” para geminação reportou $\langle |L| \rangle = 0,51$ e $\langle |L|^2 \rangle = 0,34$, valores compatíveis para dados sem geminação, neste caso.⁶⁴ A estatística de Wilson para as intensidades indicou um fator B de 80,69 Å², um fator de escala de 0,68 e um fator de anisotropia dos dados de 0,599. Uma análise do conteúdo de solvente utilizando o XTRIAGE indicava a presença de possivelmente sete cadeias da cruzaína na ASU para coeficiente de matthews (V_M) de 2,33 correspondendo a 47,3% de conteúdo de solvente. Várias tentativas de faseamento por substituição molecular (Seção 4.2.3) com Phaser-MR, utilizando 90% de identidade na sequência primária da estrutura modelo 3KKU, foram realizadas buscando por 5, 6, ou 7 cadeias na ASU e ainda grupos espaciais enantioméricos alternativos. Todos os resultados parciais obtidos convergiam para o grupo espacial P4₃2₁2 com 5 cadeias da cruzaína na ASU relacionadas por NCS. A melhor solução encontrada pelo Phaser-MR exibiu *top LLG* de 5021,2 e *top TFZ* (*translation function Z-score*) de 44,9 com R_{work} inicial de 30,9 %. O valor de *RFZ* (*rotation function Z-score*) para a primeira cadeia (A) foi de 6,8 e para a última cadeia (E) de 3,5. Em geral, as últimas cadeias posicionadas na substituição molecular para a mesma ASU apresentam o pior ajuste (menor *RFZ*) na

densidade eletrônica. O fator de correção do solvente (*bulk solvent*) utilizado foi $k_{sol} = 0,31$ e $/ \text{\AA}^3$ e $B_{sol} = 22,2 \text{ \AA}^2$. Uma análise visual, não mostrou sobreposição entre cadeias ou presença de densidade eletrônica para uma sexta ou sétima cadeia.

O conteúdo de solvente ficou então estabelecido como 63,4 % (bem elevado em relação ao observado em cristais de proteínas), sendo esta uma das razões para a sensibilidade à radiação, formação de cristais de gelo, e ao baixo poder de difração e resolução destes cristais. Após a obtenção das fases e do modelo inicial, foi realizado o refinamento de corpo rígido seguido de refinamento de coordenadas a partir de dados do espaço recíproco e mapas do espaço real mantendo os fatores B isotrópicos para grupos de átomos. Em seguida, foram adicionadas restrições de NCS para ângulos torcionais e os fatores B isotrópicos foram refinados individualmente. Os cinco operadores de NCS entre cadeias foram encontrados automaticamente pelo programa indicando um RMSD de 0,0; 0,2; 0,25; 0,28 e 0,43 para as cadeias A, B, C, D e E em relação à cadeia A respectivamente. Restrições torcionais de NCS definem grupos de resíduos dos quais ângulos diedrais serão similares, aumentando assim, a proporção de dados / parâmetros a serem refinados (Seção 4.2.4).⁵³ Estas restrições, no entanto, não afetam as cadeias laterais de aminoácidos que são refinadas livremente permitindo que diferenças conformacionais entre cadeias relacionadas por NCS fiquem evidentes.⁵⁶

Os módulos de restrições de estrutura secundária (que adiciona restrições de distância para ligações de hidrogênio entre aminoácidos envolvidos em α -hélices e folhas- β) e restrições de modelo de referência (que usa ângulos diedrais originários de uma estrutura referência de alta resolução) do Phenix-Refine foram habilitados para aumento da razão dados / parâmetros refinados.⁵³ Quando o refinamento convergiu para R_{work} de 21,42 % e R_{free} de 25,20 %, com todos os aminoácidos e rotâmeros ajustados na densidade eletrônica, foi iniciada a construção do inibidor nas cadeias A e B. Coordenadas para o inibidor foram desenhadas no programa MarvinSketch (v.5.5.1), sendo em seguida visualizadas na rotina REEL (do Phenix) para edição da valência de ligação, hibridização, estereoquímica, carga formal e planaridade. Em seguida foi realizada uma minimização no programa ELBOW para a geração dos arquivos nas extensões cif e pdb correspondentes.

O ligante foi manualmente ajustado no mapa de diferença $F_{obs}-F_{calc}$ da cadeia A inicialmente. Após alguns ciclos de refinamento, verificou-se que o mesmo estava ligado covalentemente sendo adicionada uma restrição geométrica de distância para

o *link* covalente de 1,77 Å (valor mais frequente no CSD) com σ de 0,08 Å para esta medida, e de planaridade para os três principais átomos envolvidos nesta ligação (enxofre SG da Cys25, carbono CAL e nitrogênio NAN do ligante) com um σ de 0,02°. O mesmo foi realizado para a cadeia B que exibiu densidade eletrônica considerável para o ligante no mapa $F_{obs}-F_{calc}$. Contudo, na região do *warhead* nitrila a densidade eletrônica era fraca e pouco definida para a modelagem do inibidor. Para esta cadeia, portanto, não foi construído o inibidor para evitar o sobreajuste do modelo final. Para as cadeias C, D e E não havia densidade para o inibidor no sítio. O refinamento findou com os fatores *B* refinados anisotropicamente usando TLS para 36 grupos identificados pelo módulo TLSMD do Phenix_refine. As estatísticas do modelo final depositado no PDB com o código 4QH6 também estão na Tabela 4.2. A validação no Molprobitry encontrou 97,7% dos ângulos ϕ e Ψ em regiões favoráveis no gráfico de Ramachandran. O MOB do Neq0409 na cruzaina nativa, ortogonalmente validado por esta estrutura cristalina, é exibido na Figura 4.10.

Figura 4.10 - MOB do Neq0409 em complexo com a cruzaina nativa (código PDB: 4QH6) covalentemente ligado à cisteína catalítica (Cys25) da cadeia A com mapa de densidade eletrônica $2F_{obs}-F_{calc}$ contornados ao nível de 1σ . Os contornos positivos (verde) e negativos (vermelho) do mapa de diferença $F_{obs}-F_{calc}$ ao nível 3σ também são exibidos. Interações de hidrogênio são representadas por pontilhados de cor preta.

Fonte: Autoria própria

A caracterização do MOB do Neq0409 demonstrou que o *warhead* nitrila liga-se covalentemente com a cisteína catalítica (Cys25) e que as porções P2 e P3 do inibidor estão interagindo nos respectivos sítios S2 e S3 com o padrão de interações por ligações de hidrogênio coerente com as previsões de docagem e as hipóteses

iniciais baseadas em análogos cujas estruturas cristalográficas foram obtidas para complexos com a catepsina L de humanos. Um destes análogos é a dipeptidil-nitrila AZ12878478 sintetizada e identificada pela Astra Zeneca como inibidor da catepsina L cuja estrutura do complexo foi depositada no PDB com o código 3HHA. A sobreposição desta estrutura com o MOB do Neq0409 é exibida na Figura 4.11.

Figura 4.11 - Sobreposição do MOB do Neq0409 com o análogo estrutural ligado a catepsina L (código PDB 3HHA) em superfície de Van der Waals da cruzaina. Átomos de carbono estão coloridos em verde para o composto Neq0409 e em cinza claro para o análogo. Os principais aminoácidos da cruzaina envolvidos em interações estão nomeados.

Fonte: Autoria própria.

Além da ligação covalente com a Cys25, o Neq0409 também interage no sítio catalítico por ligações de hidrogênio com a Gly66 e a Asp161 (essenciais para o correto posicionamento do substrato natural peptídico da enzima), e ainda com o aminoácido Gln19 da tríade conforme exibido na Figura 4.10 cujas distâncias médias destas interações são apresentadas em vermelho. O resíduo hidrofóbico Leu67, que está posicionado entre os anéis aromáticos P2 e P3 do Neq0409, realiza interações de *Van der Waals* não polares do tipo π -alquil com os mesmos. Tal contato hidrofóbico também ocorre entre o anel P2 e a Ala138 no início do S2. O anel aromático P2 ainda interage por um empilhamento (*stacking*) π -amida com o grupo amida entre a cadeia principal da His162 e da Gly163. Outros contatos hidrofóbicos com a Leu160 e a Met68 no S2 colaboram para a estabilização do inibidor nesta conformação e posição no sítio catalítico. O Glu208 está orientado para a região do solvente evitando assim uma potencial interação desfavorável com o resíduo

volumoso fenilalanina da posição P2 do Neq0409, conforme exibido na Figura 1.6 do Capítulo 1. A Figura 4.11 demonstra uma sobreposição muito similar entre os análogos (RMSD de 0,781) revelando, no entanto, que o grupo fenilalanina do Neq0409 não penetra tão profundamente no sítio S2 da cruzaina em relação ao observado na catepsina L de humanos. Tal fato pode ser explicado pelo espaço maior no sítio S2 da catepsina L que possui uma alanina no lugar do Glu208 sendo este resíduo, o principal responsável pela seletividade e especificidade da cruzaina.

Apesar de existir uma densidade eletrônica significativa para a região do sítio S2 e S3 na cadeia B, a mesma era muito fraca e pouco definida na região do *warhead* nitrila (ou grupo imina quando ligado covalentemente). Tentativas de se construir o inibidor na cadeia B covalentemente ligado a Cys25 foram realizadas durante o refinamento, sendo obtidos em todas as ocasiões mapas de densidade tais como os apresentados na Figura 4.12 em que contornos negativos (vermelho) do mapa de diferença $F_{obs}-F_{calc}$ ao nível de 3σ na região do *link* covalente mostravam que esta ligação seria proibitiva para uma ocupação de 100%. Adicionalmente, o contorno positivo (verde) do mesmo mapa em região vizinha a este *link*, suporta a hipótese de que o *warhead* estaria parcialmente como nitrila ($RC\equiv N$) e parcialmente como imina ($R'RC=NH$, onde R' corresponde ao átomo de enxofre da Cys25). Tal efeito pode ser justificado pela reversibilidade do inibidor, sendo este mecanismo também comprovado nos ensaios bioquímicos realizados no NEQUIMED.

Figura 4.12 - Exibição estereoscópica da tentativa de modelar o inibidor Neq0409 na cadeia B do complexo com a cruzaina. Mapa $2F_{obs}-F_{calc}$ em azul contornado a 1σ e mapa de diferença $F_{obs}-F_{calc}$ à 3σ para contornos positivos (verde) e negativos (vermelho). Ângulo de visão similar ao da Figura 4.10 com destaque para os resíduos envolvidos em interações.

Fonte: Autoria própria

Como a densidade eletrônica observada é a média da densidade no espaço (todas as cadeias B do cristal) e no tempo de coleta de dados, existe a possibilidade

de que o inibidor esteja ligado (covalentemente) em algumas destas cadeias B e se ligando na população restante de cadeias B do cristal. O fator *B* das porções P2 e P3 do inibidor eram cerca de 30 % superiores aos fatores *B* dos aminoácidos da vizinhança sugerindo ocupação inferior a 100 % para estas porções. Esta mesma comparação quando realizada na cadeia A, indicava 100 % de ocupação até mesmo para o *warhead* na sua forma imina. Como os dados cristalográficos estão em uma baixa resolução e não exibem qualidade significativa para a modelagem deste efeito na cadeia B, esta foi deixada sem o inibidor para evitar o sobreajuste do modelo.

A reversibilidade e a consequente mobilidade do inibidor no sítio durante a cristalização, associada a uma condição de cristalização não ideal e de uma auto-proteólise residual (existente, mesmo no pH 7,5 do ensaio, devido a potência de inibição na faixa baixo μM conforme discutido na Seção 3.2.2), fizeram com que os cristais se empacotassem em um mínimo local de energia. Este empacotamento não maximizava contatos entre cadeias, levando a canais de solvente volumosos que tornavam os cristais frágeis, sensíveis à radiação e com baixo poder de difração, impactando diretamente na qualidade e resolução dos dados coletados. A qualidade destes cristais também pode ter sido afetada pelo fato de se ter usado uma mistura racêmica do inibidor durante a cristalização dado a não ciência de sua existência até aquele momento. Atualmente, toda a síntese de dipeptidil nitrilas com possibilidades de racemização passam por purificações em cromatografia de coluna quiral para separação de enantiômeros. Ensaio bioquímicos posteriores demonstraram que o enantiômero *S* do Neq0409 possui uma afinidade 14 vezes maior ($K_i = 453,0 \pm 32,2$ nM) do que o enantiômero *R* ($K_i = 6,3 \pm 0,6$ μM), o que corrobora com a estereoquímica do Neq0409 exibida na Figura 4.10 e com todos os inibidores peptídicos da cruzaina disponíveis no PDB. Apesar da baixa qualidade e resolução deste modelo cristalográfico, o MOB aqui elucidado validou com êxito o uso de dipeptidil nitrilas como esqueleto básico para a síntese de uma série de inibidores mais potentes e seletivos no grupo NEQUIMED tais como os que serão apresentados no Capítulo 5.

4.4.2 Análise do MOB do Neq0176 em complexo com a cruzaina nativa

Vários ensaios de co-cristalização com a cruzaina nativa para os inibidores do tipo fragmento (Figura 1.17) foram realizados conforme discutido na Seção 3.2.2, sendo obtidos apenas cristais de complexos com o Neq0176 na condição 0,1 M Bicine pH 9,0 e 1,6 M de sulfato de amônio. O melhor cristal da série difratou a uma

resolução máxima de 2,62 Å para $I/\sigma(I) > 2$ sendo este utilizado para a elucidação do MOB. Foram coletadas 180 imagens com $\Delta\phi$ de 0,6° e cinco segundos de exposição por imagem. O cristal exibiu uma mosaicidade de 0,39°, sem a presença de anéis de cristais de gelo ou grande decaimento causado pela radiação no padrão de difração. As reflexões foram indexadas e integradas utilizando o iMosflm e escalonadas com o Scala, indicando dados de boa qualidade sem dano causado por radiação.²⁹ A estatística da coleta e do processamento dos dados é mostrada na Tabela 4.3.

Uma análise da qualidade dos dados no XTRIAGE não reportou nenhuma irregularidade ou patologia cristalográfica, sugerindo a existência de sete moléculas do monômero por ASU com 47,1% de conteúdo de solvente e V_M de 2,32, para o grupo espacial $P4_32_12$ (que contém oito cópias da ASU dentro da cela unitária). A substituição molecular com o Phaser-MR encontrou, como melhor solução, cinco monômeros da cruzaina na ASU tendo um valor máximo LLG de 7598, TFZ de 71 e um fator R_{work} inicial de 30,2%, considerados apropriados para esta faixa de resolução. Em uma inspeção visual foi observado a presença de densidade eletrônica para uma sexta cadeia. Porém esta era muito fraca e desconexa não sendo encontrada em nenhuma das tentativas com Phaser, mesmo após o uso do modelo final refinado (cuja densidade eletrônica na região da sexta cadeia exibia leve melhora) como solução parcial em um novo ciclo de substituição molecular. Considerando a existência desta sexta cadeia na ASU que não foi possível de ser modelada, o conteúdo de solvente deste cristal se alteraria dos atuais 62,2% para 54,7%, sendo este valor ainda considerado elevado para cristais de proteínas.⁶⁵

Tabela 4.3 - Estatísticas relacionadas à coleta de dados e ao refinamento do modelo para cruzaina nativa em complexo com o inibidor reversível Neq0176

Coleta e processamento de dados		Refinamento	
Grupo espacial	$P4_32_12$ (Nº 96)	Resolução (Å)	39,23 - 2,62 (2,71 - 2,62)*
Nº de mol. na ASU	5	Nº de reflexões (conj. teste)	48282 (1954)*
Cela unitária		R_{work} / R_{free} (%)	18,9 / 23,8
a; b; c (Å)	138,7; 138,7; 163,5	Nº átomos: Proteína	7960
α ; β ; γ (°)	90 90 90	Ligante	54
R_{merge} (%)	2,6 (33,5)*	Água	109
R_{meas} (%)	3,7 (49,9)*	Fator B (Å²): Proteína	33,5
Completeza (%)	99,5 (98,4)*	Ligante	74,0
$I/\sigma(I)$	15,9 (2,1)*	Água	26,3
Multiplicidade	7,6 (7,0)*	RMSD: Comp. de ligação (Å)	0,008
Reflec. obs. (únicas)	369873 (48351)	Ângulo de ligação (°)	1,08

* Valores para a camada de maior resolução

Fonte: Autoria própria.

Os programas *Phenix_Refine* e Coot foram utilizados em todas as etapas do refinamento, que seguiu o procedimento descrito na Seção 4.3. 4% das reflexões foram destinadas ao cálculo de R_{free} . Alguns ciclos de *simulated annealing* foram aplicados para melhor ajuste dos aminoácidos da cadeia principal que não estavam bem definidos, sobretudo em algumas regiões de alças e *loops*. Após alguns ciclos de refinamento com NCS torsional, foi observada densidade eletrônica pouco intensa nos mapas de diferença $F_{obs}-F_{calc}$ para os sítios catalíticos das cadeias A, B e C. Após construção e correto ajuste do inibidor nestas cadeias, foram adicionadas moléculas de água para a primeira camada de solvatação, sendo estas, analisadas com base no padrão de interações e na correlação com os mapas de densidade. Por fim, TLS foi utilizado para refinar os fatores B anisotropicamente. A geometria do modelo final foi avaliado utilizando o *MolProbity* indicando todos os valores dentro dos padrões aceitáveis para esta faixa de resolução. As coordenadas e os fatores estrutura foram depositados no PDB com o código de acesso 4KLB.

O MOB para o fragmento Neq0176 está exibido na Figura 4.13. Três variantes conformacionais deste MOB foram observadas para as cadeias A, B e C, sendo vários os fatores que poderiam ter levado a este efeito. O sítio S3 é superficial, grande e exposto ao solvente. Além disto, contém poucos pontos para interações por ligações de hidrogênio ou contatos hidrofóbicos, que somados aos cinco graus de liberdade rotacionais do pequeno Neq0176, permitem uma grande mobilidade do inibidor nesta região. Apesar disto, o padrão de interação para o grupo 2-acetamido tiofeno-3-carboxamida do Neq0176 se manteve similar nas três conformações (Figura 4.13D), sendo este, o principal responsável pelo seu alto valor de IC₅₀ e LE.⁶⁶

Este inibidor do tipo fragmento foi obtido pela simplificação molecular do Neq0042, sendo este, o de maior atividade tripanocida dentre uma série de inibidores reversíveis não peptídicos que possuíam o Neq0176 como esqueletos base. As docagens realizadas para o Neq0176 no Nequimed (tese de doutorado do Dr. Helton Wiggers) prediziam em suas poses que o anel 1,2,4-triazol do Neq0176 interagiria na base do sítio S2, sendo este, um potencial candidato a capturar interações importantes para a seletividade de novos inibidores. No entanto, com a validação experimental do MOB via cristalografia, foi observado que este grupo triazol se posicionava na base do sítio S3 interagindo principalmente com o resíduo Ser61 por ligação de hidrogênio nas cadeias B e C (Figura 4.14B e C), otimizando contatos hidrofóbicos com a superfície do sítio S3. Para a variante conformacional

da cadeia A, o anel triazol está voltado para a região do solvente realizando apenas contatos hidrofóbicos com a Leu67 tal como exibido na Figura 4.14A.

Figura 4.13 - Em **A**, **B** e **C**, são exibidos os modos de interação do fragmento Neq0176 nas cadeias A, B e C do complexo com a cruzaina nativa. O mapa $2F_{obs}-F_{calc}$ contornado a $0,7\sigma$ é exibido em ciano, e o mapa de diferença $F_{obs}-F_{calc}$ é exibido à 3σ para contornos positivos (verde) e negativos (vermelho). Pontilhados pretos representam as ligações de hidrogênio. As três variantes conformacionais estão sobrepostas em superfície de Van der Waals em **D**, com destaque para os principais aminoácidos envolvidos em interações.

Fonte: Autoria própria.

O grupo 2-acetamidotiofeno-3-carboxamida interage por ligação de hidrogênio com a Gly66 e a Asp161, que são importantes aminoácidos para o reconhecimento e o correto posicionamento do substrato peptídico no sítio. Uma interação fraca, porém significativa é a que ocorre com a Gly65 nas cadeias B e C através de uma ligação de hidrogênio não clássica. Esta interação é substituída por outra mais efetiva com a Cys25 na cadeia A (que no pH 9 de cristalização teria o átomo de enxofre protonado) devido a uma leve rotação do grupo acetamido em relação ao

anel tiofeno. Contatos hidrofóbicos também são realizados pelo anel tiofeno com os aminoácidos Leu67, Met68, Leu160 e Ala138 que delimitam a entrada do sítio S2. Todas estas interações somadas ao mecanismo não covalente observado (átomo de enxofre da Cys25 distante a pelo menos 4,2 Å do centro nucleofílico do grupo acetamido), fazem deste grupo químico uma atraente alternativa aos atuais grupos peptídeo-miméticos encontrados nos inibidores da cruzaina. A elucidação estrutural do MOB do Neq0176 foi um importante resultado para a validação ortogonal do mecanismo não covalente de uma série de inibidores não peptídicos da cruzaina análogos ao Neq0042 sendo este resultado publicado na PlosNTD (Anexo I).⁶⁶

Figura 4.14 - Análise detalhada do MOB obtido para as três variantes conformacionais do Neq0176 na cruzaina nativa para as cadeias A, B e C. Ligações de hidrogênio são representadas por pontilhados de cor verde e cinza (para não clássicas). Empilhamentos π -amida e outros contatos hidrofóbicos são também exibidos com suas respectivas distâncias.

Fonte: Autoria própria.

Uma análise do mapa de diferença $F_o - F_c$ apresentado em verde na Figura 4.13B e C, mostra grande densidade eletrônica na região da Cys25. Inicialmente foi pensado que pudesse ser alguma possível desordem estática deste aminoácido, que não se confirmou. Além disto, os fatores B para os inibidores estavam até duas vezes e meia maiores do que os observados para a proteína, indicando baixa ocupação do inibidor no sítio, sendo este o motivo pelo qual o mapa $2F_o - F_c$ foi exibido no nível $0,7\sigma$ na Figura 4.13. Em uma recente reanálise destes dados, foi encontrada uma molécula de MMTS covalentemente ligada com a Cys25 para as cadeias B e C. O resultado desta nova modelagem na região do sítio está exibida na Figura 4.15. O MMTS, era geralmente utilizado para conter à auto-proteólise durante a purificação, sendo substituído pelo inibidor de interesse em etapa anterior ao início da cristalização. O MMTS parcialmente ligado a cruzaina e o pH de cristalização

elevado, portanto, mantiveram a enzima estável e homogênea durante o processo de cristalização sendo os responsáveis pelo crescimento deste cristal.

Figura 4.15 - Remodelagem da estrutura cristalográfica que descreve o MOB do Neq0176 considerando a presença do MMTS parcialmente ligado à Cys25 para as cadeias B e C do modelo. MOB exibido para a cadeia B com mapas de densidade $2Fo-Fc$ contornados à 1σ em ciano e mapas de diferença $Fo-Fc$ contornados a 3σ em verde a vermelho. O refinamento convergiu em 70% de ocupação para o Neq0176 e 30% para o MMTS.

Fonte: Autoria própria.

O MMTS covalentemente ligado foi refinado nas cadeias B e C com 30% e 34% de ocupação respectivamente. O Neq0176 foi refinado nestas mesmas cadeias com 70% e 66% de ocupação. A média dos fatores B dos ligantes diminuiu de $74 \text{ \AA}^2$ para $46 \text{ \AA}^2$, que mesmo sendo cerca de uma vez e meia maior do que os fatores B da proteína ($33 \text{ \AA}^2$), levou a uma melhora significativa no mapa de densidade $2Fo-Fc$ tal como apresentado na Figura 4.15, desta vez contornado ao nível 1σ . O padrão de interações em relação ao exibido no modelo anterior sem MMTS (Figura 4.14), praticamente se manteve inalterado tendo apenas sutis variações no comprimento das ligações de hidrogênio de no máximo $0,06 \text{ \AA}$. A primeira vista poderia se pensar que a grande diferença conformacional observada para o Neq0176 da cadeia A em relação às cadeias B e C seria função da presença parcial do MMTS nestas cadeias. No entanto, isto seria pouco provável uma vez que o MMTS foi refinado como uma desordem estática, implicando em baixa possibilidade de haver coexistência com as moléculas do Neq0176 no sítio. Estas variantes conformacionais podem ser melhor compreendidas pela análise do empacotamento cristalino exibida na Figura 4.16.

Os sítios catalíticos das cadeias B e C, embora estejam em uma região de canal de solvente, estão mais “blindados” por moléculas vizinhas da cruzaina. Já os

sítios das cadeias D e E são bem menos acessíveis a estes canais de solvente. Por outro lado, o sítio da cadeia A está voltado para a região onde se encontra o maior do canal de solvente, mesmo considerando que exista uma sexta cadeia da cruzaina nesta região, sendo este, um dos possíveis motivos pelo qual o inibidor tenha cristalizado nesta conformação. O inibidor por ser reversível, teria maior liberdade para entrar e sair do sítio, justificando também a baixa ocupação e os elevados fatores B. Na Figura 4.16, nota-se que o inibidor na cadeia B está em uma região que interage com a cadeia E, sendo esta, menos acessível ao solvente. Na Figura 4.13D verifica-se que todos os aminoácidos envolvidos em interações se encontram praticamente nas mesmas conformações para estas três cadeias. Logo, as diferenças no MOB observadas sejam, provavelmente, decorrentes da proximidade e acessibilidade do sítio catalítico de cada cadeia aos canais de solvente.

Figura 4.16 - Em **A**, representação em superfície de Van der Waals das cinco cadeias da cruzaina relacionadas por NCS existentes na ASU com destaque para a localização dos inibidores nas cadeias B e C. Este mesmo destaque é dado em **B** para o cristal expandido 150 Å em todos os eixos, mostrando que os sítios catalíticos destas cadeias estão em regiões menos acessíveis aos canais de solvente.

Figura: Autoria própria

Assim, como o sítio catalítico da cadeia A está em uma região próxima ao maior canal de solvente, é possível que o inibidor ainda estivesse solvatado por moléculas de água, sobretudo ao redor do anel triazol, que não foram identificadas devido à média resolução desta estrutura. Uma análise no MOLGUL revela uma maior preferência para ângulos diedrais que mantenham os grupos acetamido e tiofeno no mesmo plano, embora variações de até 40° em ambas as direções também tenham sido observadas. É provável que esta seja a conformação mais

frequente desta região do inibidor quando em solução. Porém, durante o *binding* no sítio, um ganho energético muito maior possivelmente ocorra quando aconteça uma leve torção neste ângulo diedral, fazendo com que o grupo acetamido interaja por ligações de hidrogênio com a Gly65, Gly66, Asp161. Como o sítio da cadeia A está em uma região mais acessível ao solvente, provavelmente o grupo triazol ainda esteja solvatado e o grupo acetamido ainda esteja quase planar, fazendo com que a molécula entre mais profundamente no sítio e realize melhores interações nesta região (Figura 4.14A). Como consequência, o anel triazol se distancia da Ser61 e do sítio S3, passando a interagir apenas por contatos hidrofóbicos com a Leu67 e por ligações de hidrogênio com as moléculas de água que o solvata. O mesmo não ocorre para as outras cadeias, que estão menos acessíveis ao canal de solvente, fazendo com que o inibidor tenha interações otimizadas quando interaja no sítio S3. Estas variantes do MOB são, portanto, causa ou consequência do empacotamento das cinco cadeias da cruzaina na ASU e estudos envolvendo dinâmica molecular têm sido executados no Nequimed para melhor compreensão destes resultados.

4.4.3 Análise do MOB do Neq0147 em complexo com a cruzaina C25S mutante

Tentativas de explorar interações específicas com o Glu208 foram feitas no Nequimed em um projeto de mestrado que envolvia a seleção, via métodos virtuais, de fragmentos planejados para o sítio S2. Na ocasião o Neq0147 e o Neq0220 (Figura 1.17) foram identificados como inibidores da cruzaina. As predições iniciais do MOB via docagem sugeriam que o grupo carboxila tanto do Neq0147 quanto do Neq220 interagiria com o Glu208 via ligações de hidrogênio ou ponte salina (dependendo do pH) sendo um bom ponto de partida para busca por seletividade em relação a catepsina L, que apresenta uma alanina (Ala214) na mesma região. Com o objetivo de validar o MOB destes fragmentos para posterior evolução a inibidores mais potentes e seletivos, várias tentativas de cristalização foram realizadas principalmente na cruzaina C25S mutante surgindo apenas cristais para o complexo com o Neq0147 equilibrados em 0,1 M Hepes (pH 7,5) e 1,2 M malonato de sódio.

O melhor cristal deste complexo difratou a uma resolução de 2,72 Å para $I/\sigma(I)$ igual a dois. Para este cristal, foram coletadas 220 imagens com 3 segundos de exposição por imagem e $\Delta\phi$ de 0,5°. As reflexões foram indexadas e integradas com o iMosflm e escalonadas com o Scala. Uma análise das reflexões integradas no programa Pointless indicava maior probabilidade para o grupo espacial $P4_32_12$, o

mesmo observado para os cristais dos outros dois complexos aqui expostos. Uma análise no Xtriage não reportou nenhuma patologia, indicando dados de qualidade muito superior aos obtidos para os cristais com o Neq0409 e Neq0176, embora estes cristais também fossem sensíveis ao dano causado pela radiação indicando, portanto, um grande conteúdo de solvente. Uma análise no Xtriage do coeficiente de Matthews indicou seis cópias da cruzaina na ASU com Vm de 2,56 totalizando 52% de conteúdo de solvente. Estas seis cópias foram encontradas na substituição molecular com o Phaser-MR que reportou top LLG de 8237 e Top TFZ de 55,4 com valores R_{work} iniciais de 33%. Estrutura foi refinada seguindo o protocolo descrito na Seção 4.3 empregando restrições de NCS e de estrutura secundária. Os fatores B anisotrópicos foram refinados com TLS. Uma forte densidade eletrônica estava disponível para o ligante nas cadeias A, B, C e D. Após a construção do Neq0147 nestas cadeias, a densidade eletrônica também se intensificou para as cadeias E e F, onde o ligante também foi posteriormente construído. Moléculas de água foram adicionadas para a primeira camada de solvatação e manualmente checadas para a correlação com os mapas de densidade eletrônica e padrão de interações. Uma análise no Molprobit indicou 97,4% dos ângulos ϕ e ψ em regiões favoráveis gráfico de Ramachandran. As estatísticas de coleta e refinamento estão disponíveis na Tabela 4.4. Esta estrutura foi depositada no PDB com o código 4KNO.

Tabela 4.4 - Estatísticas relacionadas à coleta de dados e ao refinamento do modelo para cruzaina C25S em complexo com o fragmento Neq0147

Coleta e processamento de dados		Refinamento	
Grupo espacial	P4 ₃ 2 ₁ 2 (Nº 96)	Resolução (Å)	48,50 - 2,72 (2,79 - 2,72)*
Nº de mol. na ASU	6	Nº de reflexões (conj. teste)	40282 (1954)*
Cela unitária		R_{work} / R_{free} (%)	19,5 / 23,2
a; b; c (Å)	133,1; 133,1; 167,4	Nº átomos: Proteína	9552
α ; β ; γ (°)	90 90 90	Ligante	259
R_{merge} (%)	6,4 (82,3)*	Água	93
R_{meas} (%)	8,7 (87,9)*	Fator B (Å ²): Proteína	51,5
Completeza (%)	99,1(95,8)*	Ligante	74,8
$I/\sigma(I)$	17,5 (2,0)*	Água	38,9
Multiplicidade	8,7 (8,2)*	RMSD: Comp. de ligação (Å)	0,009
Reflec. obs. (únicas)	354341 (40771)	Ângulo de ligação (°)	1,42

* Valores para a camada de maior resolução

Fonte: Autoria própria.

Importantes aspectos estruturais foram obtidos com a elucidação estrutural do MOB do Neq0147. Este se ligou em uma orientação oposta ao que foi inicialmente

predito, demonstrando que o carboxilato interagia por três ligações de hidrogênio com resíduos da tríade catalítica “ancorando” o inibidor nesta região por um mecanismo não covalente de inibição tal como o exibido na Figura 4.17. Outro detalhe interessante obtido pelo MOB do Neq0147 foi a posição em que se encontrava o grupo oxadiazol. Este interagia por ligação de hidrogênio com a Gly66. Porém, com as devidas otimizações poderia também interagir com a Asp161 e a Gly65 atuando como um bioisóstero da ligação amídica presente nos substratos naturais desta enzima. Estes “azóis” seriam, portanto, atraentes alternativas aos tradicionais inibidores peptídeo-miméticos cujo desenvolvimento tem sido evitado.

Figura 4.17 - Exibição estereoscópica do MOB do fragmento Neq0147 em complexo com a cruzaina C25S mutante obtido para a cadeia A. Mapas de densidade eletrônica $2F_{obs}-F_{calc}$ contornados à 1σ em ciano e mapa de diferença $F_{obs}-F_{calc}$ à 3σ para contornos positivos (verde) e negativos (vermelho). As ligações de hidrogênio são apresentadas em pontilhados pretos com destaque para os principais resíduos de aminoácidos envolvidos em interações.

Fonte: Autoria própria

O inibidor ocupa a região entre a tríade catalítica e o sítio S2 para as cadeias A, B, C e D, com um RMSD de 0,08 Å entre as diferentes conformações (Figura 4.18A até D). O Neq0147 é estabilizado nesta posição por ligações de hidrogênio entre o anel oxadiazol e a Gly66, e também entre o íon carboxilato e os aminoácidos da tríade His162, Gln19 e cadeia principal da Ser25. O grupo clorofenil do inibidor realiza contatos hidrofóbicos cujo padrão de interações varia para cada cadeia conforme exibido na Figura 4.18. O átomo de cloro nas cadeias A, B e D indica a possibilidade de interação de halogênio nesta posição com a Met68 e cadeia principal da His162, ao contrário do observado para a cadeia C, onde este mesmo átomo está apontando para a região do solvente, o que possibilita melhor ajuste hidrofóbico do clorofenil na parte mais profunda do sítio S2. O MOB do Neq0147 nas cadeias E e F do complexo com a C25S foi influenciado pelo empacotamento

cristalino. Estas cadeias cristalizaram com seus sítios catalíticos relacionados por um eixo de simetria de ordem dois conforme exibido na Figura 4.19A e B, fazendo com que o inibidor assumisse diferente orientação nos sítios destas cadeias.

Figura 4.18 - Variantes conformacionais do MOB do Neq0147 obtidos nas cadeias B, C e D do complexo com a cruzaina C25S mutante com os respectivos mapas $2F_{obs}-F_{calc}$ contornados à 1σ (em ciano) e $F_{obs}-F_{calc}$ contornado à 3σ (em verde e vermelho) exibidos em **A**. O padrão de interações de cada uma destas variantes é exibido em **B**, **C** e **D** com destaque para as ligações de hidrogênio (pontilhados em verde), e contatos hidrofóbicos.

Fonte: Autoria própria

Poucos contatos entre as cadeias da cruzaina C25S foram privilegiados nesta estrutura tal como exibido na Figura 4.19C. O Neq0147 possui apenas três graus de liberdade rotacionais e cristalizou em um sítio mais profundo e mais definido para interações. Isto mantém o inibidor menos flexível nesta região, diferentemente do MOB observado para o Neq0176 (Seção 4.4.2) cujos 5 graus de liberdade rotacionais somado ao sítio S3 superficial, levaram a variantes conformacionais mais distintas. Esta mobilidade dos fragmentos em sítios superficiais pode ter contribuído para a cristalização de várias cadeias na ASU, levando a cristais com alto conteúdo de solvente e baixo poder de difração. Cada variante existe devido aos diferentes arranjos entre cadeias na ASU, que levam a sutis mudanças conformacionais no

sítio (Figura 4.20B) e a cristalização em diferentes estados energéticos. As variantes conformacionais do Neq0147 estão sobrepostas na Figura 4.20.

Figura 4.19 - A influência do empacotamento cristalino no MOB do Neq0147 das cadeias E e F do complexo com a C25S. Estas cadeias cristalizaram com seus sítios catalíticos relacionados por um eixo de simetria de ordem dois conforme exibido em **A** e aumentado em **B**. O conteúdo da ASU é exibido em **C** com cadeias relacionadas por NCS com destaque para a posição dos sítios catalíticos destas cadeias.

Fonte: Autoria própria.

Figura 4.20 - Em **A**, sobreposição das variantes conformacionais das cadeias A (verde), B (coral), C (azul), D (ciano), E (dourado) e F (rosa) em superfície de *Van der Waals*. Em **B**, os principais aminoácidos do sítio e suas mudanças conformacionais são destacadas.

Fonte: Autoria própria

A elucidação do MOB do Neq0147 mostrou uma orientação do inibidor no sítio catalítico em sentido contrário em relação às poses de docagem que levaram a sua seleção e posterior identificação como inibidor da cruzaina. Isto pode ser explicado pelo fato de ter sido utilizado um ligante da catepsina L para a minização da enzima em etapa anterior à docagem, cuja intenção era explorar uma sub-região formada na base do sítio quando o Glu208 estivesse voltado para o solvente. Quando a docagem era realizada nas mesmas condições, porém sem este artifício, foi observada grande similaridade para a pose melhor ranqueada em relação ao MOB elucidado (RMSD de 0,12 Å) conforme sobreposição exibida na Figura 4.21, cuja única diferença significativa ocorre na conformação do anel clorofenil que está rotacionado cerca de 180°. Esta estrutura cristalográfica validou ortogonalmente o MOB para o Neq0147, sendo este essencial para a identificação e síntese de uma série de potentes inibidores que serão apresentados no Capítulo 5 deste trabalho.

Figura 4.21 - Sobreposição do MOB do Neq0147 na cadeia B da C25S (em magneta) com a pose de docagem (em verde).

Fonte: Autoria própria

4.4.4 - Validação da cruzaina C25S como modelo cristalográfico.

De modo a verificar a correlação do MOB de inibidores entre a cruzaina nativa e a C25S, e validar esta como um modelo cristalográfico para estudo com inibidores de baixa afinidade, o inibidor mais avançado da cruzaina, o K777, também foi co-cristalizado na C25S. Para este complexo foram coletadas 140 imagens de difração com dois segundos de exposição cada e $\Delta\phi$ de 0,8° exibindo 0,39° de mosaicidade. A análise do coeficiente de Matthews reportou sete cópias da cruzaina na ASU com um conteúdo de solvente de 47% e V_m de 2,32. A substituição molecular com Phaser-MR posicionou corretamente apenas cinco cópias da cruzaina na ASU, com conteúdo final de solvente de 62,1%. O empacotamento obtido era similar ao

observado para o complexo com o Neq0176, não existindo densidade eletrônica para uma sexta ou sétima cadeia. Densidade eletrônica para o ligante foi observada nas cadeias A e B. A Tabela 4.5 apresenta as estatísticas da coleta e refinamento.

O K777 é estabilizado no sítio de ligação da cruzaina C25S principalmente por ligações de hidrogênio com os aminoácidos Gln19, Gly66, Asp161 e His162 tal como exibido na Figura 4.22A. O MOB do K777 na C25S apresentou padrão de interações muito similares às observadas para a cruzaina nativa na estrutura 2OZ2, conforme sobreposição exibida na Figura 4.22B. As diferenças no MOB ocorrem devido à falta da ligação covalente entre a vinilsulfona e a Ser25, fazendo com que a porção P1' do inibidor esteja rotacionada em torno de 90° deixando de interagir de maneira eficiente no sítio S1' (principalmente por empilhamento π - π com o Trp184) passando a interagir por ligações de hidrogênio com a His162 e a Gln19 e também por empilhamento π - π com o anel aromático da porção P1. Diferentes conformações também são observadas em P1 e P3, sendo estas, consequência da liberdade rotacional destas porções nestes sítios grandes e superficiais. Não foram notadas diferenças conformacionais na região P2, e na porção peptídica do inibidor, que interagem com aminoácidos de reconhecimento da enzima em um sítio mais profundo e definido. O K777 apresentou elevados fatores B e baixa ocupação no sítio da C25S indicando que mesmo este potente inibidor, que apresenta uma das melhores complementaridades com o sítio da cruzaina tanto em forma quanto em padrão de interações, não foi capaz de influenciar a obtenção de cristais de qualidade quando este não está covalentemente ligado a cisteína catalítica.

Tabela 4.5 - Estatísticas da coleta de dados e de refinamento do modelo para a cruzaina C25S mutante em complexo com o K777

Coleta e processamento de dados		Refinamento	
Grupo espacial	P4 ₃ 2 ₁ 2 (Nº 96)	Resolução (Å)	38,5 - 2,96 (3,06 - 2,96)*
Nº de mol. na ASU	5	Nº de reflexões (conj. teste)	33807 (1350)*
Cela unitária		R_{work} / R_{free} (%)	21,9 / 25,2
$a; b; c$ (Å)	137,7; 137,7; 165,6	Nº átomos: Proteína	7960
$\alpha; \beta; \gamma$ (°)	90 90 90	Ligante	82
R_{merge} (%)	10,9 (82,8)*	Água	0
R_{meas} (%)	11,6 (110,9)*	Fator B (Å ²): Proteína	41,2
Completeza (%)	99,9(99,9)*	Ligante	67,7
$I/\sigma(I)$	17,5 (2,0)*	Água	0,0
Multiplicidade	9,0 (9,2)*	RMSD: Comp. de ligação (Å)	0,005
Reflec. obs. (únicas)	303649 (33807)	Ângulo de ligação (°)	1,02

* Valores para a camada de maior resolução

Fonte: Autoria própria.

Figura 4.22 - MOB do inibidor K777 na cruzaína C25S com mapa de densidade eletrônica $2F_{obs}-F_{calc}$ contornados à 1σ (em azul) e mapa $2F_{obs}-F_{calc}$ à 3σ (verde e vermelho) e ligações de hidrogênio representadas por pontilhados pretos. Em **B**, o MOB do K777 na cruzaína nativa (código PDB 2OZ2) está sobreposto com o MOB obtido na cadeia A da C25S.

Fonte: Autoria própria

Mesmo os cristais obtidos para o potente K777 em complexo com a C25S exibiram múltiplas cadeias da cruzaína na ASU, alto conteúdo de solvente e baixa resolução. Não é incomum proteínas cristalizarem com mais de uma molécula na ASU. Neste caso patologias cristalográficas como geminação, pseudosimetria translacional e rotacional poderiam estar atuando. Situações de pseudosimetria surgem quando operadores de simetria não cristalográficos (NCS) estão próximos da verdadeira simetria cristalográfica. Cerca de 6% de estruturas depositadas no PDB exibem pseudosimetria.⁶⁷ Do ponto de vista da simetria, a cristalização de proteínas pode ser entendida como a organização de moléculas de proteína de acordo com os operadores de simetrias cristalográficos e não cristalográficos (que não obedecem às imposições da rede cristalina) nas celas unitárias que se repetem para formar a rede cristalina. Esta rede poderá, em alguns casos, apresentar defeitos (patologias cristalográficas) que frequentemente geram cristais com alto conteúdo de solvente.⁶⁸ Estes cristais geralmente exibem menor poder de difração e resolução de acordo com o grau de desordem, devido à maior difração difusa que caracteriza ruído proveniente da região do solvente e da fraca estabilização das moléculas de proteínas por interações intermoleculares.⁴³

Existia a preocupação de que a cruzaína C25S agisse como uma serino protease mantendo alguma atividade proteolítica. No entanto, uma análise estrutural indicou que a distância entre a serina (Ser25) e a histidina (His162) catalítica desta construção era pelo menos 0,7 Å maior do que a observada para outras conhecidas

serino proteases disponíveis no PDB. Mesmo apresentando toda a maquinaria para atuar como uma serino protease, esta seria limitada devido a maior distância entre os constituintes da tríade. Alguns cristais da forma apo da C25S foram crescidos no tampão 0,1M Hepes (pH 7,5), 4,3M de cloreto de sódio, 0,2M de sulfato de amônio e 25% (peso / volume) de PEG 3000. No entanto, apenas um único conjunto de dados com 600 imagens, tempo de exposição de 5 segundos por imagem e $\Delta\phi$ de 0,5° foi coletado. Este cristal exibiu mosaicidade de 0,29° difratando a 2,47 Å para $I/\sigma(I)$ igual a dois. O processamento com o Mosflm e Scala indicou um R_{merge} de 8,1% sendo 116% para a camada de maior resolução e grupo espacial $P4_32_12$ como o mais provável exibindo de cela unitária com dimensões 70,0 Å; 70,0 Å; 182,4 Å; 90°, 90° e 90°. A análise de Matthews indicou a presença de duas cópias da cruzaina na ASU com conteúdo de solvente de 46,8% e V_m de 2,31. A análise destes dados no Xtriage indicou que estes eram anisotrópicos com fatores de escala diferindo por um fator maior que dois em uma ou mais direções, apresentando ainda fortes indícios de geminação pela análise das intensidades (teste L). O Phaser-MR não encontrou soluções após várias tentativas. Espera-se que com a constante evolução dos métodos e programas cristalográficos, a geminação identificada nestes dados possa ser futuramente tratada. O processamento destes dados, no entanto, revelou que estas condições de cristalização poderiam levar a cristais da cruzaina na forma Apo de qualidade (com apenas duas cópias da cruzaina na ASU) que poderiam ser otimizados e posteriormente utilizados para experimentos de *soaking* com os inibidores reversíveis. O MOB destes inibidores na C25S poderia ser perfeitamente transladado para a cruzaina nativa, embora diferentes conformações nas regiões P1', P1 e P3 possam ocorrer devido à superficialidade dos respectivos sítios.

4.5 Referências Bibliográficas

- 1 Chatterjee, S. K. **Crystallography and the World of Symmetry**. 1 ed. USA: Springer, 2008. v. 1. 152 p.
- 2 Hahn, T. **International tables for crystallography: space group symmetry**. 5 ed. USA: Springer, 2005. v. A. 905 p.
- 3 Tilley, R. J. D. **Crystals and Crystal Structures** ed. England: John Wiley & Sons Ltd, 2006. v. 1. 270 p.
- 4 Drenth, J. **Principles of Proteins X-Ray Crystallography**. 3 ed. USA: Springer Verlag, 2007. v. 1. 332 p.
- 5 Giacovazzo, C.; Monaco, H. L.; Viterbo, D. **Fundamentals of Crystallography**. 2 ed. Oxford: IUCR - Oxford University Press, 2002. v. 1. 872 p.

- 6 Hammond, C. **The Basics of Crystallography and Diffraction**. 3 ed. Oxford: IUCr - Oxford University Press, 2009. v. 1. 264 p.
- 7 Newnham, R. E. **Properties of Materials: Anisotropy, Symmetry, Structure**. ed. New York, United States: Oxford University Press Inc., 2005. v. 1. 391 p.
- 8 Herbstein, F. H. **Crystalline Molecular Complexes and Compounds** ed. New York, United States: Oxford University Press Inc., 2005. v. 1. 1310 p.
- 9 Mcpherson, A. **Introduction to Macromolecular Crystallography**. 2° ed. New Jersey, United States: John Wiley & Sons, Inc., 2009. v. 1. 277 p.
- 10 Glusker, J. P.; Trueblood, K. N. **Crystal Structure Analysis: a primer**. 3 ed. New York, United States: Oxford University Press Inc., 2010. v. 1. 299 p.
- 11 Fernandes, W. B.; Napolitano, H. B.; Perez, C.; Lariucci, C. Aplicações tecnológicas da metodologia cristalográfica. **Revista Processos Químicos**, v. 7, n. 1, p., 2010.
- 12 Clegg, W. **Crystal Structure Analysis: Principle and Practices**. 2 ed. New York: IUCr - Oxford University Press, 2009. v. 1. 282 p.
- 13 Messerschmidt, A. **X-ray Crystallography of Biomacromolecules** 1ed. USA: WILEY-VCH, 2007. v. 1. 318 p.
- 14 Dauter, Z. Efficient use of synchrotron radiation for macromolecular diffraction data collection. **Progress in Biophysics and Molecular Biology**, v. 89, n. 1, p. 153-172, 2005.
- 15 Dauter, Z.; Wlodawer, A. Impact of synchrotron radiation on macromolecular crystallography. **Journal of Synchrotron Radiation**, v. 17, p. 433-444, 2010.
- 16 Skarzynski, T. Collecting data in the home laboratory: evolution of X-ray sources, detectors and working practices. **Acta Cryst D**, v. 69, p. 1283-1288, 2013.
- 17 Ravelli, R. B. G.; Garman, E. F. Radiation damage in macromolecular cryocrystallography. **Current Opinion in Structural Biology**, v. 16, p. 624-629, 2006.
- 18 Ravelli, R. B. G.; Mcsweeney, S. M. M. The 'fingerprint' that X-rays can leave on structures. **Structure**, v. 8, n. 3, p. 315-328, 2000.
- 19 Weik, M.; Ravelli, R. B. G.; Kryger, K.; Mcsweeney, S.; Raves, M. L.; Harel, M.; Gros, P.; Kroon, J.; Sussman, J. L. Specific chemical and structural damage to proteins produced by synchrotron radiation. **PNAS**, v. 97, n. 2, p. 623-628, 2000.
- 20 Gonzalez, A. Optimizing data collection for structure determination. **Acta Cryst D**, v. D59, n. 1, p. 1935-1942, 2003.
- 21 Garman, E. Cool data: quantity AND quality. **Acta Cryst D**, v. 55, p. 1641-1653, 1999.
- 22 Dauter, Z. Data-collection strategies. **Acta Cryst D**, v. D55, n. 1, p. 1703-1717, 1999.
- 23 Evans, P. Scaling and assessment of data quality. **Acta Cryst D**, v. 62, p. 72-82, 2006.
- 24 Kabsch, W. Integration, scaling, space-group assignment and post-refinement. **Acta Cryst D**, v. D66, n. 1, p. 133-144, 2010.
- 25 Dauter, Z. Carrying out an optimal experiment. **Acta Cryst D**, v. 66, p. 389-392, 2010.
- 26 Evans, G.; Axford, D.; Owen, R. L. The design of macromolecular crystallography diffraction experiments. **Acta Cryst D**, v. 67, p. 261-270, 2011.
- 27 Karplus, P. A.; Diederichs, K. Linking Crystallographic Model and Data Quality. **Science**, v. 336, n. 1, p. 1030-1033, 2012.
- 28 Wlodawer, A.; Dauter, Z.; Jaskolski, M. Protein crystallography for non-crystallographers, or how to get the best from published macromolecular structures. **FEBS Journal**, v. 275, n. 1, p. 1-21, 2008.
- 29 Battye, T. G.; Powell, H. R.; Leslie, A. G. iMOSFLM: a new graphical interface for diffraction image processing with MOSFLM. **Acta Cryst D**, v. D67, n. 1, p. 271-281, 2011.
- 30 Kabsch, W. XDS. **Acta Cryst D**, v. D66, n. 1, p. 125-132, 2010.

- 31 Powell, H. R.; Johnson, O.; Leslie, A. G. W. Autoindexing diffraction images with iMosflm. **Acta Cryst D**, v. 69, n., p. 1195–1203, 2013.
- 32 Leslie, A. G. W. The integration of macromolecular diffraction data. **Acta Cryst D**, v. D62, p. 48–57, 2006.
- 33 Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. **Acta Cryst D**, v. 67, p. 282–292, 2011.
- 34 Evans, P. R.; Murshudov, G. N. How good are my data and what is the resolution? **Acta Cryst D**, v. 69, p. 1204–1214, 2013.
- 35 Adams, P. D.; Afonine, P. V.; Read, R. J.; Richardson, J. S.; Terwilliger, T. C. Recent developments in phasing and structure refinement for macromolecular crystallography. **Current Opinion in Structural Biology**, v. 19, n. 1, p. 566–572, 2009.
- 36 Taylor, G. L. Introduction to phasing. **Acta Cryst D**, v. D66, n. 1, p. 325–338, 2010.
- 37 McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. **Acta Cryst D**, v. 67, p. 32–41, 2007.
- 38 Abergel, C. Molecular replacement: tricks and treats. **Acta Cryst D**, v. 69, p. 2167–2173, 2013.
- 39 Scapin, G. Molecular replacement then and now. **Acta Cryst D**, v. 69, p. 2266–2275, 2013.
- 40 McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. **Journal of Applied Crystallography**, v. 40, n. 4, p. 658–674, 2007.
- 41 Tickle, I. J. Statistical quality indicators for electron-density maps. **Acta Cryst D**, v. 68, p. 454–467, 2012.
- 42 Terwilliger, T. C.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Adams, P. D.; Read, R. J.; Hunga, L. W. Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. **Acta Cryst D**, v. 64, p. 515–524, 2008.
- 43 Acharya, K. R.; Lloyd, M. D. The advantages and limitations of protein crystal structures. **TRENDS in Pharmacological Sciences**, v. 26, n. 1, p. 10–14, 2005.
- 44 Fernandes, W. B.; Martins, F. T.; Lião, L. M.; Camargo, A. J.; Lariucci, C.; Perez, C. N.; Napolitano, H. B. Conformational variability in a new terpenoid-like bischalcone: structure and theoretical studies. **Journal of Structural Chemistry**, v. 54, n. 6, p. 1112–1121, 2013.
- 45 Alpeshkumar, K. M.; Mark, A. E. Challenges in the determination of the binding modes of non-standard ligands in X-ray complexes. **J Comput Aided Mol Des**, v. 25, p. 1–12, 2011.
- 46 Afonine, P. V.; Adams, P. D.; Urzhumtsev, A. Bulk-solvent and overall scaling revisited: faster calculations, improved results. **Acta Cryst D**, v. 69, p. 625–634, 2013.
- 47 Afonine, P. V.; Grosse-Kunstleve, R. W.; Adams, P. D. A robust bulk-solvent correction and anisotropic scaling procedure. **Acta Cryst D**, v. 61, p. 850–855, 2005.
- 48 Kantardjieff, K. A.; Rupp, B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins crystals. **Protein Science**, v. 12, n. 1, p. 1865–1871, 2003.
- 49 Murshudov, G. N.; Steiner, R. A.; Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures **Acta Cryst D**, v. D67, n. 1, p. 355–367, 2011.
- 50 Afonine, P. V.; Headd, J. J.; Adams, P. D. Towards automated crystallographic structure refinement with phenix.refine. **Acta Cryst D**, v. D68, n. 1, p. 352–367, 2012.
- 51 Urzhumtseva, A.; Afonine, P. V.; Adams, P. D.; Urzhumtsev, A. Crystallographic model quality at a glance. **Acta Cryst D**, v. 65, p. 297–300, 2009.
- 52 Tronrud, D. E. Introduction to macromolecular refinement. **Acta Cryst D**, v. 60, p. 2156–2168, 2004.

- 53 Headd, J. J.; Afonine, P. V.; Richardson, D. C.; Adams, P. D. Use of knowledge-based restraints in phenix refine to improve macromolecular refinement at low resolution. **Acta Cryst D**, v. D68, n. 1, p. 381-390, 2012.
- 54 Brunger, A. T.; Barre, B.; Davies, J. M.; Weisb, W. X-ray structure determination at low resolution. **Acta Cryst D**, v. D65, n. 1, p. 128-133, 2009.
- 55 Evans, P. R. An introduction to stereochemical restraints. **Acta Cryst D**, v. 63, p. 58–61, 2007.
- 56 Wiegels, T.; Lamzin, V. S. Use of noncrystallographic symmetry for automated model building at medium to low resolution. **Acta Cryst D**, v. 68, p. 446–453, 2012.
- 57 Zucker, F.; Champ, P. C.; Merritt, E. A. Validation of crystallographic models containing TLS or other descriptions of anisotropy. **Acta Cryst D**, v. D66, n. 1, p. 889-900, 2010.
- 58 Davis, I. W.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Richardson, J. S. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. **Nucleic Acids Research**, v. 35, n. 1, p. W375–W383, 2007.
- 59 Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. **Acta Cryst D**, v. 60, n. 12 Part 1, p. 2126-2132, 2004.
- 60 McNicholas, S.; Potterton, E.; Wilson, K. S.; Noble, M. E. Presenting your structures: the CCP4mg molecular-graphics software. **Acta Cryst D**, v. 67, n. 4, p. 386-394, 2011.
- 61 Fernandes, W. B.; Aragão, A. Q.; Martins, F. T.; Perez, N. C.; Lariucci, C.; Napolitano, H. B. The discrete role of chlorine substitutions in the conformation and supramolecular architecture of arylsulfonamides. **Acta Cryst C**, v. 67, n. 7, p. 226-229, 2011.
- 62 Bissantz, C.; Kuhn, B.; Stahl, M. A Medicinal Chemist's Guide to Molecular Interactions. **J Med Chem**, v. 53, n. 1, p. 5061-5084, 2010.
- 63 Holton, J.; Alber, T. Automated protein crystal structure determination using elves. **Proceedings of the National Academy of Sciences of the United States of America**, v. 101, n. 6, p. 1537-1542, 2004.
- 64 Padilla, J. E.; Yeates, T. O. A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. **Acta Cryst D**, v. 59, p. 1124-1130, 2003.
- 65 Chruszcz, M.; Potrzebowski, W.; Zimmerman, M. D.; Grabowski, M.; Zheng, H.; Lasota, P.; Minor, W. Analysis of solvent content and oligomeric states in protein crystals—does symmetry matter? **Protein Science**, v. 17, p. 623-632, 2008.
- 66 Wiggers, H. J.; Rocha, J. R.; Fernandes, W. B.; Sesti-Costa, R.; Carneiro, Z. A.; Cheleski, J.; Da Silva, A. B. F.; Juliano, L.; Cezari, M. H. S.; Silva, J. S.; Mckerrrow, J. H.; Montanari, C. A. Non-peptidic Cruzain Inhibitors with Trypanocidal Activity Discovered by Virtual Screening and In Vitro Assay. **PLoS NTD**, v. 7, n. 8, p. e2370, 2013.
- 67 Zwart, P. H.; Lebedev, A. A.; Murshudov, G. N.; Adams, P. D. Surprises and pitfalls arising from (pseudo)symmetry. **Acta Cryst D**, v. 64, n., p. 99-107, 2008.
- 68 Kim, W. M.; Sigalov, A. B.; Stern, L. J. Pseudo-merohedral twinning and noncrystallographic symmetry in orthorhombic crystals of SIVmac239. **Acta Cryst D**, v. 66, p. 163-175, 2010.

Capítulo 5

Estruturas Cristalográficas de Alta Resolução, Ensaio de Novos Inibidores e Mapeamento do Sítio Catalítico

Este capítulo inicia-se com a elucidação de duas estruturas de alta resolução para a cruzaína nativa em complexo com Iodoacetamida e MMTS. Estas estruturas demonstram o efeito da estabilização da auto-proteólise na qualidade e resolução de estruturas provenientes de cisteíno proteases da família da papaína. Com base nestes resultados, novas abordagens para a cristalização com inibidores reversíveis foram testadas levando aos primeiros cristais da cruzaína crescidos e coletados no Brasil. Todas as informações estruturais obtidas neste trabalho, incluindo dados provenientes da sondagem do sítio catalítico com os programas SuperStar e GRID, foram reunidas e analisadas para o mapeamento de interações específicas que devem ser realizadas para o incremento de potência e seletividade dos novos inibidores.

5 ESTRUTURAS CRISTALOGRÁFICAS DE ALTA RESOLUÇÃO, ENSAIOS DE NOVOS INIBIDORES E MAPEAMENTO DO SÍTIO CATALÍTICO

5.1 A estabilização da auto-proteólise leva a estruturas de alta qualidade e resolução para complexos com a cruzaína nativa

Durante os ensaios de cristalização, foram feitas tentativas de se estabilizar o efeito auto-proteolítico da cruzaína nativa com o uso de dois conhecidos inibidores covalentes de cisteína proteases. A Iodoacetamida e o metanotiosulfonato de metila (MMTS). Estes inibidores ligam-se covalentemente à cisteína catalítica de maneira irreversível no caso da Iodoacetamida pelo mecanismo exibido na Figura 5.1, e de maneira reversível no caso do MMTS podendo ser facilmente removido na presença de agentes redutores como o DTT, conforme mecanismo da Figura 5.2.¹

Figura 5.1 - Mecanismo de inibição irreversível de cisteína proteases por Iodoacetamida.

Fonte: Autoria própria.

Figura 5.2 - Mecanismo de inibição covalente reversível de cisteína proteases por MMTS.

Fonte: Autoria própria.

Desde os primeiros protocolos de purificação da cruzaína, já se recomendava o uso do MMTS logo após a ativação para evitar a degradação enzimática durante as etapas subsequentes de purificação. Este era posteriormente removido com DTT quando a enzima era usada em ensaios enzimáticos, ou substituído pelos inibidores de interesse em etapa anterior a cristalização no caso de estudos cristalográficos.

Em algumas situações o MMTS não era completamente removido do sítio catalítico antes de se iniciar a cristalização, aparecendo nos mapas de densidade eletrônica com uma ocupação parcial conforme observado na estrutura 3KKU (Figura 5.3) e também na estrutura 4KLB do complexo com o inibidor Neq0176 (Seção 4.4.2).^{2,3} Conforme detalhado na Seção 3.2.2, possivelmente o MMTS parcialmente ligado à cisteína catalítica da estrutura 3KKU tenha sido o responsável pela formação do cristal, uma vez que este diminuía drasticamente a auto-proteólise durante a cristalização. É provável que a presença do MMTS no sítio não tenha influenciado na diferente conformação adotada pelo inibidor B95 (código identificador do ligante no PDB) na estrutura cristalina em relação à pose de docagem que o identificou na triagem virtual, conforme Figura 5.3B. Isto porque o MMTS e o inibidor B95 foram modelados como desordem estática, significando que cada um destes complexos existe em regiões independentes do cristal. Esta diferente conformação observada é, possivelmente, função do sítio S1' ser superficial, grande e exposto ao solvente, cujas predições do MOB por docagem em tais sítios geralmente são imprecisas.⁴⁻⁶

Figura 5.3 - Em **A**, estrutura da cruzaina nativa (código PDB 3KKU) mostrando o MOB do inibidor reversível B95 com 70% de ocupação, com destaque para a molécula de MMTS ligada covalentemente com a Cys25B com 30% de ocupação. Mapa $2F_o - F_c$ a 1σ em azul e mapa de diferença $F_o - F_c$ à 3σ para contornos positivos (verde) e negativo (vermelho). Em **B**, sobreposição do MOB do inibidor B95 (em ciano) com a pose de docagem (em verde) em superfície de Van der Waals.

Fonte: Figuras próprias preparadas apartir dos dados PDB 3KKU de autoria de FERREIRA, R., 2010, p. 4900.²

Uma estrutura cristalográfica de alta resolução (1,95 Å) foi obtida para a cruzaina nativa em complexo com o MMTS ligado covalentemente à Cys25 com 100% de ocupação. Esta estrutura foi obtida por acaso em um dos ensaios de co-cristalização realizados em colaboração com o professor Brian Schoichet (co-orientador deste trabalho juntamente com o professor James McKerrow), para dois potentes inibidores reversíveis da cruzaina (código ZINC 08691187 e 3363859 com K_i de 2 µM e 6 µM respectivamente). Estes dois inibidores foram identificados pelos mesmos métodos de triagem que identificaram o B95 como inibidor da cruzaina e deram origem a estrutura 3KKU elucidada por Ferreira, Schoichet e colaboradores em 2010.² Dos ensaios de cristalização, cresceram apenas cristais para o complexo com o inibidor ZINC 3363859 por difusão de vapor em gota suspensa a 291 K na condição 0,1M Tris (pH 8,0) com 2,0M de fosfato de amônio. Interessantemente, foi observado que o ligante ZINC 3363859 cristalizou em uma região entre cadeias próxima ao sítio catalítico (Figura 5.4A) que assim como observado para o MMTS, também exibiu 100% de ocupação, indicando um possível sítio alostérico nesta posição. A possibilidade de existência de sítios alostéricos já havia sido reportada por McCammon *et al.* em 2010 em um trabalho que envolveu a identificação de novos possíveis sítios de ligação utilizando simulação de dinâmica molecular.⁷

Para o melhor cristal deste complexo, foram coletadas 193 imagens de difração com um segundo de exposição por imagem e $\Delta\phi$ de 1°. A análise no XTRIAGE indicou a presença de anéis de gelo com menores valores de completude para os dados após escalonamento nas resoluções de 3,98 Å (77%), 3,44 Å (88%) e 2,24 Å (65%). Conjuntos de dados sem anéis de gelo também foram coletados para outros cristais do mesmo complexo apresentando, no entanto, resolução menor do que 2,25 Å para $I/\sigma(I) > 2$. A análise do coeficiente de Matthews indicou uma cópia da cruzaina por ASU com conteúdo de solvente de 40,2% e V_m de 2,06. A estrutura foi elucidada utilizando procedimento similar ao descrito na Seção 4.3 do capítulo anterior. As estatísticas do processamento de dados e refinamento do modelo são exibidas na Tabela 5.1. O modelo final obtido com 97% dos ângulos ϕ e ψ em regiões favoráveis do gráfico de Ramachandran está apresentado na Figura 5.4.

O inibidor ZINC 3363859 interage na região intersuperficial (Figura 5.4B) por ligações de hidrogênio com os aminoácidos Asn79, Tyr84, e Asn80. Contatos hidrofóbicos são realizados com os resíduos Thr103, Trp74, Val108. Outra ligação de hidrogênio é realizada com a Asn80 da cadeia vizinha gerada pela simetria (0,0,-

1), que ainda realiza contatos hidrofóbicos com o inibidor através da Thr111. A área de contato interfacial do inibidor com a cadeia principal da cruzaina e com a cadeia vizinha gerada pela simetria (0,0,-1) é de 233,7 Å² e 158,4 Å² respectivamente. Este inibidor cristalizou em uma região pouco acessível aos grandes canais de solvente do cristal (Figura 5.4D), pois uma terceira cadeia da cruzaina gerada pela operação de simetria (-1, -1, -1) delimita uma espécie de bolsão onde apenas moléculas de água conseguem transitar. Apesar da área de contato interfacial do inibidor com esta terceira cadeia ser de 60 Å², nenhuma interação significativa é observada.

Tabela 5.1 - Estatísticas da coleta de dados e de refinamento do modelo para a cruzaina nativa em complexo com MMTS.

Coleta e processamento de dados		Refinamento	
Grupo espacial	P6 ₅ 22 (n°179)	Resolução (Å)	33,80 - 1,95 (2,02 - 1,95)*
Nº de mol. na ASU	1	Nº de reflexões (conj. teste)	24367 (1867)*
Cela unitária		<i>R</i> _{work} / <i>R</i> _{free} (%)	15,4 / 18,9 (20,0 / 26,8)*
<i>a</i> ; <i>b</i> ; <i>c</i> (Å)	82,34 82,34 101,60	Nº átomos:	Proteína 1623
α; β; γ (°)	90 90 120		Ligante 28
<i>R</i> _{merge} (%)	7,2 (82,0)*		Água 62
<i>R</i> _{meas} (%)	6,7 (74,9)*	Fator <i>B</i> (Å ²):	Proteína 21,2
Completeza (%)	96,2 (99,8)*		Ligante 28,2
<i>I</i> /σ(<i>I</i>)	31,8 (3,3)*		Água 25,4
Multiplicidade	10,9 (6,4)*	RMSD Comp. de ligação (Å)	0,007
Reflec. obs. (únicas)	616168 (56223)	Ângulo de ligação (°)	1,05

* Valores para a camada de maior resolução.

Fonte: Autoria própria.

O empacotamento cristalino adotado por esta estrutura (Figura 5.4D) é similar ao observado na estrutura 3KKU cuja sobreposição destas, revela sutis mudanças estruturais principalmente na alça entre os aminoácidos His106 e Cys101 que estão em contato direto ou indireto com o inibidor. A possibilidade de que esta região atue como um sítio alostérico modulador da atividade enzimática é reforçada pelo fato da Cys101 estar ligada à Cys56, via ligação dissulfeto, podendo facilmente transferir mudanças estruturais para aminoácidos próximos que exercem importantes papéis no sítio catalítico tais como a Ser61, Gly65, Gly66, Leu67 e Met68. Estes poderiam mudar de conformação ou alterar suas posições relativas como no caso da Gly65 e da Gly66 cujas alterações em décimos de Angstroms em relação a Asp161 e a Cys25, impactariam diretamente no posicionamento dos substratos peptídicos afetando a taxa de hidrólise destes pela tríade catalítica. O MMTS se liga covalentemente à Cys25 por uma ligação S-S de 1,98 Å de comprimento (Figura

5.4C). Não existe densidade eletrônica no sítio catalítico da cruzaina nesta estrutura, sendo esta uma boa oportunidade de analisar o comportamento estrutural na ausência de ligantes principalmente para o sítio S2. O sítio catalítico por outro lado é exposto ao canal de solvente, o que possibilitaria a realização de experimentos de *soaking* e cristalização tais como os apresentados na próxima seção.

Figura 5.4 - Estrutura cristalográfica da cruzaina nativa em complexo com MMTS. A cruzaina e a cadeia vizinha gerada pela operação de simetria (0,0,-1) em superfície de potencial eletrostático é exibida em **A**, com destaque para a região entre estas cadeias onde o inibidor ZINC 3363859 cristalizou (em **B**) e para a região do sítio onde o MMTS se ligou covalentemente à Cys25 (em **C**). Em **B** e **C**, são exibidos mapas de densidade $2F_o - F_c$ contornados à 1σ em azul e mapa de diferença $F_o - F_c$ à 3σ para contornos positivos (verde) e negativos (vermelho). O empacotamento cristalino visto na direção do eixo z (obtido pela rotação de 90° em torno do eixo x de **A**, e expansão do retículo cristalino em 100 Å em torno da cadeia principal) é exibido em **D** e **E**. Em **E** os átomos da cadeia principal foram omitidos para evidenciar que o inibidor ZINC 3363859 interage em uma região distante do canal de solvente principal. A cela unitária é representada pelo losango de cor preta em **D** e **E**.

Fonte: Autoria própria

Dada à impossibilidade de se obter cristais da forma apo da cruzaina nativa, vários ensaios de co-cristalização (Seções 3.2 e 3.3) foram configurados na tentativa de se determinar o MOB dos fragmentos Neq0147, Neq0176, Neq0220 e Neq0221

nesta enzima. Em um destes ensaios, fragmentos foram co-cristalizados na cruzaína nativa inibida com iodoacetamida (que formaria um complexo estável) na expectativa de que os mesmos cristalizassem em algum subsítio de ligação caso tivessem afinidade ou tamanhos apropriados para isto. Vários cristais surgiram nestes ensaios e alguns conjuntos de dados foram coletados. No entanto, estes cristais sempre cresciam na condição química 1M cloreto de lítio, 0,1M de ácido cítrico e 30% (peso / volume) de PEG 6000 com pH final 4,0 independente de qual fragmento ou coquetéis de fragmentos eram utilizados na co-cristalização. O processamento destes dados indicava o grupo espacial C2 (Tabela 5.2) que ainda não havia sido reportado no PDB para a cruzaína, e sempre exibiam densidade no sítio apenas para a iodoacetamida covalentemente ligada a Cys25 (Figura 5.5A). A falta de densidade eletrônica para os fragmentos no sítio era decorrente da natureza reversível e de baixa potência destes ligantes que levavam a baixíssima ocupação no sítio mesmo quando incubada com excesso molar do ligante.

Uma das estruturas obtidas para este complexo, no entanto, exibiu a melhor resolução (1,60 Å) de todas as estruturas elucidadas neste trabalho sendo outra prova do efeito da estabilização da auto-proteólise na obtenção de estruturas de alta qualidade e resolução. Foram coletadas 194 imagens com $\Delta\phi$ de 0,90° e cinco segundos de exposição por imagem, exibindo mosaicidade de 0,45°. Uma análise no XTRIAGE após o processamento usando iMosflm e Scala indicou que estes eram dados de altíssima qualidade. O coeficiente de Matthews sugeriu uma molécula na ASU com conteúdo de solvente de 29,8% e V_m de 1,75. Este foi o menor conteúdo de solvente obtido em todos os complexos cristalinos elucidados neste projeto, confirmando as hipóteses levantadas na Seção 4.4 sobre o efeito do alto conteúdo de solvente na sensibilidade à radiação e baixo poder de difração destes cristais.

Portanto a flexibilidade, reversibilidade e menor potência dos inibidores cujos modos de interação foram reportados no capítulo anterior, implicaram em baixa taxa de ocupação no sítio catalítico da cruzaína que levaram a uma proteólise residual e consequente heterogeneidade química e conformacional da proteína durante a cristalização. Assim, tal como discutido na Seção 3.1, a cristalização de uma proteína que perde gradativamente sua homogeneidade leva a empacotamentos que priorizam poucos contatos entre cadeias que poderão exibir defeitos na rede cristalina (ou patologias cristalográficas do ponto de vista do processamento de

dados) e alto conteúdo de solvente. O procedimento descrito na Seção 4.3 foi usado na elucidação desta estrutura que reportou as estatísticas exibidas na Tabela 5.2.

Tabela 5.2 - Estatísticas da coleta de dados e de refinamento do modelo para a cruzaína nativa em complexo com iodoacetamida.

Coleta e processamento de dados		Refinamento	
Grupo espacial	C121 (n5)	Resolução (Å)	37,81 - 1,60 (1,67 - 1,60)*
Nº de mol. na ASU	1	Nº de reflexões (conj. teste)	21532 (861)*
Cela unitária		R_{work} / R_{free} (%)	15,9 / 18,8 (18,2 / 21,1)*
<i>a</i> ; <i>b</i> ; <i>c</i> (Å)	64,62 34,80 82,23	Nº átomos: Proteína	1616
α ; β ; γ (°)	90 113,13 90	Ligante	4
R_{merge} (%)	3,5 (19,5)*	Água	124
R_{meas} (%)	5,0 (27,6)*	Fator B (Å ²): Proteína	15,3
Completeza (%)	95,4 (80,7)*	Ligante	16,1
$I/\sigma(I)$	15,5 (5,8) *	Água	23,5
Multiplicidade	8,1 (5,3)*	RMSD Comp. de ligação (Å)	0,006
Reflec. obs. (únicas)	167257 (20744)	Ângulo de ligação (°)	1,10

* Valores para a camada de maior resolução.

Fonte: Autoria própria.

Figura 5.5 - Estrutura do complexo cruzaína nativa ligada covalentemente à iodoacetamida em **A**, com destaque para os principais resíduos de aminoácidos do sítio catalítico e as moléculas de água identificadas na primeira camada de solvatação. Mapas de densidade $2Fo-Fc$ exibido à 1σ em azul e mapa $Fo-Fc$ a 3σ em verde e vermelho. Em **B** estão sobrepostas às estruturas dos complexos da cruzaína nativa com iodoacetamida em verde, com MMTS em dourado e com a estrutura 1ME4 (sem o ligante) em cinza destacando a flexibilidade para os resíduos de aminoácidos para a cruzaína no estado ligado e não ligado.

Fonte: Autoria própria

A iodoacetamida ligou-se covalente à cisteína catalítica com comprimento de ligação de 1,96 Å, sendo orientada no sítio por uma ligação de hidrogênio com a Gln19. Esta estrutura de alta resolução poderia ser considerada uma forma “pseudo-apo” da cruzaína devido à maneira como foi cristalizada e por não haver densidade eletrônica para outros ligantes no sítio catalítico. Logo, esta estrutura poderia indicar

aspectos estruturais que ainda não foram observados em relação às formas holo (em complexo com ligantes) desta enzima, dada a ausência no PDB de estruturas da cruzaina nativa na sua forma apo. Tais diferenças estruturais ficam evidenciadas na Figura 5.5B, onde foi realizada a sobreposição de estruturas da cruzaina nativa em complexo com a iodoacetamida, com o MMTS, e com a estrutura de código PDB 1ME4 que cristalizou com um inibidor reversível interagindo nos sítios S1, S2 e S3.

Nesta figura fica evidente a grande flexibilidade dos aminoácidos Leu67, Gln159 e Glu208 que atuam principalmente no sítio S2, embora a flexibilidade da Leu67 também possa impactar na potência do sítio S3. Outro destaque também é dado para o resíduo Ser61 e Leu160 que teve leve mudança estrutural (posição e conformação relativa). É provável que a flexibilidade destes resíduos nos sítios S2 e S3 estejam relacionadas com a capacidade da cruzaina em reconhecer um número muito maior de sequências de aminoácidos em relação às homólogas em humanos.^{8,9} Todos os outros aminoácidos se mantiveram praticamente inalterados com exceção do Trp184 e da Met145 que tiveram leve mudança em suas posições relativas entre as formas apo e holo. As mudanças conformacionais e posicionais mais expressivas foram observadas para os resíduos Gln159, Glu208 e Leu67 na estrutura do complexo com o MMTS em relação às outras estruturas da cruzaina. Estas mudanças estruturais são provavelmente decorrentes da influência do inibidor de código ZINC 3363859 que se ligou em um possível sítio alostérico que se comunica com o sítio catalítico via ligação de sulfeto entre a Cys101 e a Cys56. Além da comprovação da influência do mecanismo de inibição na estabilização do efeito proteolítico e consequente obtenção de estruturas de alta qualidade e resolução, as estruturas apresentadas nesta seção abriram novas possibilidades de experimentos de cristalização tais como os descritos na Seção 5.2.1.

5.2 Ensaios de cristalização dos novos inibidores planejados com base nos modos de interação elucidados para o Neq0147 e para o Neq0409

A sobreposição do MOB elucidado para os inibidores Neq0147, Neq0176 e Neq0409 (Figura 5.6) revelou informações cruciais para o planejamento de inibidores de segunda geração muito mais potentes. O grupo carboxilato como *warhead* no Neq0147 e por analogia o no Neq0220 “âncora” o inibidor nesta posição do sítio da cruzaina C25S devido à rede de interações de hidrogênio que é formada com os aminoácidos Gln19, His162 e cadeia principal da Ser25. Como a cruzaina

C25S foi validada como um modelo cristalográfico representativo para a caracterização do MOB de inibidores de baixa potência (Seção 4.4.4) existe grande probabilidade de que este também seja o MOB na cruzaina nativa. Da sobreposição com o MOB do Neq0409, fica evidente que o grupo carboxilato do Neq0147 na cruzaina C25S seria equivalente ao grupo nitrila do Neq0409 na cruzaina nativa. O anel oxadiazol do Neq0147 e o grupo carboxamida do Neq0176 estão posicionados em uma região onde geralmente se verifica o grupo amida dos inibidores peptídeo-miméticos. Estes grupos poderiam ser otimizados para interagirem eficientemente com os aminoácidos Gly65, Gly66 e Asp161 atuando como bioisósteros do grupo amida se tornando, portanto, atraentes alternativas aos esqueletos peptídicos dos atuais inibidores cujo desenvolvimento tem sido evitado devido à possibilidade de serem metabolizados antes mesmos de atuarem no sítio catalítico (Seção 1.4).

Figura 5.6 - Sobreposição do MOB dos inibidores Neq0147, Neq0176 e Neq0409 exibidos em superfície de Van der Waals.

Fonte: Autoria própria

O anel tiofeno do Neq0176, devido sua excelente posição e orientação, realiza interessantes contatos hidrofóbicos no início do sítio S2. Diferentes substituintes no anel poderiam ser testados de forma a sondar interações com o Glu208 na base deste sítio. O mesmo poderia ser feito com o grupo fenilalanina do Neq0409 que entrou mais profundamente no sítio S2. O átomo de cloro em orto no anel aromático do Neq0147 se mostrou importante para a potência observada deste inibidor devido aos contatos hidrofóbicos e ligações de halogênio com a Met68 e cadeia principal da His162. No entanto, outros substituintes em diferentes posições

deste anel aromático poderiam ser testados de forma a incrementar principalmente a seletividade dos inibidores frente à catepsina L de humanos. A exploração do sítio S3 possivelmente não resultará em melhorias em termo de seletividade para o inibidor podendo, no entanto, contribuir para o ganho na potência. A porção P3 do Neq0176 se mostrou interessante para este propósito uma vez que o grupo acetamido ligado ao anel tiofeno realiza interações de hidrogênio com a Gly66, e o anel triazol interage por ligações de hidrogênio com a Ser61 e com a cadeia principal da Thr59. No entanto, os três graus de liberdade rotacional entre estes dois grupos se mostrou problemática em um sítio tão largo e superficial como o S3 (Seção 4.4.2).

Estas informações estruturais levaram ao planejamento de novos inibidores contendo o esqueleto base do Neq147 e o grupo nitrila do Neq0409 como *warhead*. Assim, esperava-se que a nitrila se ligaria de maneira covalente reversível com a Cys25 funcionando como ancora (garantindo que o MOB dos novos inibidores fosse similar ao do Neq0147) para otimização de interações nos aneis heterocíclico e clorofenil. Poucos eram os compostos comercialmente disponíveis contendo este esqueleto, que foram então comprados e bioquimicamente avaliados contra a cruzaina no projeto de mestrado realizado por Karen Rangel no Nequimed. Alguns destes compostos exibiram afinidade até doze vezes maiores do que a do Neq0147, tais como o Neq0556 e Neq0557, apresentados na Figura 5.7.

Figura 5.7 - Inibidores da cruzaina do tipo fragmentos, planejados com base nos modos de interação do Neq0147 e Neq0409, que foram utilizados nos ensaios cristalográficos.

Fonte: Autoria própria.

O efeito da nitrila na potência destes inibidores precisava ser avaliado. Como o Neq0147 contendo o *warhead* nitrila não estava comercialmente disponível, este

foi sintetizado dando origem ao Neq0617 sendo observado que a maior parte da potência dos novos inibidores era devido à nitrila. Porém, a elevada potência do Neq0556 e Neq0557 demonstra que a otimização do anel heterocíclico e do anel clorofenil levará a compostos mais potentes do que o Neq0617. A caracterização estrutural do MOB via cristalografia permitiria determinar com precisão se estes esqueletos derivados do Neq0147 contendo nitrilas como *warhead* ainda mantém a mesma orientação e padrão de interações no sítio, deixando evidentes quais interações ainda precisariam ser otimizadas para que estes inibidores continuem o processo de evolução. Alguns dos inibidores mais potentes e interessantes desta série (Figura 5.7) foram selecionados para novos ensaios de cristalização.

O MOB elucidado para o Neq0409 (Seção 4.4.1) validou as hipóteses iniciais sobre o uso de dipeptidil-nitrilas como inibidores reversíveis da cruzaina. Uma série de novos inibidores baseados no esqueleto do Neq0409 tem sido planejadas e sintetizadas no grupo Nequimed dentro do projeto temático Fapesp nº 13/18009-4 tendo como um dos objetivos, sondar interações específicas nos sítios S2 e S3 em busca de incrementos de potência e seletividade. De modo a obter informações estruturais detalhadas que seriam fundamentais para o cumprimento deste objetivo, alguns dos inibidores mais potentes e representativos desta série de dipeptidil-nitrilas (Figura 5.8) foram também selecionados para novos ensaios de cristalização.

Figura 5.8 - Inibidores da cruzaina do tipo dipeptidil-nitrilas, planejadas a partir do composto protótipo Neq0409, que foram cristalográficamente ensaiados.

Fonte: Autoria própria.

O grupo nitrila como *warhead*, devido à sua ligação covalente reversível, aumentaria em pelo menos uma ordem de grandeza a afinidade de tais inibidores. Além disto, o tempo de residência (t_R) no sítio seria muito maior tornando estes inibidores bons candidatos para serem cristalizados na cruzaina nativa, uma vez que controlariam melhor a proteólise residual. Como estes inibidores são menos flexíveis e interagem em regiões mais profundas do sítio, as variantes do MOB observadas para o Neq0176 (Seção 4.4.2) não mais existiriam, contribuindo para a cristalização em empacotamentos de menor conteúdo de solvente com uma ou duas moléculas da cruzaina por ASU, o que refletiria em maior resolução e qualidade de dados.

5.2.1 *Proposição e teste de novas abordagens de cristalização para obtenção do MOB dos inibidores de segunda geração*

As estruturas elucidadas no início deste capítulo abriram novas possibilidades de cristalização para a série de dipeptidil-nitrilas e fragmentos com o *warhead* nitrila (Figuras 5.7 e 5.8) cuja caracterização do MOB trará informações estruturais cruciais para a identificação e o desenvolvimento de substâncias tripanocidas que sejam potenciais candidatos a fármacos. Cristais da cruzaina nativa em complexo com o MMTS poderiam ser inseridos na mesma “solução mãe” de cristalização contendo DTT (ou outro agente redutor) para remoção do MMTS e posteriormente incubados com os inibidores de interesse via *soaking*. Outra abordagem inédita seria a produção de cristais da cruzaina com iodoacetamida e posterior *soaking* com fragmentos que fossem planejados para regiões entre sítios que não sofressem a influência da cisteína catalítica (regiões entre o sítio S1 e S1', S2 e S3 ou possíveis sítios alostéricos) que estaria covalentemente ligada à iodoacetamida.

Como cristais da cruzaina nativa em complexo com iodoacetamida e MMTS empacotam-se mais densamente, estes seriam mais resistentes e menos sensíveis a tais operações de “lavagem” e *soaking*. Além disto, estes cristais poderiam ser usados como gérmenes de cristalização para experimentos de semeadura cruzada (ou *cross seeding*, Seção 3.1.2), que aumentaria a qualidade dos cristais obtidos por co-cristalizações com os inibidores de interesse, uma vez que estes cristais cresceriam obedecendo ao empacotamento do germen inicial de cristalização. Isto também contribuiria para evitar problemas tais como o surgimento de várias cadeias na ASU relacionadas por NCS. A otimização das condições de cristalização da C25S e posterior *soaking* com os inibidores de interesse também evitaria este

problema. *Seeding* ainda não havia sido utilizado nas cristalizações exibidas neste trabalho devido a dificuldade em obter o primeiro germem de qualidade.

Ensaio de co-cristalização da cruzaína nativa com os inibidores do tipo dipeptidil-nitilas (Figura 5.8) priorizaram inicialmente condições de cristalização similares às obtidas para o Neq0409, variando-se principalmente o pH, concentração do precipitante e da proteína no formato *grid*. Na sequência, foram realizados testes utilizando os *kits* da *Hampton Research* HR2-142 PCT (*pre-crystallization test*), HR2-070 *slice* pH e HR2-072 *solubility stability screen* que possibilitaram determinar as faixas de concentração de proteína, estabilidade e pH ideais para a cristalização de cada complexo de interesse. Em seguida foram realizadas triagens de condições usando os *kits* HR2-110, HR2-112 e o HR2-130 *crystal screen HT*, HR2-134 *index HT*, HR2-248 *grid screen salt HT*. Também foram utilizados *kits* específicos selecionados com base nos resultados obtidos em triagens realizadas no início deste projeto tais como o HR2-211 *grid screen ammonium sulfate* e o HR2-247 *grid Screen Sodium malonate*. Triagens de condições de cristalização da forma apo da cruzaína C25S e do zimôgeno (pro-cruzaína) nativa também foram realizadas. Todos os ensaios de cristalização com estes inibidores de segunda geração foram manualmente configurados no laboratório de cristalografia do Nequimed que estava sendo montado. Dados cristalográficos foram obtidos no Laboratório Nacional de Luz Síncrotron (LNLS, Campinas) em única coleta realizada em Outubro de 2014.

5.2.2 Resultados preliminares e os primeiros cristais da cruzaína crescidos no Brasil

De modo a testar estas novas abordagens para cristalização da cruzaína, inicialmente foram repetidas as condições de cristalização da cruzaína nativa em complexo com MMTS (0,1M Tris pH 8,0 / 2,0M de fosfato de amônio), porém sem a presença do inibidor ZINC 3363859 que cristalizou entre cadeias da cruzaína na estrutura exibida na Seção 5.1. Após algumas tentativas, não foram obtidos cristais para este complexo significando que o ligante que cristalizou entre as cadeias (Figura 5.4B) é importante para a cristalização naquele empacotamento e condições químicas. No entanto, das triagens com os *kits* da Hampton, foram obtidos cristais grandes e de faces bem definidas para este complexo em condições apresentadas na Tabela 5.3. Estes cristais foram submetidos ao experimento de “lavagem” com diferentes concentrações de DTT (0,5 mM, 1 mM, 2,5 mM e 5 mM) solubilizado diretamente na “solução mãe” em tempos que variaram entre 5, 10, 15 e 30 minutos,

para avaliar a estabilidade dos cristais em tais soluções. Em geral estes cristais resistiram bem a soluções com menores concentrações de DTT com tempos entre cinco e dez minutos se danificando rapidamente em concentrações maiores de DTT (Figura 5.9C). Após a imersão em solução com DTT, os cristais foram submetidos às nitrilas Neq0409 (que já se conhecia o MOB, porém havia um interesse em obter estruturas de melhor resolução com uma única cadeia da cruzaina na ASU) e Neq0415 (o mais potente da série). Após *soaking* com o inibidor, ainda foi necessário realizar imersão em solução crioprotetora. Estes cristais se mostraram frágeis principalmente ao excesso de manipulação entre diferentes soluções. Devido à única oportunidade de coleta de dados no LNLS e ao pouco tempo disponível para otimização de tal experimento, não foram coletados dados processáveis. No entanto, esta abordagem apresentou potencial de funcionar caso seja otimizada.

Tabela 5.3 - Condições de cristalização identificadas nas triagens iniciais com *kits* da Hampton para os complexos da cruzaina nativa com os inibidores de segunda geração

Complexo cruzaina nativa / inibidor	Condição de cristalização	
	Tampão	Precipitante, sal e outros
Neq0409*	0,1M Hepes pH 7,5	2M de sulfato de amônio / 0,1M NaCl
Neq0415*	0,1M Hepes pH 7,0	1,6M de sulfato de amônio
Neq0556*	0,1M Hepes pH 7,5	1,5M malonato de sódio
Neq0568**	1,4M fosfato de sódio / potássio pH 6,3	
Neq0568*	0,1M Bis-Tris pH 6,5	1,5M de sulfato de amônio
Iodoacetamida*	0,1M Tris pH 8,0	1,0M succinato de sódio
Iodoacetamida*	0,1M Tris pH 8,0	1,0M citrato de sódio
Iodoacetamida**	0,1M Hepes pH 7,5	1,5M sulfato de amônio / 20% (v/v) glicerol
Iodoacetamida**	0,1M Hepes pH 7,5 /	1,0M citrato de sódio / 0,2M NaCl
MMTS**	0,1M Hepes pH 7,5 /	1,0M citrato de sódio / 0,2M NaCl

*Gotas 1:1 μL (proteína : solução de cristalização ou *mother liquor*); **Gotas 2:1 μL

Fonte: Autoria própria.

Não foi possível realizar uma triagem completa de condições de cristalização para os inibidores de segunda geração usando estes *kits* da Hampton, dado o pouco tempo disponível entre a montagem do laboratório de cristalografia do Nequimed e o único *beamtime* disponível no LNLS para Outubro de 2014, considerando a grande demanda de proteína recombinante pura requerida e a configuração manual dos ensaios. Porém, alguns resultados promissores foram obtidos nestas triagens iniciais como os cristais da cruzaina nativa em complexo com o Neq0568 (Figura 5.9A) que cresceram em condições apresentadas na Tabela 5.3. Cristais da cruzaina nativa em complexo com a iodoacetamida para os novos experimentos de cristalização

foram crescidos nas mesmas condições dos cristais descritos na seção anterior. Porém quatro novas condições de cristalização também foram encontradas para este complexo utilizando os *kits* da Hampton conforme apresentadas na Tabela 5.3.

Figura 5.9 - Cristais obtidos nos ensaios de cristalização da cruzaína nativa com a segunda geração de inibidores. Cristais dos complexos com o Neq0568 em **A**, iodoacetamida em **B**, MMTS submetidos a “lavagem” com DTT em **C** e Neq0556 em **D**.

Fonte: autoria própria

Os cristais até então obtidos cresceram em maiores pHs (7,5 e 8,0) possivelmente por estarem longes do pH ideal de atividade catalítica (5,2) discutido na Seção 3.2.2. As condições contendo citrato de sódio levaram a cristais de melhor resolução. Para o melhor cristal do complexo com o Neq0568 foram coletadas 225 imagens com $\Delta\phi$ de $0,8^\circ$, 6 segundos de exposição por imagem e λ de $1,458 \text{ \AA}$ na linha de luz MX2 do LNLS. Este cristal exibiu mosaicidade de 1° e uma resolução dos dados de $2,68 \text{ \AA}$. O processamento inicial no XDS indicou o grupo espacial C222 ($n^\circ 21$ da ITC) como o mais provável exibindo c ela unitária de parâmetros $41,9 \text{ \AA}$; $79,6 \text{ \AA}$; $51,9 \text{ \AA}$; 90° ; 90° e 90° . No entanto, anéis concêntricos típicos de cristais de gelo surgiram a partir da metade das imagens coletadas, tornando estes dados praticamente impossíveis de serem processados tanto no XDS quanto no iMosflm.

Alguns conjuntos de dados foram coletados para cristais da cruzaína nativa em complexo com iodoacetamida e com inibidores derivados do Neq0147 que foram

inseridos por *soaking* (Figura 5.9B). Após inúmeras tentativas, aparentemente o *soaking* estava danificando muito a qualidade de tais cristais principalmente para as camadas de maior resolução do padrão de difração. Este experimento também necessita de maiores otimizações. Contudo, uma estrutura interessante foi elucidada para um cristal de iodoacetamida que foi submetido à *soaking* com um coquetel de fragmentos composto por: 3,5 mM do Neq0563 que possuía o *warhead* NH₂ cuja expectativa era de que este tivesse menor influência da iodoacetamida no sítio; 5,5 mM do Neq0147 cujo *warhead* carboxilato também poderia interagir em outro subsítio dada à ocupação da tríade catalítica pela iodoacetamida; 7 mM do Neq220, o qual se esperava que o grupo metoxi interagisse mais profundamente no sítio S2; e três mM do Neq0556 que foi o mais potente da série da Figura 5.7.

O processamento dos dados com o XDS indicou o grupo espacial C222 com cela unitária de 194,6 Å; 195,1 Å; 109,4 Å; 90°, 90° e 90° como a mais provável. Uma análise no Xtriage indicou oito moléculas da cruzaina por ASU e conteúdo de solvente de 48,5% com V_m de 2,39 como os mais prováveis. No entanto, a análise dos dados com este programa também indicou a presença de pseudosimetria translacional. As primeiras tentativas de faseamento com o Phaser_MR encontraram soluções parciais com três moléculas na ASU não conectadas entre si apresentando *top* LLG de 2969,7 e *top* TFZ de 46,0. Na tentativa de reprocessar os dados no grupo espacial P₄ (número 75 da ITC) novamente a pseudosimetria foi detectada. A análise do coeficiente de Matthews indicou nove cópias na ASU para conteúdo de solvente de 48,3% e V_m = 2,38. Após fatores de correção terem sido aplicados, o programa Phaser_MR encontrou uma solução parcial com seis moléculas na ASU relacionadas por NCS com *top* LLG de 8056,13 e *top* TFZ de 66,9 no grupo espacial P₄₃ não existindo densidade eletrônica para outras cadeias. O conteúdo de solvente deste cristal, portanto era de 65,5% para seis moléculas na ASU. A estrutura foi refinada seguindo procedimento apresentado na Seção 4.3, levando a uma estrutura com 97% dos ângulos ϕ e ψ em regiões favoráveis do gráfico de Ramachandram, cujas estatísticas de coleta e refinamento estão exibidas na Tabela 5.4.

Nesta estrutura, a iodoacetamida ligou-se covalentemente à Cys25 de todas as seis cadeias exibindo um padrão de interações com a Gln19 similar ao observado na estrutura apresentada na Seção 5.1. Intensa densidade eletrônica foi observada no sítio catalítico em regiões vizinhas a Cys25 tal como exibido na Figura 5.10. Tentativas de se construir os inibidores foram realizadas não sendo possível, no

entanto, distinguí-los perfeitamente no sítio dado à similaridade estrutural, a natureza reversível e a baixa afinidade destes fragmentos. Esta estrutura, no entanto, demonstra que é possível obter o MOB de inibidores por *soaking* com fragmentos individualmente, ou em coquetéis desde que possuam esqueletos químicos diversos. Talvez o *soaking* tenha sido o responsável por danificar a rede cristalina levando a pseudosimetria, maior conteúdo de solvente e menor resolução. Mas isto só poderá ser comprovado na presença de dados para cruzaina nativa em complexo com iodoacetamida apenas, crescidos nas mesmas condições químicas. A disposição das cadeias da cruzaina na ASU e o empacotamento cristalino com destaque para os grandes espaços vazios correspondentes aos canais de solventes quando visualizado na direção do eixo z, são apresentados na Figura 5.11.

Tabela 5.4 - Estatísticas da coleta de dados e de refinamento do modelo para a cruzaina nativa em complexo com iodoacetamida em grupo espacial de maior simetria.

Coleta e processamento de dados		Refinamento	
Grupo espacial	P4 ₃ (n°78)	Resolução (Å)	29,6 - 2,47 (2,56 - 2,47)*
Nº de mol. na ASU	6	Nº de reflexões (conj. teste)	21532 (861)*
Cela unitária		R_{work} / R_{free} (%)	21,2 / 24,3 (30,7 / 32,9)*
a; b; c (Å)	137,6 137,6 109,4	Nº átomos: Proteína	9552
α ; β ; γ (°)	90 90 90	Ligante	69
R_{merge} (%)	6,7 (45,8)*	Água	178
R_{meas} (%)	7,5 (53,3)*	Fator B (Å ²): Proteína	43,5
Completeza (%)	94,9 (87,2)*	Ligante	59,9
$I/\sigma(I)$	10,7 (3,1)*	Água	41,5
Multiplicidade	4,6 (3,9)*	RMSD Comp. de ligação (Å)	0,008
Reflec. obs.	323802 (70964)	Ângulo de ligação (°)	1,11

* Valores para a camada de maior resolução.

Fonte: Autoria própria.

Figura 5.10 - Exibição estereoscópica do sítio da cruzaina (cadeia A) em complexo com iodoacetamida para o grupo espacial P4₃, cujo cristal foi submetido à *soaking* com um coquetel de fragmentos. Mapa de densidade eletrônica $2Fo-Fc$ contornado a 1σ é exibido em azul e o mapa $Fo-Fc$ à 3σ é exibido em verde e vermelho.

Fonte: Autoria própria.

Figura 5.11 - Em **A**, conteúdo da ASU na estrutura da cruzaina nativa em complexo com iodoacetamida no grupo espacial $P4_3$ com destaque para os sítios catalíticos de cada cadeia (círculos em cinza) e a região onde se cristalizou moléculas do tampão Hepes (círculos em vermelho). Em **B**, **C** e **D**, são exibidos o empacotamento cristalino visualizados na direção dos eixos z , x e y respectivamente. O quadro interno representa a cela unitária deste cristal.

Fonte: Autoria própria.

Todos os resultados preliminares obtidos neste estudo servirão de base para os futuros ensaios de cristalização que serão realizados no grupo NEQUIMED.

5.3 Sondagem do sítio catalítico da cruzaina com GRID e SuperStar em busca de pontos para incremento de potência e seletividade

A exploração do espaço químico-biológico em busca de substâncias bioativas inibidoras da cruzaina pelos métodos de triagem virtual (VS) tem funcionado bem para docagens realizadas para inibidores maiores em sítios catalíticos profundos e bem definidos para interações intermoleculares como a região entre a tríade e o sítio S2. Quando se parte para fragmentos muito flexíveis em sítios grandes, superficiais e com poucos pontos de interação, predições por pose docagem tem falhado tanto para os fragmentos elucidados neste trabalho, quanto para os da literatura. É importante obter o MOB por cristalografia especialmente para os fragmentos, pois estes em geral assumem poses diferentes das docagens para o sítio grande e exposto ao solvente. As propriedades do sítio de ligação devem ser mapeadas antes de se iniciar uma corrida de triagem virtual (VS).^{10,11} Várias ferramentas tem sido descritas para elucidação dos *hot spots* para interação em um sítio catalítico

particular. Estes métodos são baseados em parâmetros de campos de força (GRID), ou em informações empíricas (Superstar).¹² O programa GRID possui a função de gerar os campos de interação molecular conhecidos como MIF's (*Molecular Interactions Fields*). O programa GRID é um procedimento computacional para se determinar sítios de ligação favoráveis em moléculas de estruturas conhecidas.

As interações de um grupo de sondas com a proteína de estrutura conhecida é computada em uma dada posição através de superfícies de contorno com níveis de energia apropriados que são calculadas para cada sonda e apresentadas como gráficos computados junto com a estrutura da proteína. Contornos negativos de níveis de energia delineiam regiões de atração entre a sonda e a proteína que são encontradas em um sítio de ligação de um ligante em particular. Quanto mais negativo o valor da energia, mais atrativa e favorável é a atração. Os contornos também permitem que outras regiões de atração sejam identificadas facilitando a interpretação da energética entre proteína e ligante sendo, portanto, valiosa para o planejamento de fármacos. De modo a identificar regiões no sítio que pudessem contribuir para incremento de potência no planejamento de novos inibidores, o software GRID foi aplicado na estrutura de 1,6 Å de resolução da cruzaina nativa em complexo com iodoacetamida apresentada na Seção 5.1.

Inicialmente, moléculas de água e do ligante foram retiradas do modelo, e átomos de hidrogênio foram adicionados na geometria correta automaticamente pelo módulo *riding-hydrogen model* do programa Hermes. O pH 5,5 foi considerado para o ensaio, sendo este, o pH de operação da cruzaina no reservossoma do parasita. Portanto, os estados de protonação e possível tautomerismo foram checados para a Cys25 e para a His162. Cadeias laterais das histidinas, asparinas e Glutaminas presentes na estrutura também foram checadas. Os aminoácidos que margeam os sítios S1, S2, S3 e S4 foram usados para geração da caixa no Hermes. O software GRID foi aplicado a esta estrutura de sítio catalítico delimitado, usando-se dez diferentes sondas cujos resultados na forma de superfícies de contorno com os respectivos níveis de energia estão exibidos nas Figuras 5.12 e 5.13.

Pela análise da Figura 5.12 verifica-se que existe grande propensão em se colocar grupos hidrofílicos em região próxima a Ser61 no sítio S3 e também na região onde se localiza a Gly66 e a Asp161. Grupos hidrofóbicos se encaixariam bem na porção inicial dos sítios S2 e S3 interagindo provavelmente com a Leu67, Met68, Ala138 e Leu160. A propensão de grupos hidrofóbicos no sítio S1' é devida a

possibilidade de interações com o Trp184. Grupos metilas e átomos de carbono aromáticos seriam bem aceitos no sítio S2. Grande propensão a grupos contendo nitrogênios de amida existe na região entre a Gly66 e Asp161 tal como geralmente observado nos inibidores da cruzaina em seus esqueletos peptídicos. Também existe uma energia favorável para tal grupo no limite do sítio S3.

Figura 5.12 - Sondagem do sítio catalítico da cruzaina com o programa GRID utilizando as sondas H₂O ou hidrofílica em **A**, dry ou hidrofóbica em **B**, C3 em **C**, C1 em **D**, N1 em **E** e O_{sp} em **F**. O sítio catalítico é apresentado em superfície de Van der Waals com destaque para a Cys25 colorida em amarelo. A superfície de potencial energético é mostrada em azul com o nível energético de atração para cada sonda destacado em amarelo.

Fonte: Autoria própria

Oxigênios de grupos carbonila seriam bem aceitos na região da Gly65 e da Cys25, e também na região entre a tríade catalítica e a Gln19. Interações favoráveis são observadas para hidroxilas fenólicas e carboxílicas na região entre Gly66 e Asp161. Grande propensão para átomos de flúor na base do sítio S2. A mesma propensão é observada para as sondas cloro e bromo que ainda apresentam energias favoráveis para interações no início do sítio S2 onde se localiza a Met68. A sobreposição do MOB obtido para o Neq0147 na C25S com os resultados obtidos no GRID para a sonda cloro é exibida na Figura 5.14 mostrando a concordância destes resultados para a região onde foi observado o átomo de cloro no modelo cristalográfico. Uma segunda região indica uma possível substituição que poderia ser feita com outro átomo de cloro possivelmente em meta fazendo com que o anel entre mais profundamente no S2 capturando alguma interação com o Glu208.

Figura 5.13 - Mapeamento do sítio catalítico da cruzaina com GRID usando as sondas OH (A), Flúor (B), Cloro (C) e Bromo (D) com destaque para o sítio catalítico da cruzaina em superfície de Van der Waals. Em amarelo, no centro da figura está destacada a Cys25. A superfície de potencial energético é mostrada em azul com o nível energético de atração para cada sonda destacado em amarelo.

Fonte: Autoria própria

Figura 5.14 - Sobreposição do MOB do Neq0147 (cadeia C) com a superfície de nível de energia à -5 kcal/mol (em verde) obtida para a sonda cloro.

Fonte: Autoria própria

De maneira a buscar por pontos de seletividade em relação à catepsina L uma sondagem com o SuperStar foi realizada. O SuperStar usa informação experimental real derivada tanto do CSD, quanto do PDB e portanto aplica uma abordagem totalmente baseada em conhecimento para produzir seus resultados sendo extensivamente validado em complexos proteína-ligante provenientes do PDB. Os resultados do SuperStar são exibidos como mapas de contorno 3D que retratam a distribuição espacial de probabilidades. Estes mapas representam a

chance de se encontrar o grupo de sondas escolhidas em certas posições do sítio de ligação da proteína ou em torno de uma molécula do ligante. As probabilidades são medidas em relação à possibilidade aleatória de se encontrar um grupo esperado em determinada posição, onde uma probabilidade de quatro significa que a chance de se encontrar aquele grupo de sondas é quatro vezes maior do que o aleatório. Tendências superiores a dois, indicam sítios de interações favoráveis. Pontos farmacofóricos exibem regiões de máxima intensidade nos mapas.

Este experimento foi realizado a partir da estrutura cristalográfica da cruzaína nativa em complexo com a iodoacetamida preparada da mesma maneira da que foi utilizada no programa GRID. Os aminoácidos que margeiam o sítio S2 foram selecionados para a delimitação da região de busca, uma vez que da literatura, se conhece que este subsítio é o principal para definição de seletividade em relação à catepsina L. Uma vez que a cavidade estava definida, os experimentos com as sondas puderam ser executados. O SuperStar trata grupos terminais de proteínas tais como o OH, NH₃ e SH como rotâmeros. Isto significa que para estes grupos os *plots* direcionais do IsoStar são convoluídos com a distribuição de torção, e então, todos os *hotspots*, vizinhos ao grupo são exibidos no mapa final independente da conformação de entrada. O banco de dados utilizado para estas análises foi o PDB em virtude da natureza dos dados, resolução, disponibilidade de sub-estruturas. A seleção de sondas provenientes do PDB é mais limitada do que a do CSD. Por exemplo, a ocorrência de átomos de halogênio em ligantes do PDB é tão pequena, que seu uso como sonda não pode ser justificado. A sonda oxigênio carbonílico, deriva suas informações dos plotes de espalhamento do IsoStar para contatos entre quaisquer grupos C=O, e produz o contorno para o oxigênio.

As sondas mais utilizadas para começar a explorar um sítio de ligação são: oxigênio de álcool onde o grupo hidroxila pode doar ou receber ligações de hidrogênio e portanto, pontos favoráveis para doadores e aceitadores de ligação de hidrogênio serão localizados; oxigênio de carbonila que indica pontos favoráveis para aceitadores de ligação de hidrogênio; NH amina que indica pontos favoráveis para doadores de ligação de hidrogênio; CH alifático e CH aromáticos que indicam regiões favoráveis para grupos apolares e hidrofóbicos. As diferenças estruturais do sítio catalítico entre a cruzaína e a catepsina L são evidenciadas na Figura 5.15.

Da sobreposição exibida na Figura 5.15 fica evidente que as principais diferenças entre os sítios catalíticos das duas enzimas é a presença da Ala214 no

lugar do Glu208 e de uma metionina (Met161) no lugar da Leu160. Assim como discutido na Seção 1.3, a flexibilidade do Glu208 e a possibilidade de realizar diferentes interações nesta porção do sítio, permitem a cruzaína reconhecer uma quantidade maior de sequências de aminoácidos para serem clivadas, fazendo com que esta exiba atividade proteolítica muito maior do que a catepsina L, o que também justifica seus efeitos auto-proteolíticos tão significativos. Outra diferença é a presença de um glutamato (Glu63) no lugar da Ser61 da cruzaína sendo este, outro potencial ponto para busca por seletividade na região do sítio S3.

Figura 5.15 - Sobreposição das estruturas da cruzaína (complexo com iodoacetamida) e catepsina L (PDB 3HHA) com destaque para os principais aminoácidos do sítio.

Fonte: Autoria própria

Os resíduos Gln19 e Trp184 na cruzaína são importantes para interações no sítio S1'. Estes resíduos estão um pouco mais distantes da Cys25 e His162 da tríade em relação ao que é observado para a catepsina L, fazendo com que o sítio S1' na cruzaína tenha um volume um pouco maior. O Trp184 é altamente conservado entre os membros da família da papaína de cisteíno proteases. Acredita-se que este resíduo desempenhe papéis críticos na clivagem do substrato pela orientação da maquinaria catalítica da enzima.¹³ O anél indol do Trp184 ajuda a posicionar a asparagina da tríade catalítica (Asn182) em uma orientação que conduz a interação com a His162 via uma interação do tipo NH- π . Esta interação facilita a formação do par iônico tiolato-imidazol necessária para a catálise. O triptofano desempenha outros papéis importantes na estrutura e função da proteína, no qual acima da

cadeia lateral possui uma grande área superficial hidrofóbica disponível que devido a sua mobilidade e a possibilidade de realizar interações aromáticas (empilhamento π - π do tipo face-face e face-canto) e ligações de hidrogênio, tornam este resíduo chave na energia de ligação entre proteína-proteína e interações proteína ligante.¹⁴ A maior parte dos esforços tem sido realizados para desenvolvimento de novos inibidores baseados na otimização de interações com os sítios de ligação S2 e S3, porém, apesar de ser amplo e superficial o sítio S1' poderia ser melhor explorado principalmente em busca de aumento de potência para futuros inibidores. As Figuras 5.16, 5.17 e 5.18 mostram os resultados da sondagem com o programa SuperStar.

Figura 5.16 - Sondagem do sítio catalítico da cruzaina e da catepsina L com as sondas OH e C=O no SuperStar. Pontilhados em vermelho representam probabilidade duas vezes maior que a aleatoriedade de se encontrar interações favoráveis, em verde probabilidade quatro, em azul probabilidade 8, e o ponto de maior probabilidade é representado pela esfera em vermelho.

Fonte: Autoria própria.

Uma grande probabilidade de que grupos hidroxilas interagem favoravelmente com o Glu208 na base do sítio S2 é demonstrada pela análise com a sonda OH hidroxila (Figura 5.16). Apesar de também existir probabilidade para a mesma região na catepsina L, esta é pelo menos quatro vezes menor, indicando possível ponto para seletividade com grupos OH. A sonda de oxigênio carbonílico reportou uma pequena probabilidade em regiões próximas a Met70 da catepsina L.

Figura 5.17 - Sondagem do sítio catalítico da cruzaina e da catepsina L com a sonda NH amino do SuperStar. Contornos com pontilhados em vermelho representam probabilidade duas vezes maior que a aleatoriedade de se encontrar interações favoráveis, em verde probabilidade quatro, em azul probabilidade 8, e o ponto de maior probabilidade é representado pela esfera em azul.

Fonte: Autoria própria.

Figura 5.18 - Sondagem do sítio catalítico da cruzaina e da catepsina L com sondas CH alifático e aromático no SuperStar. pontilhados em vermelho representam probabilidade duas vezes maior que a aleatoriedade de se encontrar interações favoráveis, em verde probabilidade quatro, em azul probabilidade 8, e o ponto de maior probabilidade é representado pela esfera em verde.

Fonte: Autoria própria.

Maiores pontos de seletividade são observadas na sonda NH amino com em região próxima ao Glu208. No sítio da catepsina L também existe uma pequena

probabilidade para este grupo em região próxima. Isto é devido à presença de uma serina ao lado da Ala214 na base do sítio S2 da catepsina L. Não foram identificados pontos para seletividade com as sondas CH alifático e aromático. Outro mapa interessante do sítio catalítico é o que está exibido na Figura 5.19, onde as superfícies hidrofóbicas, de acessibilidade do solvente (SAS), e de aceptores e doadores de ligações de hidrogênio, foram calculadas no programa DS *vizualizer* (v.4.0) para a estrutura da cruzaina nativa em complexo com o Neq0409.

Figura 5.19 - Análise do sítio da cruzaina nativa em complexo com o Neq0409 (PDB 4QH6) e catepsina L (PDB 3HHA) no DS Vizualiser. Superfície hidrofóbica em **A** e **B**, superfície SAS em **C** e **D**, e mapa com doadores e aceptores de ligações de hidrogênio em **E** e **F**.

Fonte: Autoria própria

A Figura 5.19 mostra que o sítio S2 da cruzaina é basicamente hidrofóbico principalmente devido a Met67, Leu68, Ala138 e Leu160 apresentando, no entanto, o Glu208 e a Gln159 alternam hidrofílicos sendo estes os aminoácidos mais flexíveis do sítio da cruzaina. O sítio S2 da catepsina L não é tão delimitado quanto o da cruzaina principalmente na base onde na cruzaina existe o Glu208. No entanto,

o sítio S3 da catepsina L é muito mais definido e profundo do que o da cruzaina exibindo maiores pontos para interações principalmente por ligações de hidrogênio. O mapa de acessibilidade ao solvente mostra que a região entre a tríade catalítica e o sítio S2 é a mais profunda e definida sendo pouco acessível à moléculas de água (superfície SAS) tanto para a cruzaina quanto para a catepsina L. Existem vários aceptores de ligação de hidrogênio margeando o sítio S1 e S1' embora a superfície plana destes sítios seja constituída na maioria por doadores em ambas as enzimas. A região onde mais se encontram aceptores de ligação de hidrogênio por área é a região vizinha a cisteína catalítica onde estão presentes a cadeia principal da Asp161, Met68 e Gly66. O sítio S2 que interage hidrofóbicamente, também pode aceitar interações de hidrogênio na região do Glu208 na cruzaina e na Ser213 na catepsina L. O sítio catalítico da cruzaina é côncavo e convexo ao mesmo tempo, sendo difícil atingir potência com inibidores que não sejam covalentes. A sobreposição dos modos de interação dos inibidores obtidos neste trabalho com estes mapas obtidos no GRID, SuperStar e DS Vizualiser indicarão os novos grupos químicos que poderão ser adicionados para incremento de potência e seletividade.

5.4 Referências Bibliográficas

- 1 Qu, Z.; Meng, F.; Zhou, H.; Li, J.; Cheng, J.; Greenlief, M.; Gu, Z. NitroDIGE analysis reveals inhibition of protein S-nitrosylation by epigallocatechin gallates in lipopolysaccharide-stimulated microglial cells. **J Neuroinflammation**, v. 11, n. 17, p. 1-13, 2014.
- 2 Ferreira, R. S.; Eidam, O.; Mott, B. T.; Keiser, M. J.; Mckerrrow, J. H.; Maloney, D. J.; Irwin, J. J.; Shoichet, B. K. Complementarity Between a Docking and a High-Throughput Screen in Discovering New Cruzain Inhibitors. **J Med Chem**, v. 53, n. 13, p. 4891-4905, 2010.
- 3 Wiggers, H. J.; Rocha, J. R.; Fernandes, W. B.; Sesti-Costa, R.; Carneiro, Z. A.; Cheleski, J.; Da Silva, A. B. F.; Juliano, L.; Cezari, M. H. S.; Silva, J. S.; Mckerrrow, J. H.; Montanari, C. A. Non-peptidic Cruzain Inhibitors with Trypanocidal Activity Discovered by Virtual Screening and In Vitro Assay. **PLoS NTD**, v. 7, n. 8, p. e2370, 2013.
- 4 Eidam, O.; Romagnolib, A.; Dalmassoc, G.; Barelier, S.; Caselli, E.; Bonnet, R.; Shoichet, B. K.; Prati, F. Fragment-guided design of subnanomolar β -lactamase inhibitors active in vivo. **PNAS**, v. 11, p. 1-6, 2012.
- 5 Teotico, D. G.; Babaoglu, K.; Ferreira, R. S.; Giannetti, A. M.; Shoichet, B. K. Docking for fragment inhibitors of AmpC lactamase. **PNAS**, v. 106, n. 18, p. 7455–7460, 2009.
- 6 Babaoglu, K.; Simeonov, A.; Irwin, J. J.; Nelson, M. E.; Jadhav, A.; Inglese, J.; Austin, C. P.; Shoichet, B. K. Comprehensive Mechanistic Analysis of Hits from High-Throughput and Docking Screens against Lactamase. **J Med Chem**, v. 51, n., p. 2502–2511, 2008.

- 7 Durrant, J. D.; Keranen, H.; Wilson, B. A.; Mccammon, J. A. Computational Identification of Uncharacterized Cruzain Binding Sites. **PLoS NTD**, v. 4, n. 1, p. e676, 2010.
- 8 Judice, W. A. S.; Cezari, M. H. S.; Scharfstein, J.; Chagas, J. R.; Tersario, I. L. S.; Juliano, M. A.; Juliano, L. Comparison of the specificity, stability and individual rate constants with respective activation parameters for the peptidase activity of cruzipain and its recombinant form, cruzain, from *Trypanosoma cruzi*. **Eur J Biochem**, v. 268, p. 6578–6586, 2001.
- 9 Nery, M. A. J., Scharfstein; Juliano, L. Characterization of the substrate specificity of the major cysteine protease (cruzipain) from *Trypanosoma cruzi* using a portion-mixing combinatorial library and fluorogenic peptides. **Biochem J**, v. 323, n., p. 427–433, 1997.
- 10 Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. **JMedChem**, v. 28, p. 849–857, 1985.
- 11 Gohlke, H. Knowledge-based scoring function to predict protein–ligand interactions. **J Mol Biol**, v. 295, n., p. 337–356, 2000.
- 12 Verdonk, M. L. SuperStar: a knowledge based approach for identifying interaction sites in proteins. . **J Mol Biol**, v. 289, n., p. 1093–1108, 1999.
- 13 Mcgrath, M. E. The Lysosomal Cysteine Proteases. **Annu Rev Biophys Biomol Struct**, v. 28, n., p. 181–204, 1999.
- 14 Bogan, A. A.; Thorn, K. S. Anatomy of hot spots in protein interfaces. **J Mol Biol**, v. 280, n., p. 1-9, 1998.

Capítulo 6

Conclusões e Perspectivas

Este capítulo apresenta as conclusões desta tese de doutorado que envolveu fundamentalmente o uso da cristalografia de raios X para determinação do modo de interação de inibidores reversíveis da cisteína protease cruzaina, e o uso desta informação estrutural para mapeamento do sítio catalítico em prol da proposição de novos inibidores mais potentes e seletivos. Adicionalmente, são apresentadas as perspectivas futuras para a continuação deste trabalho.

6 CONCLUSÕES E PERSPECTIVAS

6.1 Conclusões

A cristalografia de raios X é um método poderoso no estudo da função biológica e determinação do MOB de um ligante em uma proteína, tendo importante papel na descoberta de novos fármacos. Novos desafios têm surgido na elucidação estrutural devido a sistemas biológicos cada vez mais complexos que vem sendo estudados. Muitos destes desafios surgiram neste projeto de determinação estrutural do MOB de inibidores reversíveis em complexo com a cruzaina devido à auto-proteólise que provoca desordem intrínseca e flexibilidade em várias regiões da proteína durante a cristalização levando a cristais frágeis e de baixo poder de difração. No entanto, a determinação do MOB para os inibidores deste alvo validado para o tratamento da doença de Chagas, é essencial para o desenvolvimento de novos fármacos. Neste trabalho, estratégias foram aplicadas para a obtenção de estruturas da cruzaina em complexo com inibidores de menor afinidade (porém maiores LE e LLE_{AT}) do tipo fragmentos e inibidores reversíveis. Estas estratégias incluíram o desenvolvimento das enzimas com mutação sítio dirigido C25A e C25S, diferentes formas de preparo da enzima e ensaios de cristalização, aplicação de estratégias de coleta, processamento e refinamento de dados à média resolução.

Da série de compostos inicialmente testada, foram obtidas duas estruturas cristalográficas na cruzaina nativa à média resolução para os inibidores Neq0409 e Neq0176 (obtido por simplificação molecular do composto mais potente no ensaio celular contra o *T. cruzi*, o Neq0042). Este MOB obtido validou o mecanismo de inibição não covalente da série de inibidores não peptídicos testados, sendo este importante resultado, publicado na revista eletrônica Plos NTD. O MOB dos fragmentos Neq0147 selecionados por FBDD foi determinado na enzima C25S mutante, que juntamente com a estrutura para o Neq0176, representam as primeiras estruturas cristalográficas de fragmentos de baixa afinidade com uma protease da família da papaína. O MOB determinado para o K777 na enzima C25S foi um importante resultado para a validação da C25S como modelo apropriado para determinação do MOB de ligantes de baixa afinidade e translação de resultados para a enzima nativa. Outro importante resultado advém do MOB obtido para o inibidor protótipo do tipo dipeptidil-nitrila Neq0409 na cruzaina nativa, validando

ortogonalmente a hipótese do uso de dipeptidil-nitrilas como inibidores covalentes reversíveis da cruzaina que atuam competitivamente em seu sítio catalítico. O Neq0409 tem servido como esqueleto base para o planejamento de uma série de potentes dipeptidil-nitrilas que vem sendo sintetizada e avaliada principalmente para sondar interações via SAR nos sítios S2 e S3 da cruzaina em busca de incrementos de potência e seletividade principalmente em relação à homóloga em humanos, a catepsina L, sendo esta uma das etapas principais dentro do projeto temático nº13/18009-4 da FAPESP que se iniciou em maio de 2014 com o objetivo principal de identificar substâncias tripanocidas com propriedades que as tornem potenciais candidatas a fármacos.

Adicionalmente a estes resultados, foram obtidas duas estruturas cristalográficas de alta resolução para a cruzaina nativa co-cristalizada com os pequenos inibidores de cisteína protease MMTS e iodoacetamida. Estas estruturas reforçam a relação entre a estabilização do efeito proteolítico da cruzaina e o incremento na qualidade e resolução dos cristais produzidos. As estruturas cristalográficas obtidas a média resolução para os inibidores de interesse na cruzaina C25S são possivelmente decorrentes do efeito proteolítico residual do processo de trans-ativação, em que traços da cruzaina nativa foram utilizados para clivagem e digestão do prodomínio. Outro motivo que elucida a fragilidade e baixo poder de difração dos cristais obtidos, e ainda a grande quantidade de moléculas na ASU, é decorrente do tipo de inibidores utilizados. Estes são pequenos e interagem de maneira não covalente apresentando muitos graus de liberdade rotacionais e conformacionais dentro do sítio grande, superficial e exposto ao solvente da cruzaina, levando a variantes conformacionais em diferentes cadeias da ASU, conforme o caso do Neq0176 na enzima nativa também evidenciada por experimentos de dinâmica molecular. Estas variantes são causa ou consequência das ligeiras diferenças conformacionais entre as cadeias refletindo no seu arranjo supramolecular que privilegia interações específicas entre cadeias levando a empacotamentos pouco densos e grande conteúdo de solvente nos cristais. Adicionalmente a estes efeitos, provavelmente as condições de cristalização obtidas não eram as ideais e a falta do uso de *seeding* talvez tenha sido o motivo para cristais com geminação, patologias cristalográficas e outros problemas encontrados para os inibidores de interesse. Estas estruturas obtidas estão validando a qualidade dos métodos de quimio informática que estão sendo utilizados na seleção de

inibidores e ainda, levando a aspectos essenciais para o melhor entendimento da enzima e desenvolvimento de novos inibidores.

As sondagens realizadas no sítio catalítico da enzima junto com a análise os dados estruturais obtidos neste trabalho, trazem valiosas informações para o planejamento e síntese novos inibidores mais potentes e seletivos. As construções recentes da cruzaína tanto nativa quanto as mutantes C25S e C25A possibilitarão a realização de experimentos utilizando os métodos biofísicos SPR, RMN e calorimétricos ITC e DSC antes limitados, uma vez que é possível obter uma grande quantidade de proteína pura e ativa para estes ensaios que aliadas com a enzima mutante, possibilitaram um melhor estudo com ligantes fracos como fragmentos moleculares. As novas estratégias de cristalização aqui desenvolvidas e testadas (além das condições de cristalização testadas) poderão ser continuadas para a cristalização de uma série de novos potentes inibidores que vem sendo sintetizados no NEQUIMED em ordem de validar ortogonalmente os MOB e identificar regiões importantes para incremento de potência e seletividade.

6.2 Perspectivas

Como perspectivas de continuidade deste projeto, espera-se dominar o efeito proteolítico da enzima, encontrar condições de cristalização otimizadas utilizando também *seeding* para a obtenção de melhores cristais com inibidores reversíveis de interesse, além da evolução dos atuais inibidores cujo MOB foi determinado e dos novos inibidores que estão sendo sintetizados. Como sugestões de trabalhos futuros dando continuidade a este projeto e aos resultados promissores obtidos, sugere-se:

- Otimização das condições de cristalização da cruzaína C25S forma APO obtidas neste trabalho, para futuros experimentos de *soaking* com inibidores reversíveis do tipo fragmentos. A realização do *soaking* com um *cocktail* de fragmentos de esqueletos químicos diversos poderia sugerir novos esqueletos químicos tal como os obtidos para o Neq0147 e para o Neq0176 como propostas para substituir os atuais esqueletos peptídeo-miméticos;
- O grupo carboxílico do Neq0147 mostrou-se ser para a cruzaína C25S equivalente ao que o *warhead* nitrila é para a cruzaína nativa. Portanto, poder-se-ia realizar um planejamento de novos ligantes, via métodos de quiminformática, utilizando a C25S como alvo, e estes poderiam ser identificados em métodos biofísicos como o SPR que juntamente com a

cristalização na C25S, indicariam regiões com *hotspots* para interações e sugestão de novos ligantes;

- Produzir cristais da enzima nativa estabilizada com moléculas pequenas tais como o MMTS e Iodoacetamida para ensaios com os inibidores de interesse por *soaking*;
- Otimizar as estratégias de coleta de dados e também os tratamentos pós-cristalização incluindo a desidratação dos cristais, crio proteção e outros;
- Desenvolver a etapa de remoção do prodomínio da cruzaina C25A;
- Desenvolver a produção e cristalização para as novas construções da cruzaina C25S e C25A clonadas com cauda de histidina C-terminal;
- Realizar estudos supramoleculares das estruturas obtidas em busca da possibilidade de dimerização, sítios alostéricos, flexibilidade de alças incomuns e outros;
- Selecionar novos inibidores baseados nos MOB dos inibidores já obtidos;
- Utilizar de compostos com o *warhead* nitrila para ensaios de cristalização na cruzaina nativa para a sondagem de grupos nos sítios S2 e S3 por cristalografia;
- Identificar outras proteases que apresentam similaridade estrutural e funcional com a cruzaina e construir um banco de dados de inibidores de cisteíno-proteases;
- Sondar interações utilizando os compostos baseados em nitrilas atualmente sintetizados em nosso laboratório e a cristalografia de raios X para a cruzaina nativa;
- Com base em todas as informações estruturais advindas da determinação do MOB para as dipeptidil-nitrilas e para as nitrilas derivadas do Neq0147 na cruzaina nativa, além do MOB para os fragmentos em complexo com a cruzaina C25S mutante que serão obtidas ao longo do projeto, pretende-se em conjunto com as informações de relações estrutura-atividade (SAR) disponíveis e os métodos virtuais de planejamento molecular, propor novos inibidores que serão sintetizados e bioquimicamente avaliados dentro do ciclo planeja-sintetiza-testa-avalia do método baseado em hipóteses, dando continuidade a execução do projeto temático da FAPESP.

ANEXO I

OPEN ACCESS Freely available online

PLOS NEGLECTED TROPICAL DISEASES

Non-peptidic Cruzain Inhibitors with Trypanocidal Activity Discovered by Virtual Screening and In Vitro Assay

Helton J. Wiggers^{1,2}, Josmar R. Rocha², William B. Fernandes^{2,3}, Renata Sesti-Costa⁴, Zumira A. Carneiro⁴, Juliana Cheleski², Albérico B. F. da Silva², Luiz Juliano⁵, Maria H. S. Cezari⁵, João S. Silva⁴, James H. McKerrow^{3*}, Carlos A. Montanari^{1,2*}

1 Departamento de Química, Universidade Federal de São Carlos, São Carlos, São Paulo, Brazil, **2** Grupo de Química Medicinal do IQSC/USP, Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, São Paulo, Brazil, **3** University of California San Francisco, Center for Discovery and Innovation in Parasitic Diseases (CDIPD), Department of Pathology, San Francisco, California, United States of America, **4** Departamento de Bioquímica e Imunologia, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brazil, **5** Departamento de Biofísica, Escola Paulista de Medicina, Universidade Federal de São Paulo, São Paulo, Brazil

Abstract

A multi-step cascade strategy using integrated ligand- and target-based virtual screening methods was developed to select a small number of compounds from the ZINC database to be evaluated for trypanocidal activity. Winnowing the database to 23 selected compounds, 12 non-covalent binding cruzain inhibitors with affinity values (K_i) in the low micromolar range (3–60 μ M) acting through a competitive inhibition mechanism were identified. This mechanism has been confirmed by determining the binding mode of the cruzain inhibitor Nequimed176 through X-ray crystallographic studies. Cruzain, a validated therapeutic target for new chemotherapy for Chagas disease, also shares high similarity with the mammalian homolog cathepsin L. Because increased activity of cathepsin L is related to invasive properties and has been linked to metastatic cancer cells, cruzain inhibitors from the same library were assayed against it. Affinity values were in a similar range (4–80 μ M), yielding poor selectivity towards cruzain but raising the possibility of investigating such inhibitors for their effect on cell proliferation. In order to select the most promising enzyme inhibitors retaining trypanocidal activity for structure-activity relationship (SAR) studies, the most potent cruzain inhibitors were assayed against *T. cruzi*-infected cells. Two compounds were found to have trypanocidal activity. Using compound Nequimed42 as precursor, an SAR was established in which the 2-acetamidothiophene-3-carboxamide group was identified as essential for enzyme and parasite inhibition activities. The IC_{50} value for compound Nequimed42 acting against the trypomastigote form of the *Tulahuén* *lacZ* strain was found to be 10.6 ± 0.1 μ M, tenfold lower than that obtained for benznidazole, which was taken as positive control. In addition, by employing the strategy of molecular simplification, a smaller compound derived from Nequimed42 with a ligand efficiency (LE) of $0.33 \text{ kcal mol}^{-1} \text{ atom}^{-1}$ (compound Nequimed176) is highlighted as a novel non-peptidic, non-covalent cruzain inhibitor as a trypanocidal agent candidate for optimization.

Citation: Wiggers HJ, Rocha JR, Fernandes WB, Sesti-Costa R, Carneiro ZA, et al. (2013) Non-peptidic Cruzain Inhibitors with Trypanocidal Activity Discovered by Virtual Screening and In Vitro Assay. PLoS Negl Trop Dis 7(8): e2370. doi:10.1371/journal.pntd.0002370

Editor: Carlos A. Buscaglia, Instituto de Investigaciones Biotecnológicas, Argentina

Received: May 16, 2012; **Accepted:** July 4, 2013; **Published:** August 22, 2013

Copyright: © 2013 Wiggers et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Funding to carry out this work was provided by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) through the granting award number 2011/01893-3, which can be accessed at <http://internet.aquila.fapesp.br/agilis/publico/>. We are also indebted to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for providing with the grant from project number 301614/2010-5 (www.cnpq.br). In addition, the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, <http://www.capes.gov.br/>) is also acknowledged for sponsoring William B. Fernandes (Programa de Bolsas de Doutorado Sanduíche no Exterior - PDSE, Sistema de Concessões de Bolsa no Exterior - BEX process 5985/11-0) to carry out part of this work at UCSF. The sponsors had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: James.McKerrow@ucsf.edu (JHM); Carlos.Montanari@usp.br (CAM)

Introduction

Chagas disease, widespread in Latin America, is caused by the kinetoplastid protozoan parasite *Trypanosoma cruzi*. Despite efforts to reduce the transmission of the parasite by controlling the hematophagous triatomine insect vector, the World Health Organization estimates that 10 million people are infected worldwide, with another 25 million at risk. Most cases are in Latin America, where Chagas disease is endemic, but it is also found in Canada, the United States, Europe (mainly in Spain and Portugal), Japan and Australia [1–3].

T. cruzi's complex life cycle involves two replicative forms: the epimastigote, in the gut of the insect vector, and the amastigote, an intracellular form in the infected mammal. The two infective non-replicative forms are the metacyclic trypomastigote in the insect vector and the bloodstream trypomastigote released from infected cells into the blood of the mammal [4].

Chagas disease has an acute phase and a chronic latent phase. The acute phase, which occurs shortly after infection, lasts for a few weeks or months, whereas the chronic phase develops over many years. The acute phase may not be noticed because it is symptom-free or exhibits only mild symptoms that are not unique

ANEXO II

Resumo do manuscrito “*under consideration*” para publicação no periódico Plos NTD:

Design and Synthesis of Dipeptidyl Nitrile Cruzain Inhibitors and Their Evaluation as Potential Trypanocidal Agents

Leandro AA Avelar; Cristian D Camilo; Sergio de Albuquerque; William B Fernandes; Cristiana Gonzalez; Peter W Kenny; Andrei Leitão; James H McKerrow; Erika VM Orozco; Jean FR Ribeiro; Josmar R Rocha; Fabiana Rosini; Marta E Saidel; Carlos A Montanari.

Abstract

A series of compounds based on the dipeptidyl nitrile scaffold were synthesized and assayed for their inhibitory activity against the *T. cruzi* cysteine protease cruzain. Structure activity relationships (SARs) were established using three, eleven and twelve variations respectively at the P1, P2 and P3 positions. A K_i value of 16 nM was observed for the most potent of these inhibitors which reflects a degree of non-additivity in the structure-activity relationships. An X-ray crystal structure was determined for the ligand-protein complex for the structural prototype for the series. Twenty three inhibitors were also evaluated for their anti-trypanosomal effects and an EC_{50} value of 28 μ M was observed for the most potent of these. Although there remains scope for further optimization, the knowledge gained from this study is also transferable to the design of cruzain inhibitors based on warheads other than nitrile as well as alternative scaffolds.