**Resolução Estrutural Cristalográfica de Proteases
de HIV-1 de Subtipos Brasileiros**

&

**Estudos Estruturais da L-Asparaginase de
*Escherichia coli***

USP/IFSC/SBI

8-2-001647

MARIO SANCHES MATILDE JUNIOR

Tese apresentada ao Instituto de Física de
São Carlos, da Universidade de São Paulo,
para obtenção do título de Doutor em
Ciências: Física Aplicada opção Física Bio-
molecular.

Orientador: Prof. Dr. Igor Polikarpov

**São Carlos
2004**

IFSC-USP SERVIÇO DE BIBLIOTECAS
INFORMACAO

Junior, Mario Sanches Matilde

“Resolução Estrutural cristalográfica de proteases de HIV – 1 de subtipos brasileiros & Estudos Estruturais da L – Asparaginase de *Escherichia coli*.”
Mario Sanches Matilde Junior – São Carlos, 2004

Tese (Doutorado) – Física Aplicada - Instituto de Física de São Carlos
2004 - Páginas: 139

Orientador: Prof. Dr. Igor Polikarpov

1. Cristalografia de proteínas, protease HIV – 1, L – asparaginase

I. Título

MEMBROS DA COMISSÃO JULGADORA DA TESE DE DOUTORADO DE MARIO SANCHES
MATILDE JUNIOR APRESENTADA AO INSTITUTO DE FÍSICA DE SÃO CARLOS,
UNIVERSIDADE DE SÃO PAULO, EM 17/12/2004.

COMISSÃO JULGADORA:

Prof. Dr. Igor Polikarpov (Orientador e Presidente) – IFSC/USP

Prof. Dra. Débora Foguel – UFRJ

Dra. Maria Cristina Nonato Costa – FCFRP/USP

Prof. Dr. Marcos Roberto de Mattos Fontes – UNESP

Prof. Dr. Richard Charles Garratt – IFSC/USP

à minha esposa Mária

Agradecimentos

É difícil a tarefa de dar o devido reconhecimento a todas as pessoas que contribuíram para que esse projeto fosse possível, e começo esses agradecimentos já com o sentimento de estar sendo injusto com alguns. Assim, gostaria de antecipar meus agradecimentos a toda a qualquer pessoa que tenha me ajudado de alguma maneira durante a realização desse trabalho.

Em primeiro lugar gostaria de agradecer a todos, funcionários, professores, colegas de laboratório, do Grupo de Cristalografia do IFSC pela ajuda que de alguma maneira me foi prestada. Sem os esforços, pequenos ou grandes, de cada um de vocês nada disso teria sido possível. Agradeço também ao Igor pelas várias oportunidades, inclusive a de poder estagiar em Frederick e acumular experiências num laboratório diferente. A interação com os outros professores do grupo, Glaucius, Richard e Otávio, também foi muito importante na minha formação e na realização desse projeto.

É com prazer que agradeço à Adriana por compartilhar comigo várias experiências e por me agüentar por quase quatro anos reclamando e resmungando na mesa ao lado. Com ela tive conversas altamente produtivas e algumas idéias interessantes, que me mostraram vários caminhos durante as encruzilhadas que enfrentei durante esse projeto. Queria agradecer também à Sandra e ao André (casal Chuchu!), pelas discussões, tanto científicas quanto de qualquer outro gênero que sempre tivemos, além, é claro, pela amizade inestimável.

Algumas pessoas tiveram algumas idéias que foram bastante importante em algum momento, como o Alexander, que me ajudou com o passo final de reenovelamento e o Ricardo Aparício, que foi a primeira pessoa a me propôr a idéia de abaixar a simetria do grupo, que finalmente se revelou a resposta para o nosso problema de refinamento. Agradeço também à Nádia, que além de participar ativamente desse projeto, foi a primeira orientada da minha carreira.

Gostaria de agradecer também a amizade de várias pessoas do grupo, ou que passaram por aqui: Alécio, Alexandre (Vermão) e Ney (Diabão), Artur, Carol (Chatinha), Carol (Gatinha), Castilho, Eliana Grimm (#*&!@), Fred, Humberto, João Renato, Lucas, Marcos Navarro, Marcos Calgaro, Maria Auxiliadora (Dona Maria), Monique, Nádia, Rafael (Caqui), Rodrigo Portugal, Ronaldo, Susana e Wânus.

Não poderia deixar de agradecer ao pessoal do MCL, em Frederick. Acredito que o esforço pessoal do Alex e da Alla foram fundamentais para o sucesso desse projeto, e me considero em débito com eles. Também gostaria de deixar meus agradecimentos ao Dr. John Louis, de Bethesda, que infelizmente não pude conhecer pessoalmente, mas me forneceu a primeira experiência de orientação por telefone e email (com sucesso!) da minha vida, através da qual pude obter grande parte das informações sobre a produção das proteases, incluindo sobre a mutação Q7K.

Algumas pessoas, como Istvan Botos, Jerry Alexandratos, Jiri Vondrasek, Sergei

Pletnev e Jim e Susan Matheson, foram responsáveis por fazer com que nos sentíssemos (Mábia e eu) em casa num país estrangeiro. Agradeço também a alguns pesquisadores, dentro ou fora do MCL, como Alexander Zdanov (Sasha), Elena Gustchina, Eugenia Magracheva, Gary Laco, Istvan Botos, Jacek Lubkowski, Jerry Alexandratos, Mi Li, Sergei Pletnev, Sergei Gulnik, que estiveram de alguma maneira engajados em nos ajudar com esse projeto, fornecendo protocolos, analisando resultados e definindo experimentos.

Por último, mas não menos importante, gostaria de agradecer ao pessoal da antiga CPr da nossa época do Síncrotron, Brandão, Carlos Basílio, Katucha, Fábio Cupri, Lee, Paula, Ricardo, Ronaldo, Sandra Krauchenko, Sandra Martha e Valéria. Vocês foram importantes no começo e continuam sendo ainda hoje.

A todos vocês meu muito obrigado.

Conteúdo

I	Proteases de HIV	1
1	Introdução	2
1.0.1	Organização e Replicação do Vírus	3
1.1	Protease de HIV	5
1.1.1	Mecanismo de Ação	8
1.1.2	Inibidores	10
1.2	Subtipos e Mutações	13
1.3	Objetivos	19
2	Metodologia	21
2.1	Subclonagem	21
2.1.1	Reações de PCR	21
2.1.2	Reações de Digestão	24
2.1.3	Reações de Ligação	26
2.2	Expressão, Purificação e Reenovelamento	27
2.2.1	Lise Celular e Lavagem dos Corpos de Inclusão	28
2.2.2	Colunas de Troca Iônica	29
2.2.3	Reenovelamento das Proteases	29
2.3	Co-cristalização	30
2.3.1	<i>Seeding</i>	32
2.3.2	Desidratação dos Cristais	32
2.4	Coleta e Processamento de Dados	32
2.5	Resolução Estrutural	33
2.6	Refinamento	34
2.7	Análise e Comparação Estruturais	35
3	Resultados e Discussão	37
3.1	Subclonagem	37
3.2	Obtenção das Proteases de HIV	41
3.2.1	Ensaio Preliminares	41
3.2.2	Mutantes Q7K	43
3.3	Purificação e Reenovelamento das Proteases	48
3.4	Co-cristalização	52
3.4.1	Complexo <i>Bmut</i> :TL-3	52
3.4.2	Complexo <i>Fwt</i> :TL-3	54
3.4.3	Complexo <i>Fmut</i> :TL-3	55
3.4.4	Complexos com Outros Inibidores	57

3.5	Coleta de Dados e Resolução Estrutural	59
3.6	Análise e Comparação Estruturais	69
3.6.1	Comparação Global das Quatro Estruturas com TL-3	71
3.6.2	Interações entre TL-3 e PRs no Sítio Ativo	75
3.6.3	Diferenças Entre Subtipos B e F	85
3.6.4	Modificações no Inibidor	90
 II L-asparaginase		93
4	Introdução	94
4.1	Objetivos	96
5	Metodologia	97
5.1	Cristalização, Coleta de Dados e Refinamento	97
5.2	Análise e Comparação Estruturais	97
6	Resultados e Discussão	98
6.1	Comparações das Amidohidrolases	98
6.1.1	Interações com o Substrato	108
6.1.2	Especificidade das Amidohidrolases	109
 III Conclusões Gerais		111
7	Conclusões e Perspectivas	112
A	Artigo Publicado - HIV PR	135
B	Artigo Publicado - L-asparaginase	139

Lista de Figuras

1.1	Organização do genoma do vírus HIV-1.	4
1.2	Elementos de estrutura secundária da protease de HIV-1.	6
1.3	Interações HIV PR-Substrato.	9
1.4	Mecanismo de reação das proteases de HIV.	10
1.5	Isósteros não hidrolisáveis da ligação peptídica.	11
1.6	Inibidor não peptídico da protease de HIV-1.	13
1.7	Inibidores de protease comerciais.	14
1.8	Distribuição geográfica das subtipos.	16
1.9	Padrão de resistências por inibidores.	18
2.1	Programação das reações de PR.	23
2.2	Vetor de expressão pET11a.	25
3.1	Subclonagem dos genes das proteases em pCR4Blunt-TOPO.	38
3.2	Subclonagem dos genes das proteases com inclusão da mutação Q7K.	40
3.3	Análise das amostras dos ensaios de expressão.	42
3.4	Ensaio de expressão com a protease <i>Bmut</i>	44
3.5	Análise dos ensaios de expressão das proteases contendo com a mutação Q7K.	46
3.6	Ensaio de expressão das proteases em bactérias CodonPlus.	47
3.7	Purificação da protease <i>Bmut</i> pela coluna Q-sepharose.	49
3.8	Purificação da protease <i>Bmut</i> pela coluna SP-sepharose.	50
3.9	SDS-PAGE da purificação da protease <i>Bmut</i> por troca iônica.	51
3.10	Estrutura do inibidor universal TL-3.	52
3.11	Cristais da protease <i>Bmut</i> complexada com o inibidor TL-3.	53
3.12	Cristais da protease <i>Fwt</i> complexada com o inibidor TL-3.	54
3.13	Cristais do complexo <i>Fmut</i> :TL-3.	56
3.14	Estruturas dos três inibidores não-comerciais.	57
3.15	Cristais do inibidor indinavir.	58
3.16	Padrões de difração do co-cristal <i>Bmut</i> :TL-3.	60
3.17	Valores de R_{factor} e R_{free} para cada iteração.	62
3.18	Inibidor TL-3 modelado como meio ligante.	66
3.19	Representações da região do sítio ativo ($P6_1$).	68
3.20	Refinamento da cisteína modificada 67.	69
3.21	Quebra de simetria cristalográfica entre as subunidades A e B.	70
3.22	Comparação da ligação peptídica P3-P4 no complexo <i>Bwt</i> :TL-3.	71
3.23	Alinhamento das seqüências das proteases <i>Bwt</i> , <i>Bmut</i> , <i>Fwt</i> e <i>Fmut</i>	72
3.24	Vista frontal e do topo das proteases de HIV-1.	73
3.25	Gráfico de B fatores por resíduo das quatro estruturas resolvidas.	75

3.26	Superposição das coordenadas do inibidor TL-3.	76
3.27	Representação estereográfica dos bolsões S1 e S1'.	78
3.28	Representação estereográfica dos bolsões S2 e S2'.	80
3.29	Representação estereográfica dos bolsões S3 e S3'.	82
3.30	Representação estereográfica dos bolsões S4 e S4'.	84
3.31	Representação estereográfica dos resíduos da dobradiça do <i>flap</i>	86
3.32	Representação estereográfica das interação do resíduo Leu90.	88
3.33	Propostas de modificações no inibidor TL-3.	92
4.1	Representação de uma das subunidades da L-asparaginase de <i>E. coli</i>	95
6.1	Alinhamento de sequência de cinco amidohidrolases.	100
6.2	Sítio ativo da L-asparaginase de <i>E. coli</i> (EcA).	101
6.3	Alinhamento estrutural da região do sítio ativo das amidohidrolases analisadas.	103
6.4	Representação esquemática do mecanismo de reação.	106
6.5	Triades catalíticas envolvidas na reação das amidohidrolases.	107

Lista de Tabelas

2.1	Seqüências dos oligonucleotídeos utilizados nas subclonagens.	22
2.2	Oligos utilizados nas subclonagens com a mutação Q7K.	22
2.3	Reagentes utilizados nas reações de PCR.	23
2.4	Reagentes das reações de digestão.	24
2.5	Reagentes das reações de desfosforilação do vetor digerido.	26
2.6	Reagentes utilizados nas reações de ligação convencional.	27
2.7	Ensaio preliminar de cristalização para as proteases de HIV.	31
3.1	Dados cristalográficos para os conjuntos processados em P6 ₁ 22.	61
3.2	Dados cristalográficos para os conjuntos processados em P6 ₁	65
3.3	Rmsd calculado entre as estruturas <i>Bwt</i> e <i>Fwt</i> e todas as outras.	72
6.1	Estruturas de amidohidrolases utilizadas nas comparações.	99

Lista de Abreviaturas

AGA	glutaminase-asparaginase de <i>Acinetobacter glutaminasificans</i>
AIDS	Síndrome da Imunodeficiência Adquirida
<i>Bmut</i>	Protease de HIV Subtipo B Mutante
<i>Bwt</i>	Protease de HIV Subtipo B Selvagem
CA	protéina do Capsídio
DON	6-diazo-5-oxo-L-norleucina
DONV	5-diazo-4-oxo-L-norvalina
EcA	asparaginase de <i>Escherichia coli</i>
ErA	asparaginase de <i>Erwinia chrysanthemi</i>
<i>env</i>	poliproteína contendo as proteína do envelope
<i>Fmut</i>	Protease de HIV Subtipo F Mutante
<i>Fwt</i>	Protease de HIV Subtipo F Selvagem
<i>gag</i>	poliproteína contendo proteínas estruturais
<i>gag-pol</i>	poliproteína contendo as proteínas enzimáticas
HAART	<i>highly active antiretroviral therapy</i>
HIV	Virus da Imunodeficiência Humana
HIV PR	Protease de HIV
IN	Integrase
IPTG	isopropil- β -D-tiogalactopiranoside
LNLS	Laboratório Nacional de Luz Síncrotron
MA	protéina da Matriz
NC	protéina do Nucleocapsídeo
NCS	Simetria Não Cristalográfica
Nef	protéina de acessório
PCR	<i>Polimerase Chain Reaction</i>
pET3a_bwt	clone da protease <i>Bwt</i> (fusão 10 aa) em pET3a
pET11a_bmut	clone da protease <i>Bmut</i> em pET11a
pET11a_fmud	clone da protease <i>Fmut</i> em pET11a
pET11a_fwt	clone da protease <i>Fwt</i> em pET11a
pET11a_bmutq7k	clone da protease <i>Bmut</i> com a mutação Q7K em pET11a
pET11a_bwtq7k	clone da protease <i>Bwt</i> (Q7K sem fusão) em pET11a
pET11a_fmudq7k	clone da protease <i>Fmut</i> com a mutação Q7K em pET11a
pET11a_fwtq7k	clone da protease <i>Fwt</i> com a mutação Q7K em pET11a
PGA	glutaminase-asparaginase de <i>Pseudomonas</i> 7A
PGA-SO ₄ ⁻²	glutaminase-asparaginase de <i>Pseudomonas</i> 7A com sulfato
PR	Protease
Rev	protéina de acessório, regulatória
rmsd	Desvio Médio Quadrático

RT	Transcriptase Reversa
SAP	Fosfatase Alcalina de Camarão
SU	proteína de Superfície ou gp120
Tat	proteína de acessório, regulatória
TFR	Tampão Fosfato de Re-enovelamento
TLA	Tampão de Lise A
TM	proteína Transmembrânica ou gp41
Vif	proteína de acessório
Vpr	proteína de acessório
Vpu	proteína de acessório, regulatória
WsA	asparaginase de <i>Wolinella succinogenes</i>

Resumo

Um dos maiores problemas encontrados em terapias anti-retrovirais contra AIDS é a emergência de variantes virais que exibem resistência aos fármacos disponíveis e subtipos virais naturalmente mais suscetíveis ao desenvolvimento de falha terapêutica. Neste trabalho nós resolvemos as estruturas cristalográficas de quatro proteases de HIV-1 complexadas com o inibidor universal C2 simétrico TL-3. As proteases utilizadas foram extraídas de pacientes com AIDS, uma do subtipo B selvagem (*Bwt*), uma do subtipo F selvagem (*Fwt*), e um mutante de cada um dos subtipos (*Bmut* e *Fmut*), esses dois últimos de pacientes apresentando falha terapêutica. As proteínas foram produzidas por expressão heteróloga em bactérias *Escherichia coli*, purificadas à partir das proteínas dos corpos de inclusão, e reenoveladas. Os dados coletados foram processados à uma resolução de 2.1 Å para o complexo *Bwt*:TL-3, 1.75 Å para *Bmut*:TL-3, 2.1 Å para *Fwt*:TL-3 e 2.8 Å para *Fmut*:TL-3. Esses dados foram inicialmente processados em P6₁22 e, posteriormente, reprocessados em P6₁. As estruturas foram resolvidas por substituição molecular utilizando a estrutura de uma protease de HIV-1 publicada como modelo. A análise dessas quatro estruturas mostrou que o inibidor TL-3 liga-se de maneira muito próxima em todas elas. Nas proteases *Bmut* e *Fmut* a mutação V82A causa um reempacotamento do bolsão S1', que se reflete num rearranjo da cadeia lateral do inibidor em P1'. Nossas análises indicaram, ainda, que algumas substituições polimórficas entre os subtipos B e F podem levar à maior estabilização de regiões naturalmente flexíveis nas proteases do subtipo F, originando uma enzima intrinsecamente menos ativa e resistente à alguns inibidores. Nas proteases *Fwt* e *Fmut*, a substituição polimórfica M36I causa um deslocamento do *loop* entre os resíduos 35–41, que levaria à perda de flexibilidade dos *flaps* e do *loop* 76–83 no sítio ativo. Nossas comparações indicaram também que a substituição L89M nos subtipos não B pode ser equivalente à mutação de resistência L90M em subtipos B. Por fim, esses dados estruturais nos permitiram sugerir modificações na estrutura do inibidor TL-3 que poderiam levar à um aumento da sua potência. Nesta tese também apresentamos os dados de resolução e análise estruturais da L-asparaginase de *Escherichia coli*, resolvida à uma resolução de 1.95 Å no grupo espacial C2. Baseados em dados estruturais e cinéticos propusemos um mecanismo de reação geral para as amidohidrolases, que incluem as L-asparaginases, através da formação de duas tríades catalíticas Thr-Tyr-Glu e Thr-Lys-Asp, envolvendo as duas treoninas no sítio ativo. Nossas análises do volume da cavidade catalítica de três amidohidrolases também indicam que o aumento da atividade L-glutaminase em algumas enzimas é diretamente proporcional ao aumento do volume dessa cavidade.

Abstract

One of the major problems faced in antiviral therapy against AIDS is the emergence of viral variants that exhibit drug resistance, as well as viral subtypes naturally more liable to development of therapeutic failure. In this work we solved the crystal structure of four HIV-1 proteases complexed with the universal C2 symmetry-based inhibitor TL-3. These proteases were obtained from patients with AIDS: one of the subtype B wild type (*Bwt*), one of the subtype F wild type (*Fwt*), and a mutant of each subtype (*Bmut* and *Fmut*), these last two out of patients showing therapeutic failure. The proteins were produced by *Escherichia coli* bacteria heterologous expression, purified out of the inclusion bodies, and refolded. The collected diffraction data were processed to 2.1 Å resolution for the *Bwt*:TL-3 complex, 1.75 Å for *Bmut*:TL-3, 2.1 Å for *Fwt*:TL-3 and 2.8 Å for *Fmut*:TL-3. These data were initially processed in the P6₁22 space group and latter reprocessed in P6₁. The four structures were solved by molecular replacement using a published HIV-1 protease structure as a model. The structural analysis shown that the TL-3 inhibitor binds in a similar fashion in the active site of all four structures. On the proteases *Bmut* and *Fmut* the mutation V82A causes a repacking of the S1' pocket which rearranges the inhibitor's side chain at P1' subsite. Our analysis further indicate that some polymorphic substitutions between subtypes B and F could lead to a stabilization of naturally flexible regions on subtype F proteases, creating a intrinsically less active resistant enzyme. On the proteases *Fwt* and *Fmut* the polymorphic substitution M36I leads to the displacement of the loop between residues 35–41, which would cause a flexibility loss of the flaps and of the loop 76–83 in the active site. Our comparisons further indicate that the polymorphic substitution L89M on non-B subtypes could be equivalent to the L90M resistance mutation on subtype B proteases. Lastly, based on these structural data we were able to suggest a few structural modifications on the TL-3 inhibitor that could furnish a more potent inhibitor. On this thesis we also present the data of structural solution and analysis of the *Escherichia coli* L-asparaginase solved at 1.95 Å resolution in C2 space group. Based on structural and kinetical data we proposed a general reaction mechanism for amidohidrolases, which include L-asparaginases, involving the formation of two catalytic triads Thr-Tyr-Glu and Thr-Lys-Asp, which accounts for the two threonines in the active site. Our cavity volume analysis of three amidohidrolases also indicate that the increasing in the L-glutaminase activity in some enzymes is directly proportional to the increase in the cavity volume.

Prefácio

Este projeto nasceu no ano de 2000, pouco antes do início do meu doutorado, da colaboração entre o Laboratório de Virologia Molecular (LAVIMOAN-UFRJ) e do Grupo de Cristalografia do Centro de Biologia Molecular e Estrutural (CBME-USP-SC), no intuito de obter dados estruturais que levassem à compreensão das diferenças entre os subtipos e mutantes de proteases de HIV-1 existentes no Brasil. Para isso foram escolhidas quatro amostras de cDNA viral, extraídas de pacientes com AIDS, representativas de cada um dos grupos a serem estudados, ou seja, uma amostra do subtipo B selvagem, uma amostra do subtipo F selvagem, e uma amostra com mutações na sequência da protease de cada um desses dois subtipos.

No entanto, as dificuldades enfrentadas na obtenção dessas proteases foram muito maiores que as inicialmente imaginadas, por vários fatores, incluindo minha própria inexperiência na área, a falta de experiência do grupo na produção de proteínas tóxicas e insolúveis e a atividade autolítica dessas enzimas.

Eventualmente contamos com a colaboração do Dr. Alexander Wlodawer, do National Cancer Institute, cujo grupo possui experiência comprovada na produção e resolução estrutural de proteases de HIV, e em 2002 tive a oportunidade de passar oito meses neste laboratório desenvolvendo meu projeto de tese. Lá conseguimos resolver o problema da expressão e reenovelamento das proteases, chegando a realizar experimentos iniciais de cristalização. De volta ao Brasil, esses experimentos foram reproduzidos com sucesso em nossos laboratórios, permitindo-nos obter proteína em quantidade e pureza necessários à nossos experimentos de cristalização.

Enquanto os experimentos de clonagem e testes de expressão das proteases estavam

em andamento, ainda aqui no Brasil, O Prof. Igor me deu a oportunidade de trabalhar com um projeto estrutural, uma enzima L-asparaginase de *E. coli*. Este projeto contribuiu para o início do meu aprendizado sobre a análise de modelos cristalográficos e comparações estruturais, que mais tarde seriam úteis na análise dos dados estruturais das proteases de HIV.

Dessa maneira, achamos por bem dividir os trabalhos apresentados nesta tese em duas partes, a primeira contendo os dados obtidos no projeto principal, sobre proteases de HIV, e a segunda parte com os dados da L-asparaginase de *E. coli* e comparações estruturais de proteínas dessa mesma família.

Desde que os dados estruturais das proteases ainda não foram depositados, e no intuito de facilitar a análise crítica e a leitura desta tese, estamos incluindo um CD contendo os arquivos de coordenadas das quatro proteases, assim como algumas referências pertinentes aos dois projetos aqui apresentados.

Mario Sanches

São Carlos, novembro de 2004.

Parte I

Proteases de HIV

Capítulo 1

Introdução

A Síndrome da Imunodeficiência Adquirida, AIDS, é um complexo de sintomas e doenças resultantes da infecção pelo Vírus da Imunodeficiência Humana, HIV. Caracteriza-se por infecções oportunistas e disfunções neurológicas, dentre outros sintomas. No início da década de oitenta a AIDS apareceu como uma síndrome rara do sarcoma de Kaposi e da pneumonia *Pneumocystis carinii* que afetava um pequeno grupo de homens homossexuais em Nova Iorque e na Califórnia. Embora no início essa tenha sido uma síndrome que afetava homossexuais, não demorou muito até que abrangesse homens e mulheres heterossexuais e crianças nascidas de mães infectadas [Kummerle & Earlywine(1996)].

Hoje, mais de duas décadas depois do primeiro caso de AIDS, cerca de 20 milhões de pessoas já morreram, e cerca de 38 milhões de pessoas estão infectadas em todo o mundo, 1,6 milhões somente na América Latina. Estima-se que somente no ano de 2003, cinco milhões de pessoas tenham sido infectadas entre crianças e adultos [una(2003), una(2004a)]. No Brasil estima-se que por volta de 600.000 pessoas vivam hoje com HIV/AIDS. Desde 1997 observa-se uma estabilização desse número, graças à um esforço governamental sem precedentes [una(2004b)]. Atualmente mais de 135.000 pessoas têm acesso ao tratamento antiretroviral totalmente financiado pelo governo, que produz medicamentos genéricos em instituições nacionais públicas e privadas.

1.0.1 Organização e Replicação do Vírus

O Vírus da Imunodeficiência Humana, HIV, pertence à família *retroviridae*, mais especificamente, à sub-família *lentivirinae*, que inclui o vírus visna-maedi, o Vírus da Encefalite Artrite Caprina (CAEV), o Vírus da Anemia Infecciosa Equina (EIAV), entre outros [Vaishnav & Wong-Staal(1991)].

O genoma do HIV-1 codifica para nove fases de leitura aberta, como mostra a Figura 1.1. Dentre essas estão as poliproteínas *gag*, *pol* e *env* que são subsequentemente clivadas proteoliticamente em um conjunto de proteínas comuns a todos os retrovírus. A poliproteína *gag*, que contém as proteínas MA (matriz), CA (capsídeo), NC (nucleo-capsídeo) e p6, em conjunto com a poliproteína *env*, contendo SU (superfície ou gp120) e TM (transmembrana ou gp41), são componentes estruturais do núcleo do virion e a membrana do envelope externo. PR (protease), RT (transcriptase reversa) e IN (integrase), contidas na poliproteína *pol*, conferem funções enzimáticas essenciais e são encapsuladas na partícula viral. O HIV-1 codifica ainda para seis proteínas adicionais, normalmente chamadas de proteínas de acessório, três das quais (Vif, Vpr e Nef) são também encapsuladas na partícula viral. Duas outras proteínas acessório, Tat e Rev, possuem função regulatória do gene, e a última, Vpu, contribui indiretamente para a montagem do virion. O genoma retroviral é codificado por um RNA de $\approx 9\text{kb}$, sendo que duas cópias de RNA genômico é encapsulado na partícula viral. Assim, HIV-1 pode ser considerado como uma entidade molecular constituída por 15 proteínas e um RNA [Frankel & Young(1998)].

O ciclo de replicação do vírus pode ser dividido em fase inicial, que vai do reconhecimento da membrana celular pelo vírus até a integração do DNA viral ao genoma do hospedeiro, e fase final, que vai da transcrição do DNA integrado até o brotamento e maturação da partícula viral. O processo inicia-se com o reconhecimento de proteínas CD4 na membrana celular. O vírus liga-se à essas proteínas através de interações específicas com sua proteína de envelope SU. Ocorre então o processo de fusão com a membrana e desencapamento viral que leva à formação de um complexo de transcrição

Figura 1.1: Organização do genoma do vírus HIV-1. As poliproteínas *gag* e *env* contêm as proteínas estruturais do vírus HIV-1, enquanto que um deslocamento da fase de leitura gera a poliproteína *gag-pol*, contendo as enzimas PR, RT e IN, que somadas à duas cópias de RNAs genômicos formam a partícula viral. A clivagem das poliproteínas pela protease viral para gerar a partícula infecciosa. Reproduzido de Frankel & Young (1998).

reversa. Essa transcrição é catalisada pela transcriptase reversa (RT) e é auxiliada pela proteína de acessório Vif, transformando o RNA viral em DNA. Esse DNA é então transportado para o núcleo como um complexo de pré-integração, que aparentemente envolve as proteínas IN, MA, RT e Vpr. Uma vez no núcleo esse DNA é covalentemente integrado ao genoma da célula através de atividade catalítica da proteína IN.

A fase final inicia-se com a síntese de mRNAs que serão transportados para fora do núcleo para o processo de tradução. Em primeiro lugar são sintetizados pequenos pedaços de RNA que codificam as proteínas regulatórias Tat, Rev e Nef. Tat é um ativador transcripcional e Rev transporta para o citoplasma aqueles mRNA virais que contém introns funcionais que serão necessários na síntese e empacotamento das poliproteínas *gag* e *gag-pol*. A poliproteína *env* é sintetizada no retículo endoplasmático e aparentemente oligomeriza-se como um trímero. Esse complexo é clivado para produzir um complexo glicoprotéico trimérico (TM-SU)₃ numa associação não covalente. Esse complexo é então transportado para a membrana para formação das partículas virais.

A *env* pode se associar a proteínas CD4 sintetizadas no retículo endoplasmático. Para que isso não aconteça ocorre um ataque à proteína CD4, intermediado pela proteína Vpu. CD4 presentes na membrana são degradados por intermédio da Nef. As poliproteínas *gag* são sintetizadas nos ribossomos. Um deslocamento de fase de leitura gera menores quantidades das proteínas precursoras *gag-pol*, que se associam com a poliproteína *gag* na membrana celular. Aproximadamente 1200 à 2000 cópias de *gag* brotam para formar uma partícula viral imatura, que contém duas cópias do genoma viral. Seguindo-se ao brotamento ocorre a clivagem das poliproteínas pela enzima PR para produzir as outras enzimas, assim como MA, CA e NC. A liberação dessas proteínas gera um rearranjo, num processo chamado maturação, que forma a partícula infecciosa [Turner & Summers(1999)].

A transcriptase reversa, a integrase e a protease estão entre os primeiros alvos terapêuticos identificados, desde que sua inativação levaria à uma interrupção no processo de replicação viral. Os primeiros fármacos anti-HIV identificados foram inibidores de transcriptase reversa (RT), descobertos e desenvolvidos muito antes da estrutura da RT ser resolvida. No entanto a descoberta e desenvolvimento de inibidores da protease foram estruturalmente guiados, e representam um sucesso memorável no desenho racional de drogas [Wlodawer & Vondrasek(1998)].

A terapia baseada na combinação de três ou mais fármacos anti-HIV (*highly active antiretroviral therapy*, HART), incluindo inibidores de transcriptase reversa e protease, diminui de maneira significativa a taxa de progressão da doença e da mortalidade de pacientes com AIDS, permitindo uma sobrevida inimagináveis nos primórdios do tratamento da doença [Palella *et al.*(1998)]. O tratamento, no entanto, deve começar quando a presença do HIV é detectada.

1.1 Protease de HIV

A protease de HIV pertence à família das proteases aspárticas, assim chamadas devido à um par de resíduos aspárticos altamente conservados no sítio ativo dessas proteínas.

Inicialmente essa classificação baseava-se em sua sequência primária e inibição por pepsatina, sendo posteriormente confirmada pelas primeiras estruturas cristalográficas [Roberts *et al.*(1990), Wlodawer *et al.*(1989), Navia *et al.*(1989)]. Como já citado, essa protease é essencial na replicação e maturação do vírus.

As primeiras estruturas da protease nativa foram quase simultaneamente resolvidas por três grupos, duas utilizando proteína produzida por expressão heteróloga por *E. coli* (Merck Research Laboratories [Navia *et al.*(1989)] e Birbeck College [Lapatto *et al.*(1989)]) e outra com proteína produzida por síntese total (NCI-Frederick [Wlodawer *et al.*(1989)]). No entanto, somente as coordenadas dos C α da estrutura resolvida pelos pesquisadores da Merck foi publicada, e ainda assim, foi mostrado que essa estrutura possui erros grosseiros no traçado da cadeia [Wlodawer *et al.*(1989), Lapatto *et al.*(1989), Kent *et al.*(2000)].

Figura 1.2: Elementos de estrutura secundária da protease de HIV-1. O desenho mostra uma representação esquemática da protease *Bmut* complexada com o inibidor TL-3, resolvida nesta tese. Fitas- β em vermelho e hélices em verde, seguindo-se a nomenclatura de Wlodawer *et al.*(1989) para os elementos de estrutura secundária.

A protease de HIV-1 é uma proteína homodimérica, como 99 aminoácidos em cada

subunidade, constituída basicamente por folhas- β , como representado na Figura 1.2. As pequenas fitas- β N-terminais a formam a parte externa da folha- β interfacial responsável pela manutenção da estrutura dimérica da protease, que é completada internamente pelas fitas- β q . O primeiro conjunto de fitas- β $a-d$ em um mesmo monômero são relacionadas por uma quase simetria de ordem dois intramolecular com o segundo conjunto de fitas- β $a'-d'$. Uma única hélice está presente em cada subunidade, com um padrão de pontes de hidrogênio interno intermediário entre uma hélice- α e uma hélice- 3_{10} . A região entre os resíduos 44-57, chamada de *flap*, contém as fitas- β a' e b' e fecha-se sobre o sítio ativo da protease complexada com inibidores formando um teto. Esses *flaps* encontram-se abertos na estrutura *apo* e esse fechamento sobre o sítio ativo constitui a maior diferença entre as estruturas complexada e não-complexada da protease [Wlodawer *et al.*(1989), Miller *et al.*(1989b), Fitzgerald *et al.*(1990)]. No entanto, as estruturas das enzimas são bastante conservada entre os diferentes complexos publicados, com rmsd entre C α raramente excedendo 0.6Å [Wlodawer & Vondrasek(1998)].

O sítio ativo é composto por uma tríade catalítica (Asp25-Thr26-Gly27) localizada numa alça interna entre as fitas- β $c-d$, cuja estrutura é mantida por um conjunto de pontes de hidrogênio característico das proteases aspárticas. Os dois grupos carboxílicos dos resíduos Asp25 simetricamente relacionados são aproximadamente coplanares e formam um contato íntimo envolvendo seus átomos O δ 1. Ainda, as duas alças do sítio ativo são conectadas por um conjunto de pontes de hidrogênio chamada *fireman's grip* envolvendo os outros dois resíduos catalíticos, Thr26 e Gly27.

O monômero da protease de HIV é inativo, desde que a cavidade catalítica se forma na interface de dimerização, ao contrário do observado em outras proteases aspárticas retrovirais cuja estrutura é monomérica, como no caso do Rous sarcoma vírus (RSV) [Miller *et al.*(1989a)] . Têm sido proposto que essa dimerização constitui um mecanismo regulatório da sua ativação. Dadas as consequências potencialmente catastróficas de uma ativação prematura, a ativação da protease é retardada até que a partícula viral imatura esteja completamente formada e desligada da célula hospedeira.

Assim a protease é expressa como um produto de fusão *gag-pol* e é liberada dessa poliproteína precursora por um mecanismo autocatalítico, através da dimerização dos domínios da protease fusionada, quando a concentração dessas proteínas é aumentada dado o encapsulamento pela partícula viral [Debouck *et al.*(1987)].

Acredita-se também que as cisteínas 67 e 95, altamente conservadas entre as proteases de HIV, possuam um papel regulatório no processo de maturação do vírus. Estudos mostraram [Davis *et al.*(1996), Davis *et al.*(1999)] que a modificação dessas cisteínas por vários agentes pode levar à enzimas menos ativas. Aparentemente essas cisteínas podem ser oxidadas (glutacionadas, por exemplo) no precursor *gag-pol*, ainda no citoplasma celular, prevenindo a ativação prematura da protease no citoplasma. Enzimas celulares, tais como tiotransferases, poderiam então reverter essas modificações quando o precursor se acumula na membrana celular, restaurando a atividade da enzima pouco antes do brotamento do virion imaturo.

1.1.1 Mecanismo de Ação

O substrato protéico da protease de HIV-1 liga-se na cavidade catalítica da enzima de forma estendida, formando uma fita- β [Wlo(1991), Wlodawer & Erickson(1993)]. Essa estrutura é mantida por um padrão de pontes de hidrogênio entre átomos da cadeia principal do substrato e da protease, como mostra a Figura 1.3.

As cadeias laterais do substrato na porção N-terminal à partir da ligação hidrolisada são denominadas P1-Pn. Aquelas na porção C-terminal são denominadas P1'-Pn'. Os resíduos que interagem com esses subsítios formam os bolsões de ligação denominados S4-S4' correspondentes.

A protease de HIV processa as poliproteínas *gag* e *gag-pol* em pelo menos oito sítios discretos. Inicialmente acreditava-se que esse número limitado de sítios de clivagem no substrato natural da HIV PR indicava uma alta especificidade da enzima frente à sequência primária do seu substrato. Essa proposta, no entanto, mostrou-se infundada. De fato foi mostrado que a sequência além dos resíduos P1-P1' do substrato são

Figura 1.3: Esquema geral de pontes de hidrogênio entre HIV-1 PR e um substrato representativo. As flechas indicam a ligação a ser clivada. Os resíduos na porção amino-terminal da ligação a ser clivada são chamados P1-Pn, e aqueles na porção carboxi-terminal de P1'-Pn', de acordo com Schechter & Berger(1967). Os símbolos S4-S4' denotam os bolsões correspondentes a cada subsítio da enzima. Adaptado de Wlodawer & Erickson (1993).

bastante divergentes. Em P1-P1' existe a preferência por resíduos hidrofóbicos, normalmente contendo um resíduo de Phe ou Tyr [Meek(1992), Abdel-Meguid(1993), Wlodawer & Erickson(1993)].

Figura 1.4: Mecanismo de reação das proteases de HIV envolvendo a clivagem da ligação peptídica Tyr-Pro. Uma molécula de água catalítica é adicionada ao carbono carbonílico da ligação peptídica do substrato gerando o intermediário tetraédrico (2). A desprotonação de uma das hidroxilas do intermediário concomitantemente com a protonação do grupo de partida prolil libera os produtos e regenera a enzima. Adaptado de Meek(1992).

Estudos cinéticos e estruturais indicaram que ambos os resíduos Asp25A e Asp25B no sítio ativo participam da catálise, em estados opostos de protonação. O mecanismo de ação aceito para as proteases de HIV está mostrado na Figura 1.4. Uma molécula de água interage por pontes de hidrogênio com os resíduos catalíticos Asp25A e Asp25B. A ligação do substrato (1) promove a protonação do seu carbonílico da ligação peptídica pelo Asp25A e desprotonação e adição da água catalítica guiada pelo resíduo Asp25B, formando o intermediário tetraédrico (2). A desprotonação de um dos grupos hidroxil em 2 pelo Asp25A e a protonação do grupo de saída prolil pelo Asp25B ocorrem simultaneamente, formando os produtos da hidrólise (ácido carboxílico e amina), e restaurando os resíduos aspárticos (3) [Meek(1992)].

1.1.2 Inibidores

O desenho de inibidores de HIV PR baseia-se principalmente em análogos do substrato ou do estado de transição. Essa estratégia havia sido usada com sucesso no desenvolvimento de inibidores para renin, outra protease aspártica, e o conhecimento adquirido com os complexos renin-inibidores foi extensivamente usada no desenho dos primeiros

inibidores para HIV PR [Wlodawer & Erickson(1993)]. A grande maioria dos inibidores desenvolvidos baseou-se na síntese de análogos do substrato onde a ligação peptídica hidrolisável P1-P1' foi substituída por um isótero não hidrolisável com geometria tetraédrica (Figura 1.5).

Figura 1.5: Representação da estrutura química de alguns isósteros não hidrolisáveis clássicos do estado de transição, empregados na síntese de inibidores de protease de HIV-1. O composto (a) mostra a ligação peptídica normal para comparação.

Esses estudos mostraram que, embora o comprimento mínimo do substrato seja de sete aminoácidos, um mimético tri ou tetrapeptídico pode ser um inibidor potente das proteases de HIV-1. Em todos os casos, exceto para a amida reduzida, o grupo hidroxila da junção não hidrolisável é posicionada entre os grupos aspartato da protease, com distância para uma interação de ponte de hidrogênio para pelo menos um oxigênio do carboxilato dos aspartatos. As pontes de hidrogênio se formam principalmente entre

os átomos da cadeia principal da enzima e do substrato/inibidor, seguindo um padrão bastante similar entre os complexos (Figura 1.3) [Wlodawer & Erickson(1993)].

Também foi mostrado que a ordem de preferência da HIV PR por diferentes isósteros foi: hidroxietileno $[\text{CH}(\text{OH})-\text{CH}_2] > \text{difluorostatona } [\text{CO}-\text{CF}_2-\text{CO}] > \text{statina } [\text{CH}(\text{OH})-\text{CH}_2-\text{CO}] > \text{fosfinato } [\text{P}(\text{O}_2)\text{H}-\text{CH}_2] > \text{amida reduzida } [\text{CH}_2-\text{NH}]$. Normalmente, grupos hidrofóbicos e aminoácidos não naturais são incorporados à esses isósteros para aumentar a potência do inibidor. Por exemplo, o isóstero hidroxietilamina facilmente acomoda um grupo iminoácido, mimetizando assim a ligação Phe-Pro ou Tyr-Pro encontrada em vários sítios de clivagem das poliproteínas de HIV.

Os inibidores, no entanto, não precisam necessariamente ser peptídicos. Um exemplo bastante elegante do desenvolvimento de um inibidor potente, simétrico e não-peptídico, cujo desenvolvimento foi baseado em estrutura, é o motivo uréia cíclica, mostrado na Figura 1.6-a. Note, na Figura 1.3, que existe uma água interagindo com o substrato e os resíduos do sítio ativo. Essa água está presente em todas as estruturas cristalográficas publicadas de complexos protease-inibidor de análogos não hidrolisáveis. Assumindo que o deslocamento dessa água por um grupo funcional do inibidor contribuiria entropicamente para a energia de ligação, Lam *et al.*(1994) empreenderam uma busca em bases de dados tri-dimensionais por moléculas que satisfizessem os seguintes critérios: (1) um motivo central que interagisse com Asp25A/Asp25B, (2) grupos hidrofóbicos localizados nos subsítios P1, (3) um grupo contendo um doador de ponte de hidrogênio que substituísse a água no sítio ativo. O resultado desse exercício foi a identificação do composto mostrado na Figura 1.6-b. A estrutura cristalográfica do complexo da protease com esse inibidor mostrou que ele desloca eficientemente a água no sítio ativo.

O estudo de inibidores de vários tamanhos contendo os isósteros mostrados na Figura 1.5 eventualmente levou ao desenvolvimento de inibidores potentes e seletivos contra HIV-1. Esses inibidores são capazes de interagir com a protease durante o processo de formação da partícula viral *in vivo*, capturando o virion numa morfologia

Figura 1.6: Inibidor não peptídico da protease de HIV-1. O motivo uréia cíclica central possui um doador de ponte de hidrogênio (o oxigênio carbonílico da uréia) que substitui a água encontrada no sítio ativo das proteases de HIV-1. (a) representação geral do motivo uréia cíclica mostrando as posições e estereoquímica dos grupos em P₁/P₁' e P₂/P₂'; (b) exemplo de um inibidor contendo o motivo mostrado em (a).

imatura, incapaz de infectar outras células. Atualmente existem sete inibidores da protease aspártica do HIV-1 comercialmente disponíveis (Figura 1.7), são eles: amprenavir (VX-478), indinavir (MK-639), lopinavir (ABT-378), nelfinavir (AG1343), ritonavir (ABT-538), saquinavir (Ro31-8959) e atazanavir (BMS232632).

1.2 Subtipos e Mutações

Dada sua alta taxa de mutação, cepas de HIV circulando globalmente exibem um extraordinário grau de diversidade, que pode influenciar aspectos da sua biologia, tais como infectividade, transmissibilidade e imunogenicidade. Acredita-se que o HIV tenha se originado à partir da transmissão por primatas não-humanos infectados com SIV. As infecções em humanos são causadas por HIV tipo 1 e HIV tipo 2, sendo o HIV-1 o mais virulento, e o HIV-2 confinado aos países da África Ocidental [Reeves & Doms(????)].

Seqüências derivadas de cepas de HIV-1 têm sido historicamente classificadas em grupos e subgrupos com base em sua relação filogenética. À partir das seqüências dos genes *env* e *gag* o HIV-1 foi dividido em três grupos: M (*Major*), O (*Outlier*) e N (*Non-M Non-O*). Os grupos N e O contém uma minoria presente na República dos Camarões. Estudos demonstraram que o grupo N é decorrente da recombinação de SIV de chimpanzés (SIVcpz) com HIV-1 [Mauclere(2000)], e que os chimpanzés são

Figura 1.7: Representação dos sete inibidores comerciais de HIV-1 PR atualmente disponíveis.

hospedeiros e reservatórios naturais de HIV-1 [Gao *et al.*(1999)].

Os subtipos de HIV-1 são um conjunto de seqüências dentro do grupo M, que dominam a epidemia global, classificadas de acordo com a análise filogenética. Essas análises filogenéticas sugerem que os subtipos do grupo M tiveram um único ancestral, enquanto que os grupos N e O parecem ter evoluído de ancestrais diferentes [Robertson *et al.*(1999)]. Existem ainda as recombinações entre os subtipos, ou seja, a posição filogenética de alguns vírus difere dependendo da região genômica analisada, indicando que esses vírus são mosaicos gerados por recombinação.

Nos países da América do Norte e Europa Ocidental o subtipo B é dominante, representando mais de 95% na distribuição de subtipos. No restante do mundo, no entanto a diversidade é muito grande, não sendo possível definir qual subtipo é predominante, a não ser regionalmente, como mostra a Figura 1.8-a.

Na América Latina o subtipo B ainda prevalece, mas os outros subtipos não podem ser desconsiderados, como mostra a Figura 1.8-b. No Brasil, além do subtipo B são encontrados os subtipos A, C, D e F, sendo o F o subtipo não-B mais abundante. De acordo com a distribuição das seqüências no “Los Alamos HIV Sequence Database”² temos, no Brasil, 71.0% B, 5.0% C, 1.6% D, 12.1% F e 9.4% de outras seqüências complexas (Figura 1.8-b), embora Caride *et al.* (2001) tenham reportado pelo menos uma seqüência do subtipo A no Brasil. Todos os inibidores de protease de HIV disponíveis até o momento foram desenvolvidos para o subtipo B.

A resistência à inibidores é uma consequência inevitável da supressão incompleta da replicação do vírus. A alta taxa de replicação do vírus (10^{10} partículas virais/dia), aliada às altas taxas de substituições incorretas em posições de nucleotídeos durante a transcrição reversa (10^{-4} /nt) devido à falta de uma atividade de correção (*proof-reading*) da enzima, e a variabilidade genética do vírus, têm levado à identificação de variantes de HIV que exibem susceptibilidade variada à inibidores. Da lista de mutações resistentes publicada pelo “Los Alamos HIV Sequence Database” [Parikh *et al.*(2002)],

²<http://hiv-web.lanl.gov>

(a) Distribuição Global de Subtipos de HIV-1

(b) Distribuição de Subtipos de HIV-1 na América Latina

Figura 1.8: Distribuição geográfica dos subtipos de HIV-1, reproduzido de acordo com o “Los Alamos HIV Sequence Database”¹. Estão representados os subtipos A, B, C, D, F, G, H, J, O e quatro mosaicos CRF01_AE, CRF02_AG, CRF03_AB, CRF04_cpx. (a) Distribuição global de subtipos. Repare que o subtipo B é dominante na América do Norte e Europa Ocidental; (b) Distribuição dos subtipos na América Latina. Embora o subtipo B seja dominante os outros subtipos não são desprezíveis. O subtipo F é o não-B mais abundante no Brasil.

179 encontram-se na região da protease, o que demonstra a flexibilidade genética do HIV.

Algumas mutação selecionadas durante a terapia com antiretrovirais conferem resistência fenotípica mensurável *per se* e são chamadas mutações “primárias”, enquanto outras compensam a diminuição replicativa do vírus e podem ser associadas à resistência a fármacos ou causam resistência somente quando em combinação com outras mutações e são chamadas “secundárias”. No entanto, quais mutações são primárias e quais são secundárias não é estritamente definido e algumas mutações podem ser consideradas primárias para um inibidor e secundária para outros. A Figura 1.9 mostra um resumo do padrão de resistências para cada um dos inibidores comercialmente disponíveis [Shafer(2004)].

O papel de algumas mutações secundárias foi verificado por estudos de mutações pontuais. Por exemplo, a eficiência catalítica dos mutantes V82A/F/I (normalmente considerada primária) frente ao seu substrato é reduzida de um fator de 3.3x com relação à enzima selvagem, enquanto que a mutação M46I (normalmente secundária) reduz sua eficiência de 2.0x. Já o duplo mutante M46I/V82F tem sua eficiência reduzida de apenas 1.6x [Gulnik *et al.*(1995)]. Isso demonstra que algumas mutações têm um papel dinâmico e compensatório na restauração da atividade catalítica da enzima mutada. Ainda, algumas mutações na protease podem ser acompanhadas de mutações nos sítios de clivagem, no substrato natural. Essas mutações adaptativas no substrato teriam a função de restabelecer a atividade das proteases mutantes resistentes [Lamarre *et al.*(1994) apud Gulnik *et al.*(1995)³].

O quadro se complica ainda mais quando se considera que o padrão de mutações gerados por pressão por inibidores pode ser diferente entre os vários subtipos. Por exemplo, a posição 36 dos subtipos B contém uma metionina, sendo que a mutação M36I aparece como consequência de tratamento antiretroviral. Já nos subtipos não-B

³Lamarre, D., Croteau, G., Pilote, L., Rousseau, P., & Doyon, L. (1994). *Abstracts of the Third International Workshop on HIV Drug Resistance*, página 10, MediTech Media Ltd., London apud Gulnik *et al.*(1995).

Figura 1.9: Inibidores de protease e mutações de resistência. APV: amprenavir, ATV: atazanavir, IDV: indinavir, LPV: lopinavir (em combinação com ritonavir), NFV: nelfinavir, RTV: ritonavir, SQV: saquinavir. A contribuição de uma mutação para a resistência é dada pela intensidade dos retângulos azuis. O primeiro bloco de mutação consiste das mutações primárias; o segundo bloco contém um conjunto de mutações imporantes nos *flaps*; o terceiro bloco contém mutações não-polimórficas menos importantes que as dos *flaps*; o quarto bloco consiste de posições polimórficas que contribuem para a resistência quando em combinação com outras mutações de resistência. Reproduzido de Shafer (2004).

mesma posição contém uma isoleucina na quase totalidade das seqüências isoladas. Ainda, os subtipos F, comuns no Brasil, não desenvolvem as mutações I84V e L90M, comumente encontradas em subtipos B, enquanto que o polimorfismo L89M, provavelmente responsável pela exclusão da L90M, é encontrado na grande maioria dos subtipos não-B [Caride *et al.*(2001)]. Esse padrão de resistências secundárias que são consideradas polimórficas em proteases não-B poderia levar à uma falha terapêutica prematura quando comparado com o subtipo B [Perno *et al.*(2001)].

1.3 Objetivos

Até hoje, os estudos estruturais e de desenvolvimento de inibidores da HIV-1 PR foram, e estão sendo feitos, levando-se em conta somente o subtipo B. Até onde pudemos verificar não existe, até o momento, uma única estrutura de uma protease de subtipo não-B. Nossa proposta é de determinar a estrutura cristalográfica de uma protease subtipo F, desde que esse é o subtipo não-B mais abundante no Brasil, além da estrutura de mutantes de subtipos B e F decorrentes do tratamento com inibidores. Para acessar as diferenças do modo de ligação de um inibidor em cada uma dessas estruturas é necessário que todas elas sejam cristalizadas com o mesmo inibidor. Além disso é necessário que a estrutura da protease subtipo B selvagem também esteja cristalizada com o mesmo inibidor, para compara-lo com o modo de ligação dos isolados brasileiros e mostrar o que os faz únicos.

Para tal, se fez necessário o desenvolvimento da experiência de produção heteróloga e purificação de proteases de HIV, fato inédito no Brasil, além da metodologia pouco comum de reenovelamento de proteínas insolúveis. Nessa tese apresentamos a produção, a purificação e o reenovelamento das proteases subtipo B mutante (*Bmut*), subtipo F selvagem (*Fwt*) e subtipo F mutante (*Fmut*). As três proteases são provenientes de pacientes, e os mutantes foram escolhidos dentre aqueles apresentando falha terapêutica múltipla. A protease *Bmut* apresenta as mutações S37A, R41K, K45R, I54V, L63P, A71V, V82A, L90M frente à seqüência da *Bwt* a *Fmut* as mutações L10I, G16E, K20R,

A37N, P39S, M46I, I54V, N61D, V62I, S63L, L64I, K70R, I72V, T74A, V82A frente à *Fwt*. As substituições polimórficas entre *Bwt* e *Fwt* encontram-se nas posições I15V, E35D, M36I, S37A, R41K, R57K, D60E, Q61N, I62V, L63S, I64L, L89M.

Essas três proteases foram co-cristalizadas com o inibidor universal TL-3 [Li *et al.*(2000)]. Todas as três estruturas foram resolvidas e refinadas e a comparação dessas estruturas está apresentada nesta tese. A estrutura do complexo *Bwt*:TL-3, anteriormente publicada [Li *et al.*(2000)], também foi refeita para resolver um problema de grupo espacial e refinamento do ligante TL-3.

Esse projeto conta com a colaboração do grupo chefiado pelo Prof. Dr. Amilcar Tanuri, Laboratório de Virologia Molecular da UFRJ e do grupo de Dr. Alexander Wlodawer, Macromolecular Crystallography Laboratory, NCI-Frederick, EUA.

Capítulo 2

Metodologia

2.1 Subclonagem

As amostras de cDNA viral das proteases *Bmut*, *Fwt* e *Fmut* foram recebidas do Prof. Dr. Amilcar Tanuri (UFRJ) na forma de material precipitado e foram ressolubilizadas em 10 mM Tris-HCl, pH 8.0. O clone da protease *Bwt* em pET3a nos foi cedido pelo Dr. Jordan Tang (Oklahoma Medical Research Foundation), contendo a sequência da protease e mais dez aminoácidos no N-terminal, correspondentes à uma porção da sequência da poliproteína *gag-pol* [Ido *et al.*(1991)]. Esses clones foram recebidos na forma de cultura de células *E. coli* BL21(DE3)pLysS.

As amostras cedidas pelo Prof. Dr. Amilcar Tanuri foram obtidas de crianças HIV-1 soropositivas verticalmente infectadas [Brindeiro *et al.*(2002)]. As amostras dos subtipos *Bmut* e *Fmut* foram obtidas de pacientes recebendo ritonavir e nelfinavir como inibidores de protease e amivudine (3TC), stavudine (D4T) e zidovudine (AZT) como inibidores de transcriptase reversa. O paciente com *Fmut* recebeu ainda hevirapine (NVP) como inibidor de transcriptase reversa.

2.1.1 Reações de PCR

A amplificação da região da protease (PR) no cDNA do vírus HIV foi realizada por PCR (*Polimerase Chain Reaction*). Foram construídos oligonucleotídeos específicos para as extremidades 5' (*sense*) e 3' (*antisense*) para cada um dos genes, com sítios de restrição para a enzima *NdeI* na extremidade *sense* e *BamHI* na extremidade *antisense*. A

Tabela 2.1: Seqüências dos oligonucleotídeos utilizados nas subclonagens dos genes das proteases em vetor pET11a. Todos as seqüências diretas possuem um sítio de restrição para *Nde*I, e os reversos para *Bam*HI. As letras minúsculas nas seqüências correspondem à porção do gene da protease.

Gene	Oligo	Seqüência
<i>Bmut</i>	bmut_f	5'-AACTGTATCCCTTCATATGcctcagatcactcttttgga-3'
	bmut_r	5'-CAGTTTCAATAGGGGATCCTTAaaaatttaaagtcagccaa-3'
<i>Fwt</i>	fwf_f	5'-TGTACCCTCCCTTCATATGcctcaaatactcttttgga-3'
	fwf_r	5'-CAGTTTCAATAGGGGATCCTTAaaaatttaaagtacaaccaa-3'
<i>Fmut</i>	fmuf_f	5'-AACTGTATCCCTTCATATGcctcaaatactcttttgga-3'
	fmuf_r	5'-CAGTTTCAATAGGGGATCCTTAaaaatttagagtacaaccaa-3'

Tabela 2.2: Seqüências dos oligonucleotídeos utilizados nas subclonagens dos genes das proteases em vetor pET11a, com inserção da mutação Q7K. Todos as seqüências diretas possuem um sítio de restrição para *Nde*I e os reversos para *Bam*HI. As letras minúsculas nas seqüências correspondem à porção do gene da protease.

Gene	Oligo	Seqüência
<i>Bwt</i>	bwt_f7k	5'-GTATCTTTCCATATGccgcagatcactctgtggaagcgctcggtggaacc-3'
	bwt_r7k	5'-GCAGCCGGATCCGCTTAgaagttcaaagtgaaccgatctgag-3'
<i>Bmut</i>	bmut_f7k	5'-GTATCCCTTCATATGcctcagatcactctttggaaacgacccctcgtcac-3'
	bmut_r7k	5'-TAGGGGATCCTTAaaaatttaaagtcagccaatctgag-3'
<i>Fwt</i>	fwf_f7k	5'-CCCTCCCTTCATATGcctcaaatactctttggaaacgacccctcgtcac-3'
	fwf_r7k	5'-TAGGGGATCCTTAaaaatttaaagtacaaccaatctgag-3'
<i>Fmut</i>	fmuf_f7k	5'-GTATCCCTTCATATGcctcaaatactctttggaaacgacccatagtcac-3'
	fmuf_r7k	5'-TAGGGGATCCTTAaaaatttagagtacaaccaatctgga-3'

Tabela 2.1 mostra a seqüência desses oligonucleotídeos desenhados para a subclonagem dos genes das proteases brasileiras em pET11a.

A fim de aumentar a estabilidade das proteases à autoproteólise, como será explicado na seção 3.2.2, foram desenhados oligonucleotídeos para a inclusão da mutação Q7K em todas as quatro proteases, além de subclonagem da seqüência da protease *Bwt* sem fusão, como mostrado na Tabela 2.2.

As reações foram realizadas utilizando-se as quantidades descritas na Tabela 2.3 e submetidas ao programa mostrado na Figura 2.1. Realizou-se sempre um controle negativo, ou seja, uma reação sem DNA molde.

Os produtos de PCR foram analisados em gel de agarose 1% e purificados imediatamente após o término da reação utilizando-se um sistema comercial “ConcertTM Rapid PCR Purification System” (Gibco BRL) ou “QIAquick PCR Purification Kit” (Qia-gen). Quando não utilizados prontamente os insertos purificados foram armazenados

Tabela 2.3: Reagentes utilizados nas reações de PCR para subclonagem da região PR em vetor de expressão. A reação mostrada é para 100 μ L.

Reagentes	Volume	Quantidade
DNA molde (100 ng/ μ L)	2 μ L	200 ng
oligonucleotídeo <i>sense</i> (100 μ M)	1 μ L	1 μ M
oligonucleotídeo <i>antisense</i> (100 μ M)	1 μ L	1 μ M
Tampão Pfu 10x	10 μ L	—
MgSO ₄ (100 mM)	2 μ L	1 mM
dNTPs (A,C,G,T) (25 mM)	0.8 μ L	250 μ M
<i>Pfu</i> DNA polimerase (2.5 U/ μ L)	1.2 μ L	3U
H ₂ O milliQ	82 μ L	—

Figura 2.1: Programação padrão utilizada para a amplificação da sequência da protease (PR) por PCR à partir do cDNA do vírus HIV.

a -20°C .

Em alguns casos os insertos amplificados não purificados foram subclonados no vetor pCR4Blunt-TOPO antes de serem submetidos às reações de digestão. Para tal, utilizou-se sistema comercial “zero-blunt TOPO PCR cloning kit for sequencing” exatamente como descrito no manual. Brevemente, ao produto de amplificação não purificado (1–4 μL) foi adicionado 1 μL de uma solução salina e 1 μL de vetor TOPO obtendo-se um volume final de 6 μL . A mistura reagiu a temperatura ambiente por 5 min. Utilizou-se 2 μL do produto na transformação de bactérias DH5 α .

As verificações por PCR de colônia foi realizada utilizando-se o mesmo protocolo de PCR descrito trocando-se o DNA por uma colônia inteira retirada diretamente da placa e adicionada à mistura de reação. Neste caso aumentou-se os tempos de pré-PCR e pós-PCR para 10 min.

2.1.2 Reações de Digestão

As reações para digestão do inserto e do plasmídeo pET11a foram realizadas utilizando as enzimas *NdeI* e *BamHI* de acordo com as quantidades mostradas na Tabela 2.4. A Figura 2.2 mostra o plasmídeo pET11a utilizado na subclonagem de todas as proteases. Os sítios de restrição utilizados estão assinalados. No caso em que o produto de PCR foi subclonado em TOPO a reação de digestão foi realizada com esse vetor para liberação do inserto digerido.

Tabela 2.4: Reagentes e quantidade utilizados nas reações de digestão do DNA plasmidial e do produto de PCR purificado com volume final de 50 μL .

Reagentes	Volume	Quantidade
Tampão <i>BamHI</i> 10x	5 μL	—
DNA (0.1 $\mu\text{g}/\mu\text{L}$)	10 μL	1 μg
<i>BamHI</i> (20 U/ μL)	2.5 μL	50 U
<i>NdeI</i> (20 U/ μL)	2.5 μL	50 U
H2O milliQ	30 μL	—

Normalmente a enzima *NdeI* foi acrescentada uma hora após o início da reação com *BamHI*, num total de três horas de reação à 37°C .

Figura 2.2: Vetor de expressão pET11a [Mierendorf *et al.*(1994)] utilizado para a subclonagem de todas as quatro proteases desse trabalho. Os sítios de restrição utilizados, *Bam*HI e *Nde*I, estão marcados na figura.

Após a reação de ligação padronizou-se desfosforilar o vetor a fim de evitar o aparecimento de falsos positivos nas reações de ligação devido à religação do vetor. Normalmente esse procedimento foi realizado acrescentando-se 1 μL de fosfatase alcalina de camarão (SAP - *Shrimp Alkaline Phosphatase*) à cada 40 μL de reação de digestão do vetor. Alternativamente a desfosforilação pode ser realizada com o produto da reação de digestão purificado, conforme mostra a Tabela 2.5, com resultados igualmente satisfatórios.

Tabela 2.5: Reação de desfosforilação do vetor digerido purificado utilizando-se fosfatase alcalina de camarão (SAP). Normalmente todo o DNA plasmidial disponível pode ser digerido numa única reação.

Reagentes	Volume
Tampão SAP 10x	5.5 μL
SAP (10 U/ μL)	4.5 μL
vetor pET11a digerido	45 μL

A reação foi mantida a 37°C por uma hora e a SAP desativada por aquecimento a 80°C por 15 minutos.

Todos os produtos de digestão foram purificados imediatamente após a reação utilizando-se o sistema comercial “Gel Purification Kit/QIAGEN”, de acordo com as instruções do fabricante, e armazenados à -20°C caso necessário.

2.1.3 Reações de Ligação

Dois sistemas diferentes foram utilizados nas reações de ligação. O primeiro trata-se do sistema convencional usando a enzima DNA-ligase (Sigma) comercial. Estas reações foram realizadas a 15°C por 24h, utilizando-se as quantidades mostradas na Tabela 2.6. Normalmente várias reações foram realizadas variando-se as proporções vetor/inserto.

O volume total de cada reação foi utilizado para a transformação de bactérias competentes *E. coli* DH5 α .

O segundo sistema de ligação utilizado foi o “Rapid DNA Ligation” (Roche), seguindo-se as recomendações do fabricante. O DNA digerido e purificado do gel foi eluído em Tampão DNA 1x fornecido. Para realizar a ligação, 8 μL de vetor dige-

Tabela 2.6: Reagentes utilizados nas reações de ligação convencional. As proporções entre vetor e inserto são variadas.

Reagentes	Volume
vetor pET11a digerido	1–5 μ L
produto de PCR digerido	1–10 μ L
tampão ligase 10x	2 μ L
ATP (10 mM)	2 μ L
ligase (1 U/ μ L)	2 μ L
H ₂ O	para completar 20 μ L

rido (64 ng) e 1 μ L de produto de PCR (10 ng) foram misturados, reagindo por 5min a 20°C. Do produto de reação, 6 μ L foram utilizados na transformação de bactérias competentes *E. coli* DH5 α e o restante congelado a –20°C.

Em quaisquer dos sistemas utilizados realizou-se sempre uma reação controle con- tendo um plasmídeo digerido com uma única enzima e não desfosforilado para acessar o sucesso da reação de ligação. As colônias foram inicialmente verificadas através de PCR das colônias isoladas nas placas de Petri (veja Seção 2.1.1). As colônias que apresenta- ram resultados positivos tiveram seus plasmídeos extraídos utilizando-se o sistema “SV Minipreps” (Promega) ou “ConcertTM Rapid Plasmid Miniprep System” (Invitrogen) e digeridos com as enzimas *Bam*HI e *Nde*I para confirmação da presença do inserto contendo a sequência da protease (veja Seção 2.1.2 para as reações de digestão). Os plasmídeos com resultado positivo foram armazenados em água a –20°C.

2.2 Expressão, Purificação e Reenovelamento

Cada um dos plasmídeos construídos foram utilizados na transformação de células *E. coli* BL21(DE3). Vale salientar que os estoques a –80°C das bactérias transformadas não são estáveis e as bactérias tiveram que ser freqüentemente transformadas para que os níveis de expressão fossem mantidos.

Utilizou-se meio LB suplementado com 35 μ g/mL de cloranfenicol e 100 μ g/mL de ampicilina, 1% de glucose, 0.05% de antiespumante 204 (Sigma). Tipicamente, inoculou-se 150 mL de meio com uma única colônia selecionada de uma placa de meio

LB sólido e deixou-se crescer a 37°C e 250 rpm por uma noite. A suspensão foi centrifugada a temperatura ambiente, o meio descartado, e as células ressuspendidas em 30 mL de meio fresco a 37°C. Esse procedimento é importante para garantir a eficiência dos antibióticos na seleção de bactérias transformadas. Essa suspensão foi imediatamente utilizada para inocular 3 L de meio a 37°C e a cultura foi incubada a 37°C até uma $DO_{600}=0.7$ e então induzida com 1 mM de isopropil- β -D-tiogalactopiranoside (IPTG). Após 4 h as células foram centrifugadas a 4000xg por 15 min e a massa celular obtida armazenada a -80°C ou imediatamente tratada conforme descrito a seguir.

2.2.1 Lise Celular e Lavagem dos Corpos de Inclusão

Esse procedimento é uma adaptação do protocolo utilizado pelo Dr. Sergei Gulnik (comunicação pessoal) e foi realizado a 4°C ou em gelo. A massa celular foi ressuspendida em tampão de lise A, TLA (50 mM Tris-HCl, pH 8.0, 5 mM EDTA, 5 mM DTT ou 0.1% β -mercaptoetanol), na proporções de 20 mL/g e acrescido de 100 μ g/mL de lisozima e agitado por 30 min. A suspensão foi sonicada (6x30s) e centrifugada a 18000xg por 30 min, estando a protease na fração insolúvel na forma de corpos de inclusão.

A partir desse ponto passou-se à lavagem dos corpos de inclusão. Primeiramente o *pellet* foi ressuspendido no mesmo volume inicial de tampão de lise B, TLB (TLA + 2 M uréia + 1% Triton X-100), e centrifugado a 18000xg por 30 min. O *pellet* foi então lavado mais duas vezes repetindo-se o procedimento de ressuspensão e centrifugação em TLA e finalmente ressuspendido em 20% do volume inicial de TLA para armazenamento à 4°C.

Cada um dos sobrenadantes, os corpos de inclusão lavados e o extrato de proteínas totais foram analisados por SDS-PAGE.

Nos testes de expressão foram retiradas amostras da ordem de 1 mL de meio, sendo as amostras microprocessadas pelo mesmo método acima descrito.

2.2.2 Colunas de Troca Iônica

À lavagem dos corpos de inclusão seguiu-se dois passos de purificação cromatográfica com colunas de troca iônica: uma coluna de troca aniônica HiTrap Q-sepharose (Amersham-Pharmacia) seguida de uma coluna de troca catiônica HiTrap SP-sepharose (Amersham-Pharmacia). Essa metodologia é uma adaptação do método descrito por Laco *et al.*(1997).

Os corpos de inclusão foram centrifugados à 18000xg por 30 min, dissolvidos em tampão A (50 mM Tris-HCl, pH 8.0, 8 M uréia, 5 mM EDTA, 5 mM DTT ou 0.1% β -mercaptoetanol) na proporção de 1 mL para cada 10 mL de corpos de inclusão lavados e a amostra clarificada por centrifugação à 18000xg por 30 min. O sobrenadante foi injetado na coluna Q-sepharose previamente equilibrada com tampão A num fluxo ≤ 1 mL/min e a fração que não se liga à coluna, contendo a protease de interesse, coletada numa única fração. A coluna foi regenerada usando tampão A + 1 M NaCl.

À amostra proveniente da coluna Q-sepharose acrescentou-se NaOAc suficiente para uma solução 20 mM e ajustou-se o pH para 5.0. A amostra foi então injetada numa coluna HiTrap SP-sepharose (Amersham-Pharmacia) previamente equilibrada com tampão B (50 mM NaOAc, pH 5.0, 8 M uréia, 5 mM EDTA, 5 mM DTT ou 0.1% β -mercaptoetanol) num fluxo ≤ 1 mL/min. A HIV PR de interesse ligada à coluna foi eluída através de um gradiente para atingir 30% de tampão C (tampão B + 1 M NaCl) em 10 volumes de coluna, coletada em frações de 1 mL e analisadas por SDS-PAGE.

A fração solúvel da lavagem dos corpos de inclusão da protease *Fwt* foi carregada diretamente na coluna HiTrap Q-sepharose e os passos seguidos conforme descrito para os corpos de inclusão solubilizados.

2.2.3 Reenovelamento das Proteases

As proteases foram reenoveladas através da retirada do agente desnaturante por diálise [Tsumoto *et al.*(2003)]. A amostra proveniente das frações combinadas após a coluna HiTrap SP-sepharose foi dissolvida em Tampão B até uma concentração máxima de

0.2 mg/mL (Bradford) e dialisada em sacos de diálise de 3.5k MWCO ou 6–8k MWCO contra tampão fosfato de reenovelamento, TFR (20 mM Na_2HPO_4 , pH 7.4, 0.1 M NaCl, 2 mM EDTA, 0.1% β -mercaptoetanol), a 4°C. Duas metodologias foram empregadas nas diálises das amostras. Na diálise em passos as amostras foram tratadas da seguinte maneira:

diálise 1: TFR + 4 M uréia, por 2h.

diálise 2: TFR + 2 M uréia, por 2h.

diálise 3: TFR, por uma noite.

diálise 4: TFR, por 4h.

Alternativamente utilizou-se um protocolo de diluição contínua. Inicialmente dialisou-se por 2h contra TFR + 4 M uréia. Utilizando-se uma bomba peristáltica trocou-se o tampão por TFR continuamente durante a noite. Depois dialisou-se por mais 4h contra TFR.

A amostra foi concentrada tipicamente entre 4–7 mg/mL, de acordo com o método de Bradford [Bradford(1976)], utilizando-se membrana YM10 (10kDa MWCO).

2.3 Co-cristalização

Os inibidores utilizados neste trabalho foram dissolvidos em solvente orgânico apropriado (DMSO e/ou DMF) na concentração de 5 mg/mL antes da realização dos ensaios e armazenados à -20°C . Os complexos protease-inibidor foram obtidos misturando-se proteína pura concentrada com excesso molar de $\approx 2\text{x}$ de inibidor, em gelo, por um período mínimo de uma hora antes da realização dos ensaios. A amostra foi clarificada por centrifugação à 16 k rpm por 30 minutos.

Para os ensaios de cristalização das proteases complexadas com inibidores foram utilizados os métodos de gota suspensa (*hanging drop*) e gota sentada (*sitting drop*) em

caixas de cristalização de 24 poços (Hampton Research), realizados em câmaras com temperatura controlada de 4°C e 18°C.

Os primeiros testes de cristalização de cada um dos complexos protease-inibidor foram realizadas utilizando-se um ensaio contendo 24 condições e três precipitantes mecanisticamente distintos (sal, solvente orgânico e precipitante polimérico). Algumas das condições desse ensaio foram incluídas baseadas em condições descritas na literatura onde outras proteases foram previamente cristalizadas, e outras de maneira a incluir condições contendo combinações de dois ou três precipitantes diferentes, de maneira a aumentar a probabilidade de obtenção de cristais [Majeed *et al.*(2003)]. As quantidades de cada precipitante foram determinadas à partir de ensaios de precipitação do complexo *Bmut*+TL-3. A Tabela 2.7 mostra a constituição de cada uma dessas condições.

Tabela 2.7: Condições do ensaio preliminar de cristalização para as proteases de HIV. As condições 3–13 utilizam 0.1M cacodilato de sódio, pH 6.2 como tampão, e as condições 14–24 utilizam 0.1M NaOAc, pH 5.8. Somente as condições 1 e 2 não utilizam tampão.

#	Precipitante ²
1	0.96M SA
2	1.60M SA
3 e 14	0.48M SA
4 e 15	0.64M SA
5 e 16	0.80M SA
6 e 17	15% PEG 3350
7 e 18	25% PEG 8000
8 e 19	0.32M SA, 6% MPD
9 e 20	0.48M SA, 6% MPD
10 e 21	0.64M SA, 6% MPD
11 e 22	15% PEG 3350, 6% MPD
12 e 23	20% PEG 8000, 6% MPD
13 e 24	0.26M SA, 5% PEG 3350, 6% MPD

¹ NaCac.: cacodilato de sódio, pH 6.2; NaOAc : acetato de sódio, pH 5.5
² SA: sulfato de amônio; MPD: 2-metil-2,4-pentanodiol; PEG: polietilenoglicol

Os resultados obtidos nesse ensaio foram considerados na elaboração de ensaios de refinamento para cada uma dos complexos testados.

2.3.1 *Seeding*

Em alguns casos foram realizados experimentos de *micro* e *macroseeding* para aumentar o tamanho dos cristais.

Macroseeding foi realizado lavando-se o cristal primeiramente numa solução estabilizadora e depois em diluições sucessivas dessa solução com água. A solução estabilizadora consiste da solução da gota contendo o cristal (componentes da solução do poço + componentes da solução proteína) sem proteína. Esse cristal foi então semeado numa gota previamente estabilizada nas condições em que cresceu anteriormente mas com a quantidade de proteína reduzida até o limite em que não aparecem cristais, tipicamente com uma diluição de 5x do valor inicial [Bergfors(2003)].

Microseeding foi realizado por dois métodos: *streakseeding* [Stura & Wilson(1991)] utilizando-se pêlos de rabo de cavalo lavados e moagem do cristal em solução estabilizadora [Luft & DeTitta(1999)], com diluição sucessiva de 10x, aplicando-se 0.2 μL na gota (veja também [Bergfors(2003)]). Normalmente ensaiou-se seis diluições para cada tentativa de *microseeding*.

2.3.2 Desidratação dos Cristais

Desidratação é um dos métodos utilizados para melhorar a qualidade da difração de cristais de proteínas, juntamente com *annealing*. Nesse caso utilizou-se um método de desidratação rápida descrito recentemente [Abergel(2004)]. Para tal alguns cristais foram transferidos para uma gota sentada de 10 μL contendo a solução do poço original acrescida de 20% de glicerol como agente crioprotetor. Essa gota foi deixada em repouso, com o poço parcialmente tampado, e os cristais foram testados a intervalos regulares entre 15 min e quatro horas.

2.4 Coleta e Processamento de Dados

Os dados de difração foram obtidas usando-se raios-X Cu $K\alpha$ de um gerador de ânodo rotatório Rigaku ultraX 18 e adquiridas numa placa de imagem MAR 345 dtb ou na

linha CPr do Laboratório Nacional de Luz Síncrotron (LNLS) [Polikarpov *et al.*(1998a), Polikarpov *et al.*(1998b)], utilizando-se radiação de comprimento de onda 1,427Å e um detector marCCD. Foram coletadas imagens de rotação ($\varphi = 1^\circ$) a temperatura de 100K com o cristal previamente tratado com uma solução crioprotetora e montado em um *loop* de coleta. O número mínimo de imagens e o ângulo inicial necessários para se atingir completitude máxima foi determinado utilizando-se a subrotina **Strategy** dentro do programa MOSFILM 6.2.3 [Leslie(1992)].

A qualidade das imagens de difração foi melhorada utilizando de um à três ciclos de *annealing*. A metodologia utilizada consiste da transferência do cristal congelado para uma gota da solução crioprotetora a temperatura ambiente e então congelado novamente (*flash-cooled* []).

Cada conjunto de imagens coletado foi processado usando o programa MOSFILM 6.2.3 [Leslie(1992)] e escalonado com o programa SCALA 3.1.20 [Evans(1997)], separando-se uma porcentagem das reflexões para o conjunto teste (R_{free}) [Brünger(1993)].

2.5 Resolução Estrutural

O conteúdo de solvente, assim como o número de subunidades, da unidade assimétrica foi estimado calculando o coeficiente de Mathews [Matthews(1968)] utilizando o programa MATTHEWS_COEF dentro do pacote de programas CCP4 [Bailey(1994)] versão 4.2.2 ou 5.0.

Para a resolução estrutural por substituição molecular utilizou-se os programas AMoRe [Navaza(1994)] e Phaser, versão 1.2 [Storoni *et al.*(2004)]. A estrutura da protease *Bmut* foi resolvida utilizando-se a estrutura depositada de um complexo de uma protease de HIV-1 subtipo B selvagem com o inibidor JE-2143 à 1.09 Å de resolução (código PDB 1KZK) [Reiling *et al.*(2002)] como modelo. Para a resolução das outras estruturas (*Bwt*, *Fwt* e *Fmut*) a estrutura parcialmente refinada da *Bmut* foi utilizada. Em todos os casos o modelo para a resolução por substituição molecular foi preparado retirando-se o ligante e as moléculas de solvente, e mantendo somente uma das

conformações nos átomos com dupla conformação.

A solução encontrada pelo programa de substituição molecular foi aplicada ao modelo utilizado e este modelo mutado no programa O [Jones *et al.*(1991)] para conter a sequência correta da protease resolvida.

2.6 Refinamento

Após a resolução estrutural os dados foram refinados utilizando-se uma metodologia baseada nas recomendações de Kleywegt & Jone [Kleywegt & Jones(1997)] onde cada macrociclo consiste da utilização de um refinamento de restrição posicional (*restrained*) e uma inspeção visual com manipulação do modelo, também chamada de reconstrução.

Inicialmente utilizou-se o programa CNS, versão 1.1 [Brünger *et al.*(1998)] utilizando-se sempre refinamento por máxima verossimilhança. Em seguida à resolução estrutura por substituição molecular seguiu-se um refinamento de corpo rígido. O procedimento utilizado nos passos seguintes consiste em uma minimização posicional (restrição), seguido de um *simulated annealing*, mais uma minimização, incluídos no procedimento de *simulated annealing* do programa CNS e, finalmente, um refinamento de fatores de temperatura individuais. No primeiro macrociclo foi utilizada uma temperatura inicial de 5000 K (passos de 25 K) para o *simulated annealing* e temperaturas gradativamente menores nos próximos ciclos, chegando-se a uma temperatura inicial de 1000 K nos últimos macrociclos.

Em seguida continuou-se o refinamento utilizado-se o programa Refmac, versão 5.1.24 [Bailey(1994)], com a rotina TLS [Winn *et al.*(2001)]. Após o modelo ser considerado essencialmente correto, moléculas de água foram automaticamente adicionadas pelo programa ARP/wARP [Bailey(1994)], incluído no Refmac, através das posições dos picos no mapa de diferença $|F_{obs}| - |F_{calc}|$, α_{calc} entre $5-3\sigma$, e validadas através da checagem visual levando-se em consideração o ambiente químico de cada molécula de solvente.

Na reconstrução visual utilizou-se o programa O, versão 9.0.7 [Jones *et al.*(1991)]

e o ajuste manual das posições atômicas foi realizado levando-se em consideração as restrições estereoquímicas para o modelo, assim como a concordância com o mapa de densidade eletrônica $2|F_{obs}| - |F_{calc}|$, α_{calc} contornado à 1σ , e o mapa diferença $|F_{obs}| - |F_{calc}|$, α_{calc} contornado à $\mp 3\sigma$. Mapas do tipo *OMIT*, quando utilizados, foram gerados utilizando-se a metodologia de *simulated annealing* no programa *CNS*.

O sucesso de cada macrociclo foi confirmado através do cálculo dos fatores R do refinamento, R_{factor} e R_{free} , da análise dos desvios médios quadráticos das distâncias de ligação e ângulos de ligação com relação aos valores ideais, e do melhoramento da qualidade dos mapas de densidade eletrônica. Para o refinamento das estruturas no grupo espacial $P6_1$ utilizou-se simetria não cristalográfica (NCS) relacionando as subunidades A e B. Partiu-se sempre de uma energia NCS alta ($E=300$ no *CNS* e NCS nível 1 no *Refmac*) e utilizou-se o R_{free} para a validação do relaxamento do peso dessa simetria. A rotina TLS foi utilizada no penúltimo ciclo de refinamento, sem NCS, definindo-se três grupos para cada subunidade, incluindo N-terminal, *flap* (resíduos 35–60) e C-terminal. O último ciclo incluiu NCS novamente.

Para a construção da biblioteca do ligante TL-3 para o *Refmac* foi utilizado o programa *Monomer Library Sketcher* dentro do *CCP4*, com regularização da estrutura pelo próprio *Refmac* e checagem da estereoquímica. As bibliotecas para os programas *CNS* e *O* foram construídas utilizando as ferramentas do *HIC-Up*¹, assim como a cisteína modificada Tec.

2.7 Análise e Comparação Estruturais

Todas as figuras de representações de densidade eletrônica ou posições atômicas dos modelos foram feitas com o programa *PyMOL* [DeLano(2002)]. Para os cálculos de peso molecular e pI teóricos foi utilizado o programa *ProtParam* contido no *ExPASy*². O cálculo do gráfico de Ramachandran foi realizado pelo programa *Procheck* [Laskowski *et al.*(1993)].

¹<http://alpha2.bmc.uu.se/hicup/>

²<http://us.expasy.org/tools/>

O alinhamento das seqüências das quatro proteases foi feito utilizando-se o programa **ClustalW**³. Para a superposição das estruturas analisadas e cálculo de rmsd posicional foi utilizado o programa **LSQMAN** [Kleywegt(1996)] usando somente os átomos C α . Cálculo dos B fatores médios foi realizado pelo programa **Baverage** (CCP4) e os contatos pelo programa **Contact** (CCP4). Interações tipo van der Waals foram definidas para distâncias de até 4.2Å, e esse parâmetro também foi utilizado na definição dos resíduos formando os bolsões catalíticos [Miller *et al.*(1989b)].

³<http://www.ebi.ac.uk/clustalw/>

Capítulo 3

Resultados e Discussão

3.1 Subclonagem

Os clones das proteases de HIV-1 brasileiras, ou seja *Bmut*, *Fwt* e *Fmut*, foram produzidos por amplificação da sequência desejada à partir do cDNA do vírus (inserido em vetor TOPO) e ligação no vetor de expressão pET11a. Um vetor da linha pET11 foi escolhido no lugar dos pET3 previamente utilizados nas expressões de proteases de HIV (veja, por exemplo, [Ido *et al.*(1991)]) devido à seu controle mais estrito da regulação da transcrição, evitando assim o vazamento da expressão do gene tóxico da protease [Rizzo & Korant(1994)].

A reação de amplificação forneceu somente um produto, contendo ≈ 300 pb, como esperado (Figura 3.1-a). Inicialmente foram realizadas reações de digestão do inserto e vetor para ligação, para as quais não se obteve sucesso. Acreditávamos que o problema estivesse na digestão do inserto, desde que o sucesso da reação de digestão do vetor podia ser visualizado em gel de agarose. Utilizamos várias metodologias, incluindo digestões duplas ou simples, ou seja, a digestão com ambas as enzimas no mesmo tampão ou a digestão com uma das enzimas, purificação do produto e digestão com a segunda enzima, também sem sucesso. Resolvemos então subclonar os insertos amplificados por PCR primeiramente em pCR4Blunt-TOPO, a fim de garantir que o número de pares de base circundando o sítio de corte fossem suficientes, e verificar adequadamente as reações de digestão. Os clones produzidos em pCR4Blunt-TOPO foram confirmados por PCR de colônia e digestão, como exemplificado na Figura 3.1-b.

Figura 3.1: Subclonagem dos genes das proteases brasileiras em pCR4Blunt-TOPO. (a) amplificação por PCR de cada um dos genes; (b) verificação da digestão dupla dos subclones em pCR4Blunt-TOPO, identificados como positivos por PCR de colônia (não mostrado), com as enzimas *NdeI* e *BamHI*. As reações foram realizadas em duplicata e os produtos mostrados foram purificados do gel e utilizados nas reações de ligação subsequentes.

Realizamos reações de ligação com proporções vetor:inserto variando de 1:1, 1:2 e 1:3. As colônias obtidas foram verificadas por PCR de colônia. Desde que o plasmídeo foi sempre desfosforilado, não encontramos falsos positivos. A presença do inserto foi confirmada ainda por digestão dupla. Os clones obtidos pET11a_bmut, pET11a_fwt e pET11a_fmut foram seqüenciados para confirmação da seqüência da protease e armazenados a -20°C .

Visando aumentar a estabilidade das proteases à auto-proteólise, como será discutido um pouco adiante (seção 3.2.2), um conjunto de oligonucleotídeos (Tabela 2.2) foi desenhado de maneira a incluir a mutação Q7K em todas as seqüências. Ainda, a protease *Bwt*, antes construída como uma fusão com 30 nucleotídeos correspondentes à uma porção imediatamente anterior da poliproteína *gag-pol* no clone pET3a_bwt [Ido *et al.*(1991)], foi subclonada de maneira a conter somente o gene da protease de HIV-1. Como pode ser visto na Figura 3.2-a, o procedimento de PCR utilizando esses oligonucleotídeos renderam um único produto de amplificação de tamanho correspondente ao esperado para os genes das proteases. Esses insertos, assim como o vetor pET11a, foram então digeridos com as enzimas *NdeI* e *BamHI*, cujo produto de digestão pode ser visualizado na Figura 3.2-b.

Inserto e vetor foram submetidos à reação de ligação utilizando o sistema “Rapid DNA Ligation” (Roche) e o produto transformado em bactérias DH5 α . O resultado foi verificado por PCR de colônia e digestão dupla do plasmídeo extraído, gerando os clones pET11a_bwtq7k, pET11a_bmutq7k, pET11a_fwtq7k e pET11a_fmutq7k. Todas as colônias escolhidas apresentaram resultados positivos quanto à presença do plasmídeo com inserto da protease (dados não mostrados). Plasmídeos extraídos de duas colônias positivas para cada protease foram seqüenciados, confirmando-se o sucesso do procedimento, e armazenados a -20°C .

Figura 3.2: Subclonagem dos genes das proteases em pET11a com inclusão da mutação Q7K. (a) amplificação por PCR de cada um dos genes. Os produtos de amplificação das reações anteriores (sem mutação Q7K) foram utilizados como marcadores de peso molecular adicionais; (b) verificação da digestão dupla do vetor pET11a com as enzimas *NdeI* e *BamHI* e controle da reação com digestão simples.

3.2 Obtenção das Proteases de HIV

Os experimentos de expressão das proteases de HIV-1 foram sempre realizados visando rendimento e pureza adequados aos experimentos de cristalização. Como as proteases isoladas dos vírus de pacientes brasileiros nunca haviam sido expressas utilizamos sempre a expressão da protease *Bwt* como controle. Inúmeras metodologias de expressão e purificação foram testadas em muitas tentativas infrutíferas de obtenção das proteases, e muitas delas serão somente citadas nesta tese, desde que não levaram a resultados satisfatórios.

3.2.1 Ensaios Preliminares

Os experimentos iniciais utilizaram o clone pET3a_bwt da protease *Bwt* que codifica para uma proteína de fusão de 14 kDa fornecido pelo Dr. Jordan Tang. Com esse sistema foram definidos os tempos adequados para expressão, quantidade de IPTG e temperatura. Desde que as proteases são tóxicas às bactérias, foram utilizadas células BL21(DE3)pLysS para reprimir a expressão basal da protease antes da indução. Após 4h de indução as células foram microprocessadas, os produtos separados em sobrenadante (proteínas solúveis) e corpos de inclusão (proteínas insolúveis) lavados e analisados em gel de poliacrilamida desnaturante. Todos esses expedientes foram suficientes para alcançar uma expressão somente detectável, como pode ser visto no gel da Figura 3.3-a.

Já havia sido descrito que a proteína de fusão poderia ser clivada gerando a proteína nativa durante o processo de expressão, o que levaria à uma mistura entre a protease de 11 kDa e a fusão de 14 kDa, visualizadas na Figura 3.3-a [Ido *et al.*(1991)]. A confirmação da expressão foi realizada por *western blotting* utilizando-se um anticorpo produzido em coelho para 15 aminoácidos da sequência N-terminal da protease selvagem. Como pode ser visto na Figura 3.3-b, as duas bandas indicadas na Figura 3.3-a correspondem à proteína de fusão e à clivada. Do experimento de *western blotting* observa-se também que o sistema utilizado permite a expressão basal da protease na

Figura 3.3: Análise das amostras microprocessadas dos ensaios de expressão da protease *Bwt*. (a) Análise em SDS-PAGE. O gel encontra-se aboulado. As canaletas representam 1 e 2: extratos solúvel e *pellet* antes da indução, 3-6: extratos solúvel e *pellet* (alternados) de dois experimentos de expressão utilizando duas colônias isoladas distintas. Os retângulos pretos indicam a banda com tamanho correspondente à protease fusionada de 14 KDa e os vermelhos indicam a posição relativa à protease clivada de 11 KDa; (b) Experimento de *western blotting* utilizando as mesmas amostras do gel no item (a). A legenda das canaletas também é a mesma.

forma solúvel (canaleta 2), como mostram as bandas antes da indução, mesmo com a presença do plasmídeo pLysS. É possível notar também na Figura 3.3-a que a banda de ≈ 11 kDa indicada pelos retângulos vermelhos desdobra-se em duas, que poderia indicar degradação da enzima.

Esse procedimento de produção da protease *Bwt* foi aplicado à produção da *Bmut* com o clone pET11a-bmut, onde esperávamos obter uma única banda de 11 kDa (proteína não fusionada). Não foi possível observar qualquer expressão utilizando-se células BL21(DE3)pLysS. Como o plasmídeo pLysS reprime a expressão total de proteínas, nos foi sugerido (Elena Gustchina, comunicação pessoal) que fosse utilizada a cepa BL21(DE3), onde obtivemos resultados idênticos ou seja, sem expressão visível. Ao contrário da sequência da *Bwt*, que foi construída de maneira a otimizar os códons preferenciais para *E. coli*, os clones das proteases *Bmut*, *Fwt* e *Fmut* foram produzidos à partir do cDNA de vírus extraídos de pacientes com AIDS e, portanto, possuem códons raros para *E. coli*. Para tentar resolver o problema foram utilizadas bactérias *E. coli* BL21(DE3)-RIL e BL21(DE3)-RP (CodonPlus - Stratagene), que possuem tRNAs ex-

tras para códons raros de arginina, isoleucina, leucina (RIL) e arginina, prolina (RP). O resultado desses testes de expressão também não foram animadores, como mostra a Figura 3.4-a.

A análise desse experimento por *western blotting* revelou que a protease estava sendo produzida, mas em quantidades muito pequenas, como mostra a Figura 3.4-b. É interessante notar que, além da banda relativa à protease em 11 kDa no experimento de *western blotting*, existem também algumas bandas por volta de 6 kDa que indicam degradação da protease durante a expressão. Essa banda de protease degradada é especialmente forte na amostra contendo os corpos de inclusão lavados. Esses dados nos levaram a acreditar que a baixa expressão da protease mutante se devesse essencialmente à degradação da proteína durante a produção em *E. coli*. Esse problema foi tratado como descrito à seguir.

3.2.2 Mutantes Q7K

Os resultados observados nos experimentos de expressão da protease *Bmut* (clone pET11a_bmut) nos levaram a crer que, nas condições em que esses experimentos estavam sendo conduzidos, as proteases estavam sofrendo um processo de autólise, o que justificaria a baixa expressão observada. De fato é sabido que a protease de HIV-1 se auto-digere, levando à inativação da enzima, através da clivagem de três sítios de autólise, sendo eles Leu5-Trp6, Leu33-Glu34 e Leu63-Ile64 [Mildner *et al.*(1994)]. De acordo com os estudos de Rose e col. [Rose *et al.*(1993)] a introdução de uma mutação na glutamina na posição 7 para lisina (Q7K) leva à uma enzima 100 vezes mais estável à auto-degradação quando comparada com a enzima selvagem, desde que essa mutação anula o sítio de clivagem primário localizado na região N-terminal (Leu5-Trp6). Embora nenhum dos três sítios de clivagem encontrem-se acessíveis na proteína dimérica, a estrutura do sítio Leu5-Trp6 mantém-se graças à interações intermoleculares, ou seja, na forma monomérica essas interações são quebradas e essa região pode adotar uma conformação estendida. Os dois outros sítios de clivagem encontram-se engajados em

(a) Ensaio de Expressão

(b) western blotting

Figura 3.4: Ensaio de expressão com a protease *Bmut*. (a) Análise por SDS-PAGE de amostras coletadas no decorrer do ensaio de expressão e também da amostra processada. As canaletas representam LR: marcador de peso molecular (low range - Sigma) *: proteínas totais antes da indução, 1.5h/2.5h/3.5h: tempo de coleta das amostras durante os ensaios (proteínas totais), SPN: sobrenadante da amostra final microprocessada, IB: corpos de inclusão não lavados da amostra final; (b) Análise por *western blotting* das mesmas amostras mostradas no item (a), com as legendas correspondentes.

pontes de hidrogênio de fitas- β intramoleculares, o que lhes garante maior estabilidade conformacional. Assim, a clivagem nos sítios mais internalizados Leu33-Glu34 e Leu63-Ile64 deveria ocorrer somente após a sua exposição devido à clivagem no sítio primário Leu5-Trp6, que bloquearia a formação do dímero ativo [Mildner *et al.*(1994)].

Cada um dos clones obtidos para as proteases *Bwt*, *Bmut*, *Fwt* e *Fmut* com inclusão da mutação Q7K, ou seja, pET11a_bwtq7k, pET11a_bmutq7k, pET11a_fwtq7k e pET11a_fmutsq7k, foram utilizados na transformação de bactérias BL21(DE3) para testes de expressão em microescala. Nesse ponto é importante ressaltar que a adição de 1% de glicose ao meio de cultura foi importante na supressão da expressão basal num sistema sem pLysS. O extrato de proteínas totais provenientes desses ensaios podem ser vistos na Figura 3.5-a. Aparentemente, somente a protease *Bwt* foi expressa em quantidades consideráveis.

As amostras finais microprocessadas foram analisadas em gel de poliacrilamida desnaturante, como mostra a Figura 3.5-b. O microprocessamento confirma que somente a protease *Bwt* apresentou expressão significativa, e mostra ainda a expressão de uma proteína de ≈ 11 kDa, como esperado para a proteína sem fusão.

Como já citado anteriormente, os clones das proteases brasileiras possuem códons raros para *E. coli*. Assim, esses clones foram utilizados na transformação de bactérias *E. coli* BL21(DE3)-RIL e BL21(DE3)-RP (CodonPlus). Realizou-se novamente um ensaio de expressão em microescala de duas colônias para cada clone. Como pode ser visto na Figura 3.6-a, as proteases *Bmut* e *Fwt* podem ser visualizadas já no extrato de proteínas totais das células BL21(DE3)-RIL. O microprocessamento das amostras confirmou a expressão das três proteases em células RIL, como mostra a Figura 3.6-b.

Surpreendentemente, a protease *Fwt* foi parcialmente solubilizada durante a primeira lavagem dos corpos de inclusão e pode ser vista também na fração solúvel do processamento. Aparentemente não se tratou de um artefato, desde que essa solubilidade parcial reproduz-se em todos os experimentos de expressão e purificação da protease *Fwt*.

Figura 3.5: Análise em SDS-PAGE dos ensaios de expressão das proteases contendo a mutação Q7K. (a) extratos de proteínas totais, antes e depois da indução, para duas colônias utilizadas no ensaio de expressão, de acordo com as legendas sobre os géis. As canaletas marcadas com “*” contém a amostra não induzida; (b) produtos de microprocessamento das células após 4h de indução. LAR: marcadores de peso molecular (low range - Sigma), SPN: sobrenadante, IB: corpos de inclusão lavados.

Figura 3.6: Ensaios de expressão das proteases contendo a mutação Q7K em bactérias CodonPlus (RIL e RP). (a) extrato de proteínas totais, antes e depois da indução. Os “*” indicam a amostra não induzida; (b) produto de microprocessamento das células BL21(DE3)-RIL após 4h de indução. As canaletas contém, de acordo com as legendas, MM: marcadores de peso molecular (multimarker - Invitrogen), SPN: sobrenadante, Lav 1 e Lav 2: sobrenadante das lavagens dos corpos de inclusão, IB: corpos de inclusão lavados, LAR: marcadores de peso molecular (low range - Sigma). Os produtos de cada colônia estão marcados pelo número em vermelho.

A análise das proteínas expressas foram realizadas por três métodos: espectrometria de massas, seqüenciamento do N-terminal e *western blotting*, todos indicando que a banda de proteína expressa em ≈ 11 KDa tratava-se da protease de HIV-1 (dados não mostrados). O seqüenciamento de aproximadamente dez resíduos da porção N-terminal confirma também o sucesso da inclusão da mutação Q7K. Desde que o protocolo de expressão das quatro proteases foi estabelecido em microescala, realizou-se, com sucesso, um escalonamento desse processo para a produção das proteases em quantidade suficiente para os ensaios de cristalização.

3.3 Purificação e Reenovelamento das Proteases

Desde que as proteases são expressas em corpos de inclusão, o passo de lavagem dos mesmos pode ser considerado como um passo de purificação, como mostram os géis nas Figuras 3.5-b e 3.6-b. Além disso, o processo de solubilização desses corpos de inclusão compreende uma purificação adicional, desde que uma porção dos contaminantes protéicos dos corpos de inclusão é insolúvel no tampão com 8 M de uréia, como mostra a Figura 3.9-a.

Inicialmente foram realizadas algumas tentativas de purificação por gel filtração preparativa com uma coluna sephacryl S-100 26/60 como primeiro passo de purificação. Esse procedimento não se mostrou eficiente e a pureza alcançada não justificou a inclusão desse passo de purificação, e a gel filtração foi abandonada. Vários experimentos foram então realizados incluindo-se colunas de troca iônica como parte do protocolo de purificação, baseado em dados já publicados de utilização desse procedimento (veja, por exemplo, [Ido *et al.*(1991), Margolin *et al.*(1991)]). O protocolo finalmente estabelecido envolve dois passos de troca iônica na purificação cromatográfica das proteases.

No primeiro passo os corpos de inclusão solúveis foram injetados em uma coluna de troca aniônica (HiTrap Q-sepharose) em pH 8.0. O pI teórico das proteases é de aproximadamente 8.5. Em pH 8.0 a grande maioria dos contaminantes possuía carga total negativa, que fez com que esses contaminantes, incluindo DNA, ligassem à matriz

da coluna, enquanto que a protease passou no *void*. Esse procedimento está ilustrado na Figura 3.7 para a protease *Bmut*. A protease encontra-se no pico **A** no cromatograma, e a amostra analisada em gel pode ser visualizada na Figura 3.9-a. Os picos assinalados como **B** correspondem às impurezas eluídas da coluna com tampão A + 1 M NaCl. A inversão na proporção entre as curvas de absorbância 280 nm/254 nm indica que essas impurezas contém DNA ou proteínas que ligam DNA.

Figura 3.7: Cromatograma da purificação da protease *Bmut* pela coluna HiTrap Q-sepharose. A letra **A** indica a fração contendo a protease, que não se liga à coluna. Os picos indicados pela letra **B** correspondem às impurezas eluídas pela passagem de tampão com NaCl. A curva de concentração no cromatograma indica a porcentagem desse tampão com NaCl.

Num segundo passo, o pH da solução contendo a protease proveniente da coluna de troca aniônica foi ajustado para 5.0. Nesse pH as proteases apresentam carga total positiva e podem ser ligadas à matriz de uma coluna de troca catiônica (HiTrap SP-sepharose), para depois serem eluídas com um gradiente de tampão C contendo NaCl. O cromatograma da Figura 3.8 mostra a eluição da protease da coluna. O ombro marcado com a letra **A** e o pico **B** contém a protease. O ombro **A** é característico nessa etapa de purificação e aparentemente contém uma fração degradada da protease,

conforme analisado por *western blotting* (dados não mostrados). Esse produto de degradação é praticamente invisível no gel da Figura 3.9-b. O pico marcado com a letra C contendo impurezas pode ser mais intenso, dependendo da qualidade da amostra. Essas impurezas, juntamente com aquelas que saem no *void*, são eliminadas por esse procedimento de troca catiônica.

Figura 3.8: Cromatograma da purificação pela coluna HiTrap SP-sepharose da amostra proveniente da etapa mostrada na Figura 3.7. Os picos eluídos marcados com as letras A e B contêm a protease *Bmut* purificada. A curva de concentração no cromatograma indica a porcentagem de tampão C.

A fração da protease *Fwt* solubilizada no tampão de lavagem foi purificada como descrito acima e forneceu proteína com pureza equivalente àquela obtida com os corpos de inclusão lavados. Nenhuma diferença pôde ser observada nos cromatogramas de purificação da protease *Fwt* solubilizada, a não ser por uma maior diluição da amostra inicial carregada na coluna de troca aniônica. Para o ensaio de cristalização não foram discriminadas as amostras da protease *Fwt* de acordo com sua origem (fração dos corpos de inclusão solubilizados ou corpos de inclusão lavados).

As amostras provenientes da segunda etapa cromatográfica encontravam-se, típica-

Figura 3.9: SDS-PAGE da purificação cromatográfica dos corpos de inclusão da protease *Bmut* por troca iônica. (a) Solubilização dos corpos de inclusão em tampão A. Nas legendas temos PP: precipitando após adição de tampão aos corpos de inclusão, SPN: fração solubilizada no tampão A. As amostras provenientes da coluna Q-sepharose (troca aniônica) (Figura 3.7) e as frações eluídas da coluna SP-sepharose (troca catiônica) também estão mostradas nesse gel; (b) Amostras eluídas da coluna SP-sepharose (troca catiônica). Veja cromatograma na Figura 3.8.

mente, em uma concentração entre 0.1-0.2 mg/mL. Amostras mais concentradas que 0.2 mg/mL costumavam apresentar muito precipitado durante a etapa de reenovelamento. Assim, adotou-se como padrão manter a diluição das amostras nesse nível antes da diálise. Normalmente um pouco da amostra se precipita durante o reenovelamento. Aparentemente a quantidade de precipitado é uma função da velocidade do gradiente, ou seja, quão rápido o agente desnaturante é retirado. O procedimento de diluição contínua, com a retirada da uréia continuamente por diluição, ao invés da diluição em passos, foi o que rendeu menos precipitado ao final da diálise [Tsumoto *et al.*(2003)]. A inclusão de 100 mM de NaCl no TFR também ajudou a diminuir a quantidade de precipitado.

Nos primeiros experimentos de reenovelamento foi utilizado um tampão 50 mM NaOAc, pH 5.5, no entanto, ao final da diálise, a protease apresentou-se quase que totalmente degradada. Aparentemente, nesse pH as primeiras proteases enoveladas estão

altamente ativas e encontram uma grande concentração de proteases ainda desenoveladas, com os sítios de auto-clivagem expostos. Para resolver esse problema o tampão foi trocado para pH 7.4 e a temperatura da diálise mantida próxima de zero sempre que possível, de acordo com os dados que mostram que a protease é muito pouco ativa nessa temperatura e nesse pH [Mildner *et al.*(1994)].

3.4 Co-cristalização

Neste trabalho foram cristalizados os complexos das proteases *Bmut*, *Fwt* e *Fmut* com o inibidor TL-3 [Li *et al.*(2000)]. Também foram realizados ensaios com outros inibidores, comerciais e não comerciais, como descrito a seguir. Todas as fotos de cristais mostradas nesta seção estão na mesma escala.

3.4.1 Complexo *Bmut*:TL-3

O inibidor TL-3, que pode ser visto na Figura 3.10, pode ser considerado um inibidor universal, desde que inibe tanto HIV-PR quanto FIV-PR, selvagens e mutantes [Li *et al.*(2000)]. Este inibidor nos foi cedido pelo Dr. Alexander Wlodawer (NCI-Frederick).

As primeiras tentativas de cristalização do complexo *Bmut*:TL-3 foram realizadas utilizando-se sulfato de amônio como precipitante, desde que esse é o precipitante consistentemente utilizado para a cristalização de proteases de HIV-1 desde as primeiras publicações (veja, por exemplo, [Wlodawer *et al.*(1989)] e [Miller *et al.*(1989b)]). Dado

Figura 3.10: Estrutura do inibidor universal TL-3 [Li *et al.*(2000)]. A linha pontilhada indica o eixo de simetria de ordem 2.

o insucesso dessas tentativas um ensaio preliminar para cristalização das proteases foi desenvolvido (veja Tabela 2.7) incluindo-se as condições já estabelecidas para a cristalização das proteases com sulfato de amônio e algumas condições contendo mais de

um precipitante. Para esses ensaios com a protease *Bmut* o tampão de reenvolvimento foi trocado, no concentrador, por outro contendo 50 mM NaOAc, pH 5.0, 2 mM EDTA, 2 mM DTT. Já no primeiro ensaio envolvendo o complexo *Bmut*:TL-3, a 18°C, obtivemos os dois cristais mostrados na Figura 3.11, que apareceram após dois dias da montagem das gotas e continuaram a crescer por aproximadamente duas semanas. O cristal na Figura 3.11-a foi obtido na condição 8 do ensaio (0.1 M cacodilato de sódio, pH 6.2, 0.32 M sulfato de amônio, 6% MPD), e o cristal da Figura 3.11-b foi obtido na condição 13 (0.1 M cacodilato de sódio, pH 6.2, 0.26 M sulfato de amônio, 5% PEG3350, 6% MPD). Nesse ensaio foi utilizada proteína na concentração de 4.3 mg.mL^{-1} ($\approx 0.20 \text{ mM}$) misturada com excesso molar de duas vezes do inibidor TL-3 em DMF/DMSO (1:1 v/v). Os cristais possuíam um tamanho de $0.5 \times 0.15 \times 0.15 \text{ mm}$ (condição 8) e $0.3 \times 0.15 \times 0.15 \text{ mm}$ (condição 13) e cresceram em gotas de $2 \mu\text{L}$ de volume total (1:1 v/v solução de proteína:solução do poço).

Figura 3.11: Cristais da protease *Bmut* complexada com o inibidor TL-3. (a) condição 8 do ensaio preliminar (0.1 M cacodilato de sódio, pH 6.2, 0.32 M sulfato de amônio, 6% MPD); (b) condição 13 do ensaio preliminar (0.1 M cacodilato de sódio, pH 6.2, 0.26 M sulfato de amônio, 5% PEG 3350, 6% MPD).

É interessante notar que a cristalização desse complexo só foi possível graças à inclusão de condições contendo dois ou três precipitantes mecanisticamente distintos.

Nessa época já havíamos tomado conhecimento das dificuldades de cristalização de alguns mutantes de proteases utilizando-se sulfato de amônio ou PEG como precipitantes (Dr. Alexander Wlodawer e Dr. Mi Li, Frederick, MD, comunicação pessoal). Esse fato, e também os trabalhos realizados por Majeed e colaboradores (2003) mostrando a importância de um sistema de multiprecipitantes na cristalização de algumas proteínas, nos levou a formular o ensaio onde combinações de precipitantes foram utilizadas.

3.4.2 Complexo *Fwt*:TL-3

Inicialmente realizou-se o ensaio preliminar a 18°C para o complexo *Fwt*:TL-3 (em TFR, sem troca de tampão) com a proteína na concentração de 5 mg.mL⁻¹ (≈ 0.23 mM) e um excesso molar de duas vezes do inibidor TL-3 em DMF/DMSO (1:1 v/v), gotas de 2 μ L volume final (1:1 v/v solução de proteína:solução do poço).

Figura 3.12: Cristais da protease *Fwt* complexada com o inibidor TL-3. (a) cristais obtidos na condição 4 do ensaio preliminar (0.1 M cacodilato de sódio, pH 6.2, 0.64 M sulfato de amônio) a 18°C; (b) cristais obtidos em 0.1 M cacodilato de sódio, pH 6.2, 0.8 M sulfato de amônio, proporção 2:1 proteína:solução do poço, cristais nucleados a 4°C e posteriormente transferidos para 18°C.

Os melhores resultados foram obtidos nas condições 4 e 5, com gotas polinucleadas com monocristais, como mostrado na Figura 3.12-a. Esses cristais não possuíam tamanho suficiente para coleta de dados. Na condição 15 e 16 (equivalentes das condições 4 e 5 à pH 5.8) também apareceram cristais, mas na forma de *clusters* (dados não

mostrados). Um ensaio de refinamento dessas condições foi realizado baixando-se a concentração de proteína para 3.9 mg.mL^{-1} ($\approx 0.18 \text{ mM}$) mudando-se a proporção proteína:solução do poço para 2:1 (também mantivemos 1:1) em gotas de $9 \mu\text{L}$. Essas mudanças permitiram controlar o processo de nucleação na gota e obtivemos menos cristais de tamanho ligeiramente maior.

A melhor das condições refinadas continha 0.1 M cacodilato de sódio, $\text{pH } 6.2$, 0.8 M sulfato de amônio com proporção 2:1 proteína:solução do poço. Como esses cristais ainda não eram grandes o suficiente, esse mesmo ensaio de refinamento foi repetido a 4°C na esperança de se diminuir ainda mais o número de cristais na gota. A melhor condição nesse caso foi a mesma do ensaio a 18°C , e obtivemos somente dois bastões, que atingiram tamanho máximo de pouco mais de 0.1 mm de comprimento em aproximadamente duas semanas. Esperávamos, no entanto, obter cristais maiores do que os observados, dada a menor quantidade de núcleos presentes na gota em comparação com os resultados obtidos a 18°C . Esse resultado indicou que a proteína é mais solúvel a 4°C que a 18°C . Dado isso, a caixa de cristalização foi transferida para 18°C e esses dois cristais cresceram até um tamanho de $0.4 \times 0.1 \times 0.1 \text{ mm}$, suficiente para a coleta de dados, como mostra a Figura 3.12-b.

3.4.3 Complexo *Fmut*:TL-3

O ensaio preliminar com a protease *Fmut* foi realizado com a proteína na concentração de 3.5 mg.mL^{-1} ($\approx 0.16 \text{ mM}$ em TFR) e um excesso molar de duas vezes do inibidor TL-3 em DMF/DMSO (1:1 v/v). O ensaio foi realizado a 18°C com gotas de $4 \mu\text{L}$ de volume final (1:1 v/v solução de proteína:solução do poço). Apareceram cristais em várias condições do teste, incluindo as condições 3–5, 8, 13–16. Dentre essas condições, as que apresentaram cristais maiores foram as condições contendo somente sulfato de amônio como precipitante (equivalente aos mostrados na Figura 3.12-a para o complexo *Fwt*:TL-3), sendo que nas condições com mais de um precipitante apareceram microcristais bipiramidais.

Baseado na experiência adquirida com a cristalização do complexo *Fwt*:TL-3, re-estruturamos o ensaio e refizemos a 4°C. Esse novo ensaio continha somente sulfato de amônio como precipitante [0.54 – 1.00 M] numa faixa de pH entre 4.8 e 6.2 utilizando-se cacodilato de sódio 0.1 M ou acetato de sódio 0.2 M, dependendo do pH escolhido, e proporção proteína:solução do poço de 1:1 e 2:1. O efeito obtido com esse ensaio foi semelhante ao observado para *Fwt*:TL-3, ou seja, poucos cristais por gota, um pouco maiores que a 18°C. As caixas de cristalização foram então transferidas para 18°C para crescimento adicional dos cristais. Nas melhores condições apareceram bastões entre 0.2–0.3 mm de comprimento. As melhores condições, de onde foram recolhidos cristais para coleta de dados, foram 0.2 M NaOAc, pH 4.8, sulfato de amônio [0.54 M, 0.6 M e 0.8 M], as três usando 1:1 (v/v) solução de proteína:solução do poço; e 0.1 M cacodilato de sódio, pH 6.2, sulfato de amônio [0.64 M, 0.8 M (2:1 e 1:1) e 1 M]. Dois exemplos de cristais que apareceram nessas condições estão mostrados na Figura 3.13.

Figura 3.13: Cristais do complexo *Fmut*:TL-3, nucleados a 4°C e posteriormente transferidos para 18°C. (a) cristal crescido em 0.1 M cacodilato de sódio, pH 6.2, 0.64 M sulfato de amônio, proporção 2:1 (v/v) proteína:solução do poço (0.25 x 0.1 x 0.1 mm); (b) cristal crescido em 0.2 M acetato de sódio, pH 4.8, 0.8 M sulfato de amônio, proporção 1:1 (v/v) proteína:solução do poço (0.3 x 0.15 x 0.15 mm).

Experimentos de *macroseeding* foram realizados em várias condições a fim de aumentar o tamanho dos cristais do complexo *Fmut*:TL-3 e melhorar a resolução dos

dados de difração, sem sucesso.

3.4.4 Complexos com Outros Inibidores

Foram realizadas algumas tentativas de cristalização com outros nove inibidores de protease, seis comerciais e três não comerciais.

Inibidores Não-Comerciais

Ensaio foram realizados com o inibidor LP-149 [Wlodawer *et al.* (1995)], preparado pela Lilly Research Laboratories e gentilmente cedido pelo Dr. Alexander Wlodawer (NCI-Frederick), e dois outros inibidores, TLEU e OAI2, sintetizados pelo Prof. Dr. Octávio Antunes (UFRJ). As estruturas desses inibidores podem ser vistas na Figura 3.14. A integridade dos inibidores TLEU e OAI2 foram confirmadas por experimentos de ^1H -RMN, ^{13}C -RMN e espectroscopia de massa (dados não mostrados).

Figura 3.14: Estruturas dos três inibidores não-comerciais utilizados em ensaios de cristalização.

Ensaio de cristalização dos complexos *Bwt*:TLEU e *Bmut*:TLEU; *Bwt*:OAI2 e *Bmut*:OAI2; *Bmut*:LP-149 e *Fwt*:LP-149 foram realizados utilizando-se o ensaio preliminar obtendo somente precipitado amorfo. Ensaio adicionais foram realizados utili-

zando somente NaCl como precipitante [0.6 M, 0.75 M, 0.9 M e 1.0 M] e variando o pH [4.0, 4.5 e 5.0], também sem resultados positivos.

Inibidores Comerciais

Os inibidores comerciais amprenavir, indinavir, lopinavir, nelfinavir, ritonavir e saquinavir nos foram cedidos pela Fundação Oswaldo Cruz, Farmanguinhos (veja Figura 1.7 na Introdução).

O primeiro inibidor a ser ensaiado foi o indinavir com a protease *Fwt*, utilizando-se o ensaio preliminar. Em várias condições foram obtidos cristais pequenos em forma de bastão (< 0.05 mm de comprimento na maior dimensão), normalmente em *clusters* formando feixes de bastões.

Figura 3.15: Cristais do inibidor indinavir (Figura 1.7-b) obtido por *macroseeding* em 0.1 M cacodilato de sódio, pH 6.2, 0.8 M sulfato de amônio.

A condição 5 (0.1 M cacodilato de sódio, pH 6.2, 0.8 M sulfato de amônio) foi escolhida para fornecer sementes para um experimento de *microseeding*. Para esse experimento a concentração da proteína foi diminuída pela metade e foram ensaiados tanto *streakseeding* quanto *seeding* por moagem. O experimento por moagem rendeu monocristais maiores que puderam ser utilizados em um experimento de *macroseeding*, rendendo o cristal mostrado na Figura 3.15. Infelizmente, após o experimento de difração, constatou-se que esse cristal continha somente o inibidor indinavir e não o complexo *Fwt*:indinavir.

Posteriormente foram realizados ensaios preliminares de todos os seis inibidores com a proteases *Bmut*. Primeiramente foi determinado o melhor solvente para cada inibidor, realizando-se testes de precipitação em meio aquoso, para evitar precipitação/cristalização do inibidor. Chegou-se às seguintes combinações: amprenavir 5% DMF/DMSO (1:1 v/v), indinavir 10% DMF, lopinavir 10% DMF, nelfinavir 5% DMSO, ritonavir 10% DMF, saquinavir 5% DMF/DMSO (1:1 v/v). A porcentagem indica a quantidade final mínima de solvente

na gota para manter o inibidor solúvel. Para esses ensaios acrescentou-se essas quantidades de solvente às soluções do poço de cada uma das condições do ensaio.

Um único cristal pequeno foi obtido com amprenavir (≈ 0.05 mm de comprimento na maior dimensão) na condição 14 do ensaio preliminar (0.1 M cacodilato de sódio, pH 5.8, 0.48 M sulfato de amônio, 5% DMF/DMSO). Infelizmente esse cristal foi perdido durante a transferência do cristal congelado para o fluxo de nitrogênio do sistema crio no LNLS, e assim não pudemos confirmar se se tratava de um cristal de proteína. Com nelfinavir obtivemos vários microcristais de tamanho muito pequeno sequer para testes, na condição 15 do ensaio preliminar (0.1 M cacodilato de sódio, pH 5.8, 0.64 M sulfato de amônio). De maneira geral todas as gotas apresentaram precipitado cristalino, com mais precipitado a pH 6.2 que a 5.8.

3.5 Coleta de Dados e Resolução Estrutural

A seguir apresentaremos os dados relativos às quatro estruturas de proteases de HIV-1 resolvidas nesse trabalho: *Bwt*, *Bmut*, *Fwt* e *Fmut* em complexo com o inibidor TL-3.

Os dados de difração dos complexos *Bmut*:TL-3, *Fwt*:TL-3 e *Fmut*:TL-3 foram inicialmente coletados em nosso equipamento de ânodo rotatório e posteriormente recolhidos no LNLS. O primeiro cristal coletado foi o co-cristal *Bmut*:TL-3 mostrado na Figura 3.11-b. As primeiras imagens de difração coletadas apresentavam alta mosaicidade ($\approx 1.20^\circ$), como mostrado na Figura 3.16-a. Aplicamos então, com sucesso, um procedimento de *annealing* do cristal, após a coleta de um conjunto completo (116 imagens). Os melhores resultados foram observados após três ciclos de *annealing*, que levaram à uma diminuição da mosaicidade da amostra para $\approx 0.40^\circ$ e aumento da resolução, como mostra a Figura 3.16-b. Todos os outros co-cristais coletados posteriormente sofreram um processo de *annealing*, com uma melhora significativa do padrão de difração.

Os conjuntos das proteases *Bmut*:TL-3 (cristal da Figura 3.11-a) e *Fwt*:TL-3 (cristal da Figura 3.12-b) foram finalmente coletados no LNLS. Esses conjuntos foram inequív-

Figura 3.16: Padrões de difração do co-cristal *Bmut*:TL-3 (Figura 3.11-b) coletados em um equipamento de ânodo rotatório. (a) sem *annealing*, mosaicidade de 1.20° e resolução de $2.50 \text{ \AA}$; (b) após três ciclos de *annealing*, mosaicidade de 0.39° e resolução de $2.10 \text{ \AA}$.

camente indexados no grupo hexagonal. A análise das ausências sistemáticas mostrou que existem condições de difração $h\ k\ l = 0\ 0\ 6n$, o que caracteriza um eixo 6_1 , ou seu enantiomórfo 6_5 . Os dados foram escalonados em $P6_122$ e $P6_522$ com estatísticas bastante razoáveis (Tabela 3.1). O coeficiente de Matthews [Matthews(1968)] calculado para cada um dos conjuntos sugeriu a existência de uma única subunidade por unidade assimétrica, com 38.3% e 38.2% de solvente para *Bmut*:TL-3 e *Fwt*:TL-3, respectivamente. Esses dados concordavam com a estrutura previamente publicada do complexo *Bwt*:TL-3, indexada em $P6_122$ [Li *et al.*(2000)].

Dada a grande quantidade de estruturas depositadas existentes de proteases de HIV-1 com diferentes inibidores, optou-se pela resolução estrutural por substituição molecular. O modelo para substituição molecular foi escolhido através de uma busca por estruturas de proteases de HIV no *RCSB Protein Data Bank*, PDB¹, utilizando-se a similaridade seqüencial como parâmetro de busca. Das estruturas com maior

¹<http://www.rcsb.org/pdb>

Tabela 3.1: Dados cristalográficos e estatísticas para os conjuntos *Bmut*:TL-3 e *Fwt*:TL-3 coletados no LNLS e processados em P6₁22. Os valores relativos à última camada de resolução estão mostrados em parênteses, quando apropriado. Os valores para a segunda melhor solução de substituição molecular também estão mostrados entre parênteses.

Dados e Parâmetros	<i>Bmut</i>	<i>Fwt</i>
Grupo Espacial	P6 ₁ 22	P6 ₁ 22
Parâmetros de Cella [<i>a</i> = <i>b</i> , <i>c</i>]	60.94 Å, 82.48 Å	61.46 Å, 80.90 Å
Resolução (Å)	52.71—1.80 (1.90—1.80)	53.22—2.05 (2.16—2.05)
n ^o subunidades na u.a.	1	1
V _m (Coef. de Matthews)	2.0 Å ³ Da ⁻¹	2.0 Å ³ Da ⁻¹
Completeza (%)	99.6 (99.6)	99.9 (99.9)
I/σ(I)	27.6 (6.2)	21.9 (4.9)
R _{sym} (%)	5.3 (36.9)	5.6 (43.0)
N ^o de Reflexões Únicas	8892	6091
Multiplicidade	9.9 (10.1)	7.8 (8.0)
Substituição Molecular*		
R _{fator}	42.8% (55.5%)	37.2% (54.6%)
Correlação F	59.4% (36.9%)	71.3% (35.7%)
Correlação I	35.9% (33.4%)	72.5% (36.4%)

* O programa AMoRe foi utilizado, utilizando-se dados entre 10–3 Å de resolução.

similaridade foi escolhida a de maior resolução. Tratava-se de um complexo de uma protease de HIV-1 subtipo B selvagem com o inibidor JE-2143 à 1.09 Å de resolução (código PDB 1KZK) [Reiling *et al.*(2002)].

A ambigüidade entre os grupos enantiomórfos P6₁22 e P6₅22 foi resolvida após o passo de translação na substituição molecular realizado pelo programa AMoRe, que mostrou claramente tratar-se do grupo espacial P6₁22. Inicialmente resolveu-se a estrutura da protease *Bmut*:TL-3 e posteriormente utilizou-se essa estrutura parcialmente refinada para resolver a *Fwt*:TL-3 (veja discussão mais adiante). A segunda parte da Tabela 3.1 mostra o contraste entre as duas primeiras soluções de substituição molecular para cada um dos conjuntos resolvidos.

No início do refinamento dos dados da protease *Bmut*:TL-3 (conjunto coletado em ânodo rotatório) encontramos um problema de instabilidade do parâmetro R_{free}, como mostra a Figura 3.17-a. Esse problema foi resolvido atribuindo-se mais reflexões ao conjunto utilizado para o cálculo de R_{free}. Conforme um consenso existente (veja, por exemplo, [Dodson *et al.*(1996), Chen *et al.*(1999)]), o conjunto de validação cru-

zada (R_{free}) deve conter entre 500 – 1000 reflexões para que possua alguma relevância estatística. O conjunto inicial da protease *Bmut*:TL-3 coletado em ânodo rotatório continha 6035 reflexões únicas (resolução de 2.1Å) o que fornecia um total de ≈ 300 reflexões para um conjunto de R_{free} de 5% das reflexões totais (atribuição *default* do programa SCALA). A atribuição de 10% das reflexões totais para o conjunto teste resolveu o problema. Posteriormente, com o conjunto coletado no LNLS à maior resolução (Tabela 3.1) essa porcentagem pôde ser reduzida para 7%. A Figura 3.17-(b,c) mostra um exemplo típico de um refinamento estável à partir da re-atribuição do número de reflexões do conjunto teste.

Figura 3.17: Valores de R_{factor} e R_{free} para cada iteração do programa de refinamento. (a) Refmac com ≈ 300 reflexões atribuídas ao conjunto teste (5%); (b) CNS com ≈ 600 reflexões para o conjunto teste (10%), primeiro ciclo, sem inibidor. **A:** corpo rígido, **B:** minimização, **C:** *simulated annealing*, **D:** minimização, **E:** minimização B fatores individuais; (c) Refmac com ≈ 600 reflexões para o conjunto teste (10%).

A estrutura da *Bmut*:TL-3 coletada no LNLS pôde ser refinada até o ponto em que a inclusão de moléculas de solvente se fazia necessária, com $R_{factor} = 0.223$ e $R_{free} = 0.265$, passando-se pelo programa CNS e depois para o Refmac. Ao acrescentar as primeiras moléculas de água, embora a distância entre os R fatores se mantivessem aceitáveis, como no macrociclo sem águas, havia um ligeiro aumento no valor final do R_{free} em oposição à uma diminuição no valor de R_{factor} , comportamento inaceitável num processo de refinamento. Vários procedimentos foram utilizados na tentativa de se resolver esse problema, entre eles o acréscimo de apenas algumas águas manualmente

ao invés de utilizar o programa **ARP_waters**, ainda sem sucesso. Nesse momento, uma análise cuidadosa das saídas do programa **SCALA** indicou alta anisotropia dos dados, que poderia explicar o comportamento anormal dos valores de R durante o refinamento. A fim de verificar se esse era um comportamento particular desse conjunto de dados, refinou-se a estrutura da *Fwt*:TL-3 resolvida à partir de substituição molecular utilizando-se a estrutura parcialmente refinada da *Bmut*:TL-3 como modelo. O programa **SCALA** mostrava o mesmo comportamento de espalhamento anisotrópico, e os dados puderam ser refinados também até o ponto em que as moléculas de solvente deveriam ser adicionadas, com $R_{factor} = 0.236$ e $R_{free} = 0.288$. Novamente tentamos adicionar somente algumas águas manualmente, igualmente sem sucesso. Em ambas as estruturas os desvios de ângulo e distância de ligação foram monitorados de forma a obter valores apropriados para cada resolução. Observamos também, em ambos os conjuntos de dados, que o comportamento dos valores de R_{free} por faixa de resolução não correspondia ao esperado, com valores excessivamente altos em algumas faixas de resolução intermediária.

Na época esse problema foi postado na lista de discussão do CCP4² onde foi proposto que esse tipo de comportamento poderia advir da atribuição equivocada do grupo espacial, muito embora os valores de R_{sym} não indiquem uma anomalia em nenhum dos dois casos. Mesmo assim os conjuntos foram reprocessados de maneira a regenerar o dímero biológico na unidade assimétrica, ou seja, a retirada de um eixo cristalográfico de ordem 2. Como já mencionado anteriormente, a análise das extinções sistemáticas indicam claramente tratar-se de um dos dois grupos enantiomórfos $P6_1$ ou $P6_5$, excluindo assim o grupo espacial $P3_122$. Curiosamente, o reprocessamento dos dados em $P6_1$ eliminou o problema da anisotropia e permitiu o refinamento dos dados das quatro estruturas deste trabalho. Como para o grupo $P6_1$ existe uma lei de *twinning*, os valores dos segundos momentos foram checados na saída do programa **SCALA** e também pelo *site* mantido pelo Dr. Todd Yeats, *The Merohedral Crystal Twinning Server*³, e em

²CCP4bb; <http://www.ccp4.ac.uk/ccp4bb.html>

³<http://www.doe-mbi.ucla.edu/Services/Twinning/>

nenhum dos casos foram detectadas a presença de *twinning* merohedral perfeito. Já os testes para frações parciais de *twinning* merohedral parecem indicar uma fração muito próxima de meio. A presença de simetria não cristalográfica e dados anisotrópicos podem levar à uma falsa interpretação da fração de *twinning*, e como teste resolvemos refinar os dados sem levar em conta a possibilidade de *twinning*, que no final revelou-se realmente infundada, dado o sucesso no refinamento das estruturas. As estatísticas relativas aos dados reprocessados em $P6_1$ estão apresentadas na Tabela 3.2.

O complexo *Fmut*:TL-3 também foi coletado no LNLS obtendo-se, no entanto, uma resolução moderada de 2.8Å. Experimentos de desidratação desses cristais, aliados aos procedimentos de *annealing*, foram realizados na tentativa de melhorar a qualidade dos cristais, sem melhora no padrão de difração. Acreditamos que o insucesso na desidratação se deva ao baixo conteúdo de solvente desses cristais ($\approx 40\%$).

A Tabela 3.2 inclui também os dados do conjunto *Bwt*. Esse conjunto havia sido processado e refinado em $P6_122$ pelo grupo do nosso colaborador Dr. Alexander Wlodawer (NCI-Frederick) [Li *et al.*(2000)] e com meio inibidor TL-3 modelado no sítio ativo. A diferença nos valores dos fatores R_{factor} e R_{free} nessa estrutura era de 9.5%, o que poderia indicar o mesmo problema de atribuição equivocada do grupo espacial. O Dr. Wlodawer gentilmente nos forneceu as imagens de difração desse conjunto para serem reprocessados e re-refinados no grupo espacial $P6_1$. O reprocessamento nos permitiu confirmar que realmente havia ocorrido um erro na atribuição do grupo $P6_122$ também para esse conjunto.

Tabela 3.2: Dados cristalográficos e estatísticas para as quatro proteases de HIV-1 deste trabalho: *Bwt*, *Bmut*, *Fwt* e *Fmut*, todas co-cristalizadas com o inibidor TL-3 [Li *et al.*(2000)]. Os dados das proteases *Bmut* e *Fwt* mostrados na Tabela 3.1 foram reprocessados no grupo espacial *P6₁* (veja texto para discussão detalhada). Os valores relativos à última camada de resolução estão mostrados em parênteses, quando apropriado. Os valores para a segunda melhor solução de substituição molecular também estão mostrados entre parênteses.

Dados e Parâmetros		Bwt	Brotatório	Bmut	Síncrotron (LNLS)	Fwt	Síncrotron (LNLS)	Fmut	Síncrotron (LNLS)
Fonte de Luz		P6 ₁		P6 ₁	P6 ₁	P6 ₁		P6 ₁	
Grupo Espacial		63.16 Å, 83.41 Å		60.93 Å, 82.46 Å	61.43 Å, 80.89 Å			61.25 Å, 82.23 Å	
Parâmetros de Cella [<i>a</i> = <i>b</i> , <i>c</i>]		54.72–2.10		52.78–1.75	53.23–2.10			53.07–2.80	
Resolução (Å)		(2.21–2.10)		(1.84–1.75)	(2.21–2.10)			(2.95–2.80)	
nº subunidades na u.a.		2		2	2			2	
<i>V_m</i> (Coeficiente de Matthews)		2.2 Å³ Da ^{−1}		2.0 Å³ Da ^{−1}	2.0 Å³ Da ^{−1}			2.0 Å³ Da ^{−1}	
Completeza (%)		99.7 (99.7)		99.6 (99.6)	99.7 (99.7)			99.9 (99.9)	
<i>I</i> / <i>σ</i> (<i>I</i>)		13.6 (2.7)		17.8 (3.4)	13.0 (3.1)			9.8 (2.0)	
<i>R_{sym}</i> (%)		7.5 (41.4)		3.7 (34.8)	4.6 (31.8)			8.8 (48.2)	
Nº de Reflexões Únicas		11056		17548	10169			4354	
Nº de Reflexões Conjunto Teste		1086		887	1014			419	
Multiplicidade		3.7(3.7)		4.8 (4.8)	2.5 (2.4)			3.2 (3.2)	
Refinamento									
<i>R_{factor}</i>		0.159		0.183	0.194			0.183	
<i>R_{free}</i>		0.219		0.232	0.265			0.260	
rmsd distância de ligação (Å)		0.015		0.015	0.013			0.015	
rmsd ângulo de ligação (º)		1.575		1.777	1.593			1.690	
Gráfico Ramachandran									
Favoráveis (%)		94.3		94.2	95.6			90.2	
Permitidos (%)		5.7		5.8	4.4			7.3	
Generosamente Permitidos (%)		0.0		0.0	0.0			1.2	
Não Permitidos (%)		0.0		0.0	0.0			1.2	
moléculas de água		98		189	117			50	
número de duplas conformações		8		20	13			2	
número de cisteínas modificadas		1		2	2			0	

Como citado no parágrafo anterior, na referência original [Li *et al.*(2000)] somente metade do inibidor TL-3 foi refinado, fazendo uso do eixo de ordem 2 dessa molécula. Seguindo-se essa metodologia nossos refinamentos iniciais também utilizaram meio ligante, com as mesmas bibliotecas utilizadas no refinamento do complexo *Bwt*:TL-3. No entanto, ao analisar a densidade eletrônica para o TL-3, e o ligante completo gerado por simetria cristalográfica, algumas anomalias foram notadas. De acordo com as regiões indicadas na Figura 3.18 observamos: *i*) existiam três lóbulos na densidade eletrônica da região central do ligante TL-3 (sobre o eixo de ordem 2), enquanto que seria de esperar somente dois, correspondentes às duas hidroxilas do diol; *ii*) um pico no mapa diferença observado à 3σ , sobre o lóbulo central, indicava a ausência de modelo nessa posição; (*iii*) a ligação entre os carbonos que formam o diol, que encontrava-se em posição especial, era muito longa (2.85Å) para uma ligação carbono-carbono. Além disso, por esse procedimento, não haveria como descrever a quiralidade dos carbonos envolvidos na ligação diol.

Figura 3.18: Inibidor TL-3 modelado como meio ligante no sítio ativo da protease *Bmut* (P6₁22). A densidade eletrônica $2|F_{obs}| - |F_{calc}|, \alpha_{calc}$ (1σ , em azul) e $|F_{obs}| - |F_{calc}|, \alpha_{calc}$ (3σ em verde) está representada, assim como parte do ligante gerado por operação de simetria cristalográfica. As marcações na figura estão explicadas no texto.

Para resolver esses problemas foram construídas bibliotecas para o inibidor completo, o que possibilitou restringir a distância da ligação carbono-carbono do diol, além da quiralidade desses carbonos. O refinamento do inibidor completo com ocupação meio nos mostrou que uma das hidroxilas do diol ocupava o lóbulo central da densi-

dade eletrônica do inibidor, enquanto que os outros dois lóbulos laterais eram ocupados pela outra hidroxila em cada uma das duas conformações distintas gerada pela operação de simetria cristalográfica, como mostra a Figura 3.19-b, resolvendo os problemas mostrados na Figura 3.18. Dada a simetria da protease e do próprio inibidor, não existe favorecimento de nenhuma das conformações do inibidor frente à outra. Assim ambas as conformações foram refinadas com ocupação 0.5 para cada. É interessante notar a alta ocupação do inibidor TL-3 nas estruturas resolvidas. A densidade para o inibidor completo pode ser observada logo após a substituição molecular, antes do primeiro macrociclo de refinamento, ainda sem modelar o ligante, como exemplificado na Figura 3.19-a.

Uma particularidade do empacotamento da protease no grupo hexagonal é a desordem da Phe53 em ambas as subunidades (veja, por exemplo, [Kervinen *et al.*(1998)]), que devem ser modeladas em duas orientações mutuamente exclusivas devido ao posicionamento espacialmente conflitante das cadeias laterais de duas subunidades cristalograficamente relacionadas.

Durante o refinamento observou-se uma extensão da densidade da cisteína 67 em algumas das proteases refinadas (Figura 3.20-a). Tanto o cacodilato de sódio quanto o β -mercaptoetanol, ambos presentes nas condições de cristalização, poderiam ter modificado este resíduo. Por esse motivo foram testadas tanto a cisteína modificada com cacodilato de sódio, chamada Cas no HIC-Up (Figura 3.20-b), como modificada com β -mercaptoetanol, construída para esse propósito e chamada Tec (Figura 3.20-c). O refinamento mostrou tratar-se da Tec. A Tabela 3.2 mostra o número de cisteína modificadas por β -mercaptoetanol em cada uma das estruturas, assim como os dados mais relevantes no final do refinamento.

Uma análise cuidadosa das estruturas refinadas mostra o motivo da quebra de simetria cristalográfica que obrigou o abaixamento da simetria para $P6_1$. Ambas as subunidades são relacionadas por uma simetria não cristalográfica paralela à cristalográfica, porém algumas conformações das cadeias laterais são diferentes entre as duas

Figura 3.19: Representações estereográficas da região do sítio ativo da protease *Bmut* (P6₁). (a) densidade eletrônica $2|F_{obs}| - |F_{calc}|, \alpha_{calc}$ (1σ , em azul) e $|F_{obs}| - |F_{calc}|, \alpha_{calc}$ (3σ em verde) logo após a substituição molecular (Phaser); (b) modelo final com o inibidor TL-3 (dupla conformação) na densidade mostrada em (a).

Figura 3.20: Refinamento da cisteína modificada 67, mostrando os mapas de densidade eletrônica $2|F_{obs}| - |F_{calc}|, \alpha_{calc}$, contornados à 1σ (em azul), $|F_{obs}| - |F_{calc}|, \alpha_{calc}$, contornados à 3σ (em verde) e à -3σ (em vermelho). (a) cisteína 67, antes de incluir a modificação; (b) modificação com cacodilato de sódio (Cas); (c) modificação com β -mercaptoetanol (Tec), mostrando dupla conformação entre Cys-Tec.

subunidades, como a modificação da cisteína 67 na protease *Bwt*, por exemplo. A diferença mais marcante, no entanto, ocorre na estrutura das moléculas de solvente. Isso explica o aumento do R_{free} no refinamento em $P6_122$ ao acrescentar moléculas de água, onde padrões diferentes de solvente foram forçados a se relacionar por simetria cristalográfica. Alguns exemplos de regiões não idênticas entre as duas subunidades estão mostrados na Figura 3.21. Vale ressaltar que todas as estruturas foram refinadas usando-se simetria não cristalográfica, o que indica que as diferenças observadas são genuínas e que o abaixamento da simetria não foi artificialmente incluído para facilitar o refinamento.

3.6 Análise e Comparação Estruturais

A grande maioria dos inibidores de protease de HIV-1 liga-se no sítio ativo em uma conformação estendida. O inibidor TL-3 não é uma exceção. Neste trabalho o inibidor foi refinado com dupla conformação em cada uma das quatro estruturas resolvidas (veja Figuras 3.19-b e 1.2). Como já citado anteriormente (seção 3.5) não existe favorecimento de uma conformação do inibidor TL-3 frente à outra, no entanto, embora o inibidor tenha sido refinado com dupla conformação, a promediação dos resíduos do

Figura 3.21: Diferenças entre padrões de solvente e das conformações de cadeias laterais entre as subunidades A e B, gerando a quebra de simetria cristalográfica. (a) alinhamento da cadeia A com a cadeia B da protease *Bmut*. Observe a quebra de equivalência no padrão de águas; (b-c) protease *Bwt*, em torno do resíduo Thr4; (d-e) protease *Fwt*, em torno do resíduo Ser63. Observe que a Ser63B apresenta dupla conformação, que não ocorre na cadeia A.

sítio ativo pela falta de especificidade conformacional na cristalização pode ter levado à um mascaramento de pequenas diferenças existentes entre as duas subunidades.

Figura 3.22: Comparação da ligação peptídica entre os subsítios P3 e P4 nas estruturas do complexo *Bwt*:TL-3. (a) representação esquemática da dos subsítios P3 e P4 do inibidor TL-3; (b) estrutura do complexopreviamente publicado em P6₁22 (magenta) e resolvido em nosso grupo em P6₁ (verde). A densidade eletrônica $2|F_{obs}| - |F_{calc}|, \alpha_{calc}$, foi contornada à 1σ para a estrutura em P6₁.

No artigo original do complexo *Bwt*:TL-3 [Li *et al.*(2000)] os autores descrevem uma rotação de $\approx 180^\circ$ da ligação peptídica entre o grupo Cbz em P4/P4' e a Ala em P3/P3' com relação à orientação normalmente adotada por esse ligação nos complexo de proteases retrovirais com inibidores peptidiomiméticos. Nas estruturas resolvida em nosso grupo, inclusive a *Bwt*:TL-3 reprocessada em P6₁, não foi encontrada nenhuma evidência dessa rotação anormal da ligação peptídica (Figura 3.22). Acreditamos que essa rotação observada pelos autores se deva, ao menos em parte, à atribuição errônea do grupo pontual e do refinamento de meia molécula de TL-3 promediada no sítio ativo.

3.6.1 Comparação Global das Quatro Estruturas com TL-3

As diferenças existentes entre cada uma das seqüências de proteases pode ser observada no alinhamento apresentado na Figura 3.23. Para facilitar a visualização as diferenças nas proteases *Bmut*, *Fwt* e *Fmut* com relação à *Bwt* foram representadas na Figura 3.24-(a-c). Dentre essas, as mutações M46I, I54V, V82A e L90M são direta-

mente associadas com resistências a diversos inibidores de proteases (veja Figura 1.9 [Shafer(2004)]). Outras mutações secundárias, também envolvidas com resistência a inibidores, quando associadas com mutações primárias , incluem L10I, K20R, M36I, L63P/S e A71V. Ainda, algumas posições são altamente conservadas em indivíduos não-tratados, e mutações nessas posições têm sido associadas a tratamento com anti-retrovirais. Dentre as mutações existentes nessas quatro proteases, a K45R na *Bmut* e a T74A na *Fmut* incluem-se nessa categoria [Shafer(2004)].

Figura 3.23: Alinhamento das seqüências das proteases *Bwt*, *Bmut*, *Fwt* e *Fmut*. Os símbolos “!” indicam pontos da seqüência conservados entre todas as quatro proteases.

Tabela 3.3: Rmsd calculado entre as estruturas *Bwt* e *Fwt* e todas as outras estruturas resolvidas após alinhamento estrutural.

	<i>Bmut</i>	<i>Fwt</i>	<i>Fmut</i>
<i>Bwt</i>	0.43 Å	0.58 Å	0.55 Å
<i>Fwt</i>	0.44 Å	—	0.47 Å

O rmsd entre os 198 Cα das estruturas selvagens e todas as outras estão mostradas na Tabela 3.3. Esses valores indicam uma alta similaridade estrutural entre as prote-

Figura 3.24: Vista frontal ou face (esquerda) e do topo (direita) das proteases de HIV-1. (a-c) Posições das mutações nas proteases *Bmut*, *Fwt* e *Fmut* de acordo com a sequência da protease *Bwt*. Veja também o alinhamento na Figura 3.23; (d) Representação $C\alpha$ das quatro proteases deste trabalho alinhadas: *Bwt*, *Bmut*, *Fwt*,

ases, e são, como era de se esperar, comparáveis àqueles observados entre estruturas de HIV-1 protease cristalizadas em grupos espaciais diferentes [Kervinen *et al.*(1996)]. As maiores diferenças observadas encontram-se entre os resíduos 35 e 41 localizados num *loop* entre os as fitas- β *d* e *a'* (veja Figura 1.2) e que constituem uma dobradiça do *flap*. Esses dados nos levam a esperar diferenças muito pequenas entre os modos de ligação do inibidor TL-3 em cada uma das estruturas resolvidas. Para ilustrar essas diferenças, as estruturas alinhadas estão representadas na Figura 3.24-d.

Os fatores de temperatura por resíduo das estruturas *Bwt*, *Bmut* e *Fwt* são comparáveis, como mostra a Figura 3.25. A exceção é a protease *Fmut*, com fatores de temperatura ligeiramente mais elevados, provavelmente devido à menor resolução dessa estrutura. Os valores médios dos fatores de temperatura para cada uma das estruturas são de 32.83 Å², 30.06 Å², 36.11 Å², 47.49 Å², respectivamente para *Bwt*, *Bmut*, *Fwt* e *Fmut*. Os valores médios para os inibidores TL-3 no sítio ativo dessas mesmas estruturas são de 20.10 Å², 20.72 Å², 24.87 Å² e 21.05 Å². Nota-se que mesmo para a estrutura *Fmut* o valor do fator de temperatura médio para o inibidor é comparável às outras três. Como mostra a Figura 3.25 esses valores também são mais baixos para os resíduos que compõem o sítio ativo, desde que essa é uma região rígida e bem definida em todas as estruturas. Ainda, os resíduos com valores médios mais elevados são aqueles que compõem o *loop* entre os resíduos 35-45, que também apresentam o maior rmsd entre as estruturas analisadas.

A superposição dos inibidores TL-3 ligadas no sítio ativo das proteases *Bwt*, *Bmut*, *Fwt* e *Fmut* está mostrada na Figura 3.26. Para facilitar a visualização, as duplas conformações do ligante foram separadas em alinhamentos distintos. Desse alinhamento pode-se verificar que o inibidor TL-3 liga-se às quatro PRs de maneira bastante similar. Inclusive, as conformações A e B do inibidor sobrepõem-se quase perfeitamente quando a cadeia A é sobreposta à cadeia B da mesma estrutura. Os átomos nos subsítios P1-P3 e P1'-P3' sobrepõem-se sem diferenças aparentes, enquanto que os resíduos Cbz nos subsítios P4 e P4' possuem conformações ligeiramente distintas entre as conformações

Figura 3.25: Gráfico de B fatores por resíduo das quatro estruturas resolvidas. Os resíduos numerados 1-99 pertencem à cadeia A e aqueles numerados 101-199 à cadeia B. (a) média dos valores para a cadeia principal; (b) média dos valores para as cadeias laterais (exceto glicinas).

A e B, provavelmente pela falta de fechamento dos bolsões S4 e S4'. Essa equivalência posicional dos inibidores demonstra que, embora o inibidor seja obrigado a interagir com a enzima de maneira assimétrica, os dois modos de ligação desses inibidores, hora desviado na direção de P1, hora na de P1', são equivalentes e, portanto, não houve um favorecimento de uma conformação perante a outra durante a cristalização dos complexos. Essa equivalência é ainda mais notável se considerarmos que essas duas conformações foram refinadas independentemente, sem nenhuma relação de simetria imposta entre elas.

3.6.2 Interações entre TL-3 e PRs no Sítio Ativo

Como as conformações A e B do inibidor são equivalentes, para facilitar a visualização somente uma das conformações será mostrada nas figuras a seguir. No sítio P4 e P4', onde essas conformações são distintas ambas serão representadas. Por definição [Schechter & Berger(1967)], os subsítios P1-Pn correspondem àqueles resíduos do lado amino-terminal da ligação hidrolisável e, de maneira similar, P1'-Pn' àqueles resíduos do lado carboxi-terminal. Assim os bolsões de ligação correspondentes a cada subsítio são chamados Sn e Sn'. No caso de inibidores simétricos essa atribuição é

(a) conformação A

(b) conformação B

Figura 3.26: Superposição das coordenadas do inibidor TL-3 no sítio ativo das proteases *Bwt*, *Bmut*, *Fwt*, (respeitando as mesmas cores da Figura 3.24-d). As fitas transparentes representam o enovelamento da cadeia principal da protease *Bwt*. (a) cadeias marcadas como conformação A com indicação dos resíduos catalíticos Asp25 e dos subsítios P1'-P4' no inibidor (P1-P4 são idênticos por simetria) ; (b) cadeias marcadas como conformação B.

arbitrária, no entanto, uma vez definidas as cadeias A e B, os bolsões ficam automaticamente definidos pelos resíduos que os compõem, devendo os subsítios no inibidor seguir a nomenclatura correspondente.

Os resíduos que compõem um bolsão S_n e seu simétrico S_n' são essencialmente os mesmos nas quatro estruturas. Observando a Figura 3.27-(a,b) nota-se que há uma assimetria no modo de ligação em $P1-S1$ e $P1'-S1'$, ou seja, $P1$ não interage com seu bolsão $S1$ da mesma maneira que $P1'$ com $S1'$. Isso ocorre porque, embora o inibidor TL-3 seja simétrico, seu modo de ligação é assimétrico, pois somente uma das hidroxilas do diol pode se orientar adequadamente entre os dois resíduos de ácido aspártico catalíticos. Essa assimetria força a hidroxila que não interage diretamente com os dois Asp25 a se posicionar no interior do bolsão $S1'$. Esse deslocamento empurra a cadeia lateral da Phe em $P1'$ nas estruturas *Bwt* e *Fwt* na direção de uma cavidade formada entre os resíduos Gly49A-Ile50A na ponta do *flap* e Pro81B (veja Figura 3.27-(b), à esquerda), agindo como uma cunha. Já em $P1$, as cadeias laterais posicionam-se no meio do bolsão $S1$ e estão ligeiramente deslocadas na direção do resíduo na posição 82 (topo da Figura 3.27-a).

No bolsão $S1$ e $S1'$ a mutação V82A é a única existente (veja Figura 3.23). Nas estruturas *Bwt* e *Fwt*, com Val82, o anel da Phe em $P1/P1'$ interage com a cadeia lateral da Val82 através de interações van der Waals. Nessas estruturas, o anel em $P1'$ se afasta da Val82 ao se inserir na fenda formada entre a ponta do *flap* e o *loop* contendo o resíduo Pro81B, como descrito no parágrafo anterior. Esse afastamento, no entanto, não altera o número total de interações van der Waals que cada subsítio $P1/P1'$ faz com seu respectivo bolsão. Nas estruturas *Bmut* e *Fmut*, onde ocorre a mutação V82A, a cadeia lateral em $P1'$ sofre um rearranjo e se desloca na direção do buraco gerado pela mutação. Esse rearranjo compreende uma rotação de 53° do anel em $P1'$ entre as estruturas *Bwt* e *Bmut*, e 87° para as estruturas *Fwt* e *Fmut*. Essa mudança na conformação do inibidor é acompanhado de um rearranjo também na proteína, através de um ligeiro colapso dos resíduos na ponta do *flap* e do resíduo Pro81B para dentro

Figura 3.27: Representação estereográfica dos bolsões S1 e S1' ocupados pelos resíduos de Phe em P1 e P1', respectivamente, com a posição do subsítio vizinho P2/P2' indicada. Cada estrutura está representada nas cores *Bwt*, *Bmut*, *Fwt*, A identificação dos resíduos está feita somente para a protease *Bwt*. Nesse bolsão existe a mutação V82A nas duas estruturas mutantes, o que causa um rearranjo da cadeia lateral em P1', nas estruturas *Bmut* e *Fmut*, com uma rotação do anel da fenilalanina na direção do buraco deixado pela mutação. No bolsão S1 as cadeias laterais da Phe em P1 ocupam posições estruturalmente equivalente, no meio do bolsão.

do bolsão S1'. Esse estreitamento aumenta a interação do bolsão com o anel da Phe e, como consequência, diminui o tamanho da abertura na fenda gerada entre esses *loops*, alterando a forma do bolsão. Esse re-empacotamento em S1' só é possível graças à diminuição do volume da cadeia lateral na posição 82, que gera um espaço e permite o deslocamento do resíduo Pro81B na direção da Ile84B. Na protease *Bwt* a cadeia lateral da Ile84B possui dupla conformação, com o átomo C δ 1 apontando para dentro do sítio ativo em uma conformação, e com uma rotação de $\approx 180^\circ$ na outra conformação. O rearranjo do bolsão S1' aparentemente leva à um enrijecimento da cadeia lateral da Ile84B, com o C δ 1 em uma única conformação e apontando para fora do sítio ativo.

Em S1 existe o rearranjo somente na proteína, estreitando a fenda entre o *flap* e o *loop* que forma o bolsão, mas sem um rearranjo no inibidor, que nesse caso não se insere nessa fenda em nenhuma das estruturas (Figura 3.27-a). Esse estreitamento da fenda contribui para o aumento do número de interações desse bolsão com o subsítio, muito embora o buraco deixado pela mutação por alanina na posição 82, que não é ocupado pelo anel da Phe em P1, deveria diminuir o número de interações totais nesse bolsão, quando comparado com as estruturas selvagens. No entanto, o que se observa é que o número de interações entre os átomos em P1 com o bolsão S1 nas estruturas selvagens e mutantes não diferem significativamente. Acreditamos que trata-se de um artefato, onde a assimetria existente nas interações P1-S1 e P1'-S1' está parcialmente mascarada pela promediação da dupla conformação do sítio ativo, como discutido ao final da seção 3.6.

Todas essas observações, relatadas para as interações em S1/S1', estão em conformidade com as previamente publicadas por Baldwin *et al.* (1995), onde um inibidor simétrico (A-77003) com uma fenilalanina em P1 e um diol como ligação não hidrolisável foi cristalizado com a protease subtipo B selvagem e um mutante pontual V82A. Nesse estudo o modo de ligação do inibidor também é assimétrico, como observado para o TL-3, e a mutação V82A tende a alterar mais a conformação em P1 que em P1'. É interessante que essas assimetrias nas interações P1-S1 e P1'-S1', assim como os re-

Figura 3.28: Representação estereográfica dos bolsões S2 e S2' ocupados pelos resíduos de valina em P2 e P2', respectivamente, com a posição do subsítio vizinho P1/P1' indicada. Cada estrutura está representada nas cores *Bwt*, *Bmut*, *Fwt*, *Fmut*. A identificação dos resíduos está feita somente para a protease *Bwt*. Em ambas as representações a cadeia lateral da Val em P2/P2' está voltada para trás, na direção da Val32. Não existe nenhuma mutação nos resíduos compondo esses bolsões nas quatro proteases desse trabalho. Vale ressaltar, no entanto, que as cadeias laterais em P2 tendem a se deslocar ligeiramente na direção da ponta do *flap* que separa os bolsões S1' e S2 nas estruturas contendo a mutação V82A. Esse deslocamento é mais evidenciado na protease *Fmut* mostrada em (a).

arranjos observados pela inclusão da mutação V82A, também foram observadas com um inibidor simétrico cujo modo de ligação é também simétrico com relação aos dois Asp25, com um eixo de simetria não-cristalográfica passando pelo nitrogênio da porção aminodiol central do inibidor (BMS-182193) [Kervinen *et al.*(1996)], o que indica que a assimetria é gerada por um re-empacotamento dos bolsões S1/S1' da enzima, e não pelo rearranjo das cadeias laterais do inibidor.

As diferenças observadas no modo de ligação do anel da Phe em P1' do inibidor TL-3 entre as enzimas selvagem e mutante são maiores que aquelas observadas nesses dois trabalhos citados no parágrafo anterior, e em nenhum desses casos o anel da Phe se deslocou tanto na direção da Ala82 como no caso do TL-3. As diferenças observadas para o inibidor A-77003 são tão pequenas que são dificilmente distinguíveis, segundo os autores do trabalho [Baldwin *et al.*(1995)].

Os resíduos da ponta do *flap* delimitam, em parte, o bolsão S1 (S1') de um lado e o bolsão S2' (S2) do outro (Figura 3.26). Não existe nenhuma mutação nos resíduos que compõem os bolsões S2/S2' em todas as quatro estruturas analisadas. Da mesma maneira que em P1/P1', o resíduo de valina em P2/P2' ocupa grande parte do volume da cavidade, maximizando as interações van der Waals nesse bolsão (Figura 3.28). Parece existir um pequeno deslocamento dos átomos em P2 nas estruturas contendo a mutação V82A na direção da ponta do *flap* que separa os bolsões S1' e S2, principalmente na *Fmut* (Figura 3.28-a). É possível que esse pequeno rearranjo seja consequência do colapso gerado pela mutação V82A em S1'. Em geral as posições dos resíduos em P2/P2' são bastante coerentes entre as quatro estruturas.

O *loop* contendo os resíduos da tríade catalítica Asp25-Thr26-Gly27 delimitam tanto os bolsões S1/S1' quanto os S3/S3' vizinhos. O resíduo em P3/P3' é uma alanina, que ocupa um volume muito menor que o volume da cavidade (Figura 3.29), permitindo assim rearranjos mais pronunciados na posição desses átomos. Em geral as posições dos resíduos em P3' são mais divergentes que em P3. Essa divergência maior em P3' pode ser um reflexo do re-empacotamento do bolsão S1' nas estruturas com a mutação

Figura 3.29: Representação estereográfica dos bolsões S3 e S3' ocupados pelos resíduos de alanina em P3 e P3', respectivamente, com a posição do subsítio vizinho P2/P2' indicada. Cada estrutura esta representada nas cores *Bwt*, *Bmul*, *Fwt*, . A identificação dos resíduos está feita somente para a protease *Bwt*. Note que a cadeia lateral nesse subsítio não ocupa completamente o bolsão, permitindo uma ligeira divergência entre as posições desses átomos entre as quatro estruturas. As maiores divergências ocorrem entre os resíduos em P3', mostrados em (b), provavelmente devido ao aumento de volume do bolsão S3' gerado pelo re-empacotamento de S1' nas estruturas com a mutação V82A.

V82A, que aumentaria o volume do bolsão S3', diminuindo assim as interações do resíduo de alanina nesse bolsão e permitindo uma maior flexibilidade conformacional.

Embora não existam mutações nos resíduos formando o bolsão S2/S2', duas mutações nas vizinhanças desse bolsão, M46I e I54V, conferem resistência cruzada moderada a quase todos os inibidores comerciais (veja Figura 1.9 na Introdução). Nenhum desses dois resíduos parecem ter qualquer influência na conformação dos resíduos do bolsão S2/S2' nas estruturas cristalográficas resolvidas neste trabalho. A cadeia lateral da Phe53 é ligeiramente deslocada devido à repulsão com a cadeia lateral da Ile46 na protease *Fmut*, no entanto esse resíduo de fenilalanina participa somente nas interações com o subsítio P4/P4', já com acesso à região externa da protease. O bolsão S4/S4' é ocupado pela extensão do inibidor pelo grupo protetor Cbz, que se dobra na direção dos *flaps* nas estruturas de protease de HIV-1. O anel do grupo Cbz ocupa orientações bastante diversas em cada uma das proteases e em cada uma das conformações do inibidor, como mostra a Figura 3.30. Essa variabilidade conformacional ocorre pela falta de fechamento do bolsão S4/S4', que faz com que a porção final do inibidor oriente-se para fora da cavidade catalítica da enzima. As últimas pontes de hidrogênio formadas entre o inibidor e a enzima ocorrem entre o oxigênio da carbonila em P4/P4' e o nitrogênio da cadeia principal da Gly48A/B, e entre o oxigênio do grupo éster do Cbz com a água estrutural W4/W'4, orientada pelas pontes de hidrogênio com as cadeias laterais dos resíduos Asp29A/B. Essa extensão do inibidor está de acordo com as necessidades de um inibidor maior para a protease de FIV, e aparentemente não contribui significativamente para a constante de ligação com as proteases de HIV-1 além do grupo éter.

Nesse bolsão existe uma mutação K45R na protease *Bmut*, provavelmente induzida pela terapia antiretroviral, além da M46I na *Fmut*, como mencionado no parágrafo anterior. Nenhum dos inibidores comerciais se estendem até esse bolsão, dessa maneira acreditamos que essas mutações devam ter um efeito compensatório na atividade da protease frente a seu substrato e não interferem diretamente no modo de ligação dos

(a) Subsítio P4

(b) Subsítio P4'

Figura 3.30: Representação estereográfica dos bolsões S4 e S4' ocupados pelo grupo protetor Cbz em P4 e P4', respectivamente, com a posição do subsítio vizinho P3/P3' indicada. Cada estrutura está representada nas cores *Bwt*, *Bmut*, *Fwt*, A identificação dos resíduos está feita somente para a protease *Bwt*. Nesses bolsões, que levam à região exposta ao solvente, ocorrem os maiores rearranjos entre os subsítios das quatro proteases, provavelmente devido à falta de fechamento desses bolsões. As últimas pontes de hidrogênio entre o grupo Cbz e o bolsão são formadas com os oxigênios do grupo éster. As mutações K45R em *Bmut* e M46I em *Fmut* não parecem influenciar diretamente no modo de ligação do inibidor.

inibidores.

3.6.3 Diferenças Entre Subtipos B e F

Nenhum dos pontos de mutação entre os subtipos B e F selvagens ocorrem em aminoácidos diretamente envolvidos na formação dos bolsões de ligação (Figura 3.24-b). A maior diferença estrutural entre elas, no entanto, ocorre na região dos *flaps* (resíduos 33–62), especificamente entre os resíduos 33–42 (Figura 3.24-d). Nessa região, o deslocamento entre os $C\alpha$ chega a ser de 2.85Å para o resíduo 35. Esses resíduos encontram-se em uma região que se pode chamar de dobradiça dos *flaps*, e sofre enorme rearranjo entre as enzimas *apo* e complexada com inibidores [Miller *et al.*(1989b)]. Nessa região encontram-se as mutações E35D, M36I, S37A e R41K, além da mutação R57K na outra dobradiça do *flap*. Na posição 57 não existem diferenças significativas entre as posições atômicas dos átomos envolvidos, provavelmente devido à seu envolvimento numa folha- β . Dentre essas mutações no *flap*, somente a M36I está associada à resistência a inibidores comerciais, enquanto que todas as outras são consideradas polimorfismo entre os dois subtipos. Como já descrito anteriormente, essa é uma mutação secundária decorrente do tratamento com inibidores de protease nos subtipos B (veja Figura 1.9 na Introdução), entretanto, é comum nos subtipos não-B selvagens. Tem sido proposto que a presença dessa mutação em indivíduos não tratados aumentaria o risco de falha terapêutica [Perno *et al.*(2001)]. Até onde pudemos verificar não existe na literatura nenhuma proposta para o papel dessa mutação na resistência à inibidores de protease. Nossa proposta é de que a mutação M36I enrijece esse *loop* dobradiça do *flap* na estrutura das proteases subtipo F, quando comparadas com a subtipo B selvagem. De fato, a porção da dobradiça do *flap* (resíduos 33–42) faz 25 pontes de hidrogênio com o restante dos resíduos da protease *Bwt*, contra 30 pontes de hidrogênio na protease *Fwt*. Isso implica numa maior estabilização desse *loop* na estrutura da protease *Fwt*. Ainda, dessas pontes de hidrogênio possíveis, 10 são realizadas com os resíduos do *loop* 76–83 na *Bwt* contra 13 na *Fwt*, o que implica também numa maior estabilização do

loop-β 76–83, que compões a cavidade catalítica, na estrutura da protease *Fwt*.

Na Figura 3.31 observa-se que o final das cadeias laterais dos resíduos 36 de ambas as estruturas tendem a ocupar a mesma posição, forçando a cadeia principal a se deslocar. Esse posicionamento da cadeia lateral é estabilizado por contatos van der Waals com as cadeias laterais dos resíduos vizinhos Leu33, Leu38 e Lys20. O encurtamento da cadeia lateral na posição 36 da protease *Fwt* causaria o deslocamento da cadeia principal do *loop* contendo esse resíduo na direção do resíduo Lys20 (na parte de baixo da Figura 3.31) e também na direção do *loop* entre os resíduos 76–83, que participam da formação dos bolsões P1/P1'.

Figura 3.31: Representação estereográfica dos resíduos da dobradiça do *flap* das estruturas *Bwt* e *Fwt*. Os resíduos com os maiores deslocamento da posição dos $C\alpha$ estão indicados por linhas mais grossas e o nome de cada um desses resíduos está colocado ao lado deles com a mesma cor de cada uma das estruturas. Observe que o final da cadeia lateral do resíduo na posição 36 ocupa a mesma posição no espaço em ambas as estruturas, fazendo com que o *loop* correspondente à dobradiça do *flap* se desloque para baixo na representação da protease *Fwt*. Esse pequeno deslocamento pode ser responsável pelo enrijecimento das dobradiças.

Os fatores de temperatura também fornecem uma indicação desse enrijecimento da dobradiça 33–42 na protease *Fwt*. Na *Fwt*, a magnitude dos fatores de temperatura são consistentemente maiores que aqueles da protease *Bwt*, tanto na cadeia principal como nas laterais (Figura 3.25), exceto pelo região entre os resíduos 33–45 onde a situação se inverte e esses resíduos apresentam fatores de temperatura menores para a *Fwt*.

Esse enrijecimento das dobradiças dos *flaps* poderia levar à um sítio ativo menos

flexível e dinamicamente menos preparado para receber o inibidor, desde que os *flaps* e o *loop* 76–83 são as duas regiões que sofrem as maiores modificações estruturais pela ligação do inibidor [Miller *et al.* (1989b)], contribuindo assim para a diminuição da atividade de todos os inibidores, como de fato acontece (Figura 1.9). De maneira geral, as dobradiças das estruturas mutantes possuem um enovelamento semelhante à sua contrapartida selvagem, assim espera-se que a protease *Fmut* possua um comportamento semelhante ao observado na *Fwt*.

Todas as outras mutação entre *Bwt* e *Fwt* são consideradas polimorfismos entre os subtipos. Cinco dessas mutações, D60E, Q61N, I62V, L63S e I64L, encontram-se muito distantes do sítio ativo, no final da folha- β *b'*, enquanto que a mutação I15V encontra-se no último resíduo da fita- β *b*, em contato direto com a *b'* (Figura 3.24 e Figura 1.2 na Introdução).

Um fato curioso é que as proteases do subtipo F não desenvolvem a mutação primária L90M, nem mesmo em indivíduos sob tratamento. Em contrapartida, a mutação L89M faz parte da grande maioria dos subtipos não-B isolados [Caride *et al.* (2001)]. De fato, o aparecimento das duas mutações L89M/L90M ao mesmo tempo é muito rara.

Hong *et al.* (2000), através da análise do duplo mutante G48V/L90M com o inibidor saquinavir, sugerem que a cadeia lateral da metionina na posição 90 possui interações van der Waals adicionais com a cadeia principal dos resíduos do sítio ativo (24–26), o que levaria à um decréscimo no volume e na flexibilidade estrutural dos bolsões S1/S1', quando comparado com a proteína selvagem. Embora as diferenças observadas sejam pequenas (da ordem de décimos de angstroms), os autores sugerem que essas pequenas perturbações na região bem estruturada do sítio ativo podem comprometer a complementaridade entre o inibidor e a enzima, reduzindo a potência do inibidor.

O que se observa entre as estruturas *Bwt*:TL-3 e *Fwt*:TL-3 é bastante semelhante ao descrito acima, muito embora nesse caso a mutação esteja na posição 89 e não na 90. Aparentemente, o aumento do volume da cadeia lateral pela inclusão da metionina na posição 89 obriga o resíduo Leu90 vizinho a se deslocar na direção dos átomos da

Figura 3.32: Representação estereográfica das interação entre os átomos C δ 1 e C δ 2 do resíduo Leu90 com os resíduos 24–26 do sítio ativo para as proteases *Bwt* e *Fwt*. As linhas pontilhadas indicam interações tipo van der Waals. Note o maior número de interações van der Waals na protease *Fwt* como consequência do deslocamento do resíduo Leu90 na direção do sítio ativo pela mutação L89M, levando à uma redução no volume e na flexibilidade do sítio ativo.

cadeia principal do sítio catalítico (Figura 3.32). Esse deslocamento faz com que o número de interações van der Waals entre os átomos desses resíduos aumente de cinco para sete, além de deslocar os resíduos catalíticos 24–26 para dentro do bolsão S1/S1', com um deslocamento máximo de 0.3Å da carbonila da cadeia lateral do Asp25.

Se esse raciocínio está correto a mutação L89M seria equivalente à L90M, descrita como uma mutação primária levando à resistência direta ou indireta à cada um dos sete inibidores comercialmente aprovados [Shafer(2004), Johnson *et al.*(2004)]. Até onde pudemos verificar, não existem dados que correlatem a mutação L89M com resistência a inibidores, desde que ela ocorre em subtipos não-B, principalmente o F, e a grande maioria dos estudos se concentram nos subtipos B. Além disso, como a L89M aparece já na protease F selvagem, e a mutação L90M exclui a L89M nos subtipos B, é muito difícil correlacionar essa posição com a resistência a inibidores.

Até onde pudemos verificar, existe uma única ocorrência da mutação L89M relacionada com resistência a inibidores, nesse caso devido ao tratamento com atazanavir antes de sua aprovação pelo FDA [Gong *et al.*(2000), Parikh *et al.*(2002)].

Uma observação mais cuidadosa do subsítio P1 na Figura 3.27-a revela que os anéis da Phe, embora tendam a ocupar posições muito próximas e muito menos discrepantes que aquelas observadas no subsítio P1', apresentam um pequeno deslocamento nas proteases *Bwt/Bmut* com relação às *Fwt/Fmut*. Essa diferença entre os subtipos é consequência de diferenças intrínsecas entre os subtipos F e B, e não da mutação V82A no sítio ativo. Dentre as diferenças existentes entre os subtipos, a mutação M36I é provavelmente a responsável por essa particularidade, desde que nossa proposta é de que essa mutação têm consequências diretas na flexibilidade do *loop* contendo o resíduo Pro81, que está em contato direto com o anel da Phe em P1. L89M aparentemente não interfere nesse caso, desde que a protease *Bmut*, cuja posição do anel concorda com aquela da protease *Bwt*, carrega a mutação L90M.

Essa diferença no posicionamento dos anéis da Phe entre os subsítios também pode ser observada no subsítio P1', embora as grandes mudanças conformacionais geradas

pela mutação V82A tornem a identificação mais difícil.

3.6.4 Modificações no Inibidor

Os dados obtidos com as quatro estruturas deste trabalho, em conjunto com outros previamente publicados, nos levam a propôr algumas modificações na estrutura do inibidor TL-3 que poderiam levar à uma maior potência e especificidade desse inibidor.

Baldwin *et al.* (1995) já haviam proposto que um resíduo mais volumoso em P1 deveria aumentar a potência e especificidade do inibidor com relação à mutação V82A, desde que nesse subsítio não ocorre o rearranjo observado em P1', responsável pela equivalência das interações nesse bolsão⁴. De fato os autores descrevem que a substituição da Phe em P1/P1' por uma Tyr leva à um inibidor equipotente para as enzimas selvagem e mutante V82A.

Aparentemente a otimização das interações em P2/P2' são críticas na determinação da potência dos inibidores de HIV-PR [Wlodawer & Erickson(1993)]. O deslocamento da cadeia lateral na protease *Fmut* em P2 é bastante pronunciado, o que pode indicar que a cadeia lateral nessa posição possa ser aumentada a fim de maximizar a ocupação do volume desse bolsão, muito embora a resolução dos dados da *Fmut* levem a considerações cautelosas nesse sentido. No entanto, quatro dos inibidores atualmente aprovados pelo FDA — saquinavir, nelfinavir, indinavir e atazanavir — possuem um grupo t-butil nessa posição, o que favorece nossa hipótese. Essa metila adicional aumentaria as interações entre o grupo em P2/P2' e as cadeias laterais dos resíduos na ponta dos *flaps*, se não nos dois bolsões pelo menos no bolsão S2 onde a liberdade posicional parecem ser um pouco maior.

As águas estruturais nos bolsões são outra possibilidade a ser explorada no melhoramento do K_i do inibidor. Modificações incluindo átomos que interagissem através de pontes de hidrogênio com essas águas contribuiriam com um ganho entálpico para

⁴Embora a nossa definição dos resíduos que compõem os bolsões seja a mesma utilizada por Baldwin *et al.* (1995), os modos de interação do inibidor nos bolsões P1/P1' está invertido, ou seja, os autores observam os maiores rearranjos em P1 e não em P1' como no nosso caso. Acreditamos que isso se deva unicamente à arbitrariedade na atribuição da nomenclatura das cadeias A e B.

a constante de ligação. Ainda, modificações que viessem a substituir a posição dessa águas por átomos equivalentes do inibidor contribuiriam entropicamente para a constante de ligação através da liberação dessas moléculas de água ordenadas.

Todas as quatro estruturas aqui apresentadas possuem uma água estrutural nos bolsões S3/S3' (W2/W'2), cuja estrutura é estabilizada através de pontes de hidrogênio com os resíduos do *loop* contendo a Gly27 (Figura 3.29). Essas águas são também conservadas nas estruturas do inibidor TL-3 com a protease de FIV [Li *et al.*(2000)], onde os autores propõem que um resíduo de Ser no lugar de Ala em P3/P3' poderia fazer uma ponte de hidrogênio com essa água, aumentando a potência do inibidor. Um resíduo ainda mais longo que uma Ser, contendo uma hidroxila na cadeia lateral, poderia deslocar essa água. Essa análise é válida não somente para a protease selvagem como para cada uma das quatro estruturas apresentadas nesse trabalho. Ainda mais, a análise dessas quatro estruturas mostra que existe uma outra água neste mesmo bolsão (W3/W'3) em pelo menos duas das estruturas (*Bmut* e *Fwt*), e que a inclusão de um resíduo com um grupo ácido (ácido glutâmico, por exemplo) tiraria vantagem dessas duas águas, W2/W'2 e W3/W'3, substituindo-as pelos oxigênios da cadeia lateral. Mesmo nas estruturas onde a água W3/W'3 não está presente, um resíduo ácido se beneficiaria da ponte de hidrogênio com o N ϵ da cadeia lateral da Arg8.

No bolsão S4/S4' existe uma molécula de água (W4/W'4) nas estruturas de PR do subtipo B, interagindo com os resíduos do *loop* contendo o Asp29 (Figura 3.30). Desde que o anel aromático em P4/P4' não interage especificamente com o bolsão S4/S4', a substituição do grupo Cbz por um grupo cuja hidroxila substituísse a água W4 seria vantajosa, de acordo com as propostas que têm sido feitas até agora. Um resíduo de Ser nessa posição serviria a esse propósito, além da possibilidade de realizar pontes de hidrogênio com a cadeia lateral da Arg45, que substitui a Lys45 na protease *Bmut*, o que aumentaria a potência do inibidor frente a esse mutante.

Essas propostas de modificação para cada subsítio estão sumarizadas na Figura 3.33.

R1 = grupo volumoso
R3 = CH₂OH (Ser), (CH₂)₂COOH (Glu)

R2' = t-butil, isopropil (Ala)
R3' = CH₂OH (Ser), (CH₂)₂COOH (Glu)

Figura 3.33: Propostas de modificações no inibidor TL-3 baseadas em dados estruturais. O aumento do volume em P1 aumentaria o número de interações no bolsão S1, mesmo com a inclusão da mutação V82A. Em P2/P2' a inclusão de uma cadeia lateral t-butil também aumentaria a complementaridade com o bolsão S2/S2', enquanto que em P3/P3' e P4/P4' a substituição por resíduos polares tiraria vantagem das interações com moléculas de água presentes nos respectivos bolsões.

Parte II

L-asparaginase

Capítulo 4

Introdução

Amidohidrolase (E.C.3.5.1.1) é a denominação genérica para enzimas que catalisam a reação de desaminação da cadeia lateral da L-asparagina ou L-glutamina, transformando-os em seus respectivos aminoácidos ácidos, e amônia.

Essas enzimas podem ser específicas para L-asparagina, chamadas L-asparaginases, ou podem catalisar tanto L-asparagina como L-glutamina, quando então recebem a denominação de glutaminase-asparaginases.

As L-asparaginases são utilizadas como agentes quimioterápicos contra alguns tipos de câncer, tais como leucemia linfoblástica aguda e linfossarcoma, principalmente em crianças. A enzima age retirando L-asparagina circulante em células malignas, desde que essas são asparagina-dependentes. Embora sejam agentes antitumorais importantes, a utilização das L-asparaginases é normalmente limitada por sua toxicidade [Crowther(1971), Chakrabarti & Scuster(1997)]. Os efeitos colaterais das L-asparaginases é freqüentemente atribuído à atividade L-glutaminase secundária dessas enzimas. Assim, enzimas glutaminase-asparaginase não são úteis como agentes anti-câncer. As L-asparaginases de *E. coli* e *E. carotova* são os principais agentes quimioterápicos utilizados. Duas L-asparaginases, chamadas L-asparaginases I e II, foram isoladas de *E. coli*. A L-asparaginase tipo II localiza-se perto da membrana celular

e é altamente ativa contra tumores asparagina-dependentes em mamíferos, enquanto que a L-asparaginase tipo I é constitutiva e não inibe o crescimento tumoral [Cedar & Schwartz(1967)].

As amidohidrolases bacterianas são enzimas tetraméricas globulares contendo por volta de 330 aminoácidos por subunidade e quatro sítios ativo. Cada subunidade da enzima possui dois domínios conectados por um *loop*, ambos contendo uma folha- β torcida cercada por hélices- α , como mostra a Figura 4.1. No início do domínio N-terminal existe um *loop* flexível exposto ao solvente, provavelmente atuando como uma tampa do sítio ativo. Este *loop* contém dois resíduos importantes para a catálise, uma treonina e uma tirosina. O restante dos resíduos compondo o bolsão catalítico são mais rígidos e localizados no interior da molécula. Duas subunidades são necessárias para formar dois sítios ativos, e a junção dessas subunidades é chamada de dímero íntimo, muito embora a enzima ativa encontre-se na forma tetramérica [Swain *et al.*(1993)].

Figura 4.1: Representação de uma das subunidades da L-asparaginase de *E. coli* resolvida em nosso grupo. O aspartato no sítio ativo está representado com sua densidade eletrônica $2|F_{obs}| - |F_{calc}|$, α_{calc} à 1σ . O domínio C-terminal encontra-se à esquerda e o N-terminal à direita. O *loop* flexível no início do domínio N-terminal está representado em cinza.

Atualmente várias estruturas de amidohidrolases de bactérias foram cristalograficamente resolvidas, incluindo *Erwinia chrysanthemi* [Miller *et al.*(1993), Jaskólski *et al.*(2001)], *Pseudomonas* 7a [Lubkowski *et al.*(1994b), Jakob *et al.*(1997)], *Acinetobacter glutaminasificans* [Lubkowski *et al.*(1994a)], *Wolinella succinogenes* [Lubkowski *et al.*(1996)] e *Escherichia coli* [Swain *et al.*(1993), Sanches *et al.*(2003)].

4.1 Objetivos

A compreensão detalhada do mecanismo de ação das amidohidrolases, em especial das L-asparaginases, permitiria a inclusão de modificações na sequência da proteína selvagem que levariam à uma diminuição dos efeitos colaterais em terapias anti-câncer. Assim, o objetivo deste trabalho é de fornecer informações sobre o mecanismo reacional das amidohidrolases, baseado principalmente em dados cinéticos e estruturais, disponíveis na literatura.

Nossa contribuição nesse sentido vem também na forma de uma estrutura resolvida da L-asparaginase de *E. coli*, num grupo espacial diferente daquele já existente, e a comparação com a previamente publicada, no intuito de explicar a estabilidade e o estado de oligomerização da enzima, além da ocupação dos sítios ativos. Estes resultados estão apresentados na forma de um artigo publicado na *Acta Crystallographica* [Sanches *et al.*(2003)], em anexo nessa tese.

Capítulo 5

Metodologia

5.1 Cristalização, Coleta de Dados e Refinamento

A cristalização, coleta de dados e resolução estrutural da L-asparaginase estão descritas em Polikarpov *et al.* (1999) e no artigo anexado à esta tese [Sanches *et al.*(2003)].

5.2 Análise e Comparação Estruturais

Às estruturas contendo uma ou duas subunidade na unidade assimétrica foi aplicada a simetria cristalográfica apropriada para a geração da unidade biológica tetramérica utilizando-se o programa O [Jones *et al.*(1991)]. As representações das posições atômicas dos modelos foram feitas com o programa PyMOL [DeLano(2002)].

Para a superposição das estruturas analisadas e cálculo de rmsd posicional foram utilizados os programas LSQMAN [Kleywegt(1996)] e Lsqkab [Bailey(1994)], enquanto que o programa ClustalW¹ foi utilizado no alinhamento seqüencial.

Os resíduos envolvidos nas cavidades do sítio ativo, e o volume dessas cavidades, em cada uma das estruturas contendo o *loop* flexível fechado, foram calculados pelo programa VOIDOO [Kleywegt & Jones(1994)]. O volume corresponde à acessibilidade por rolagem de uma sonda (*probe rolling accessibility*) usando sempre a posição do Ca do ligante Asp na estrutura EcA como ponto de partida.

¹<http://www.ebi.ac.uk/clustalw/>

Capítulo 6

Resultados e Discussão

Não temos a intenção de traduzir aqui o texto apresentado no artigo em anexo [Sanches *et al.*(2003)]. No entanto, apresentaremos algumas comparações decorrentes da análise de estruturas disponíveis de amidohidrolases de vários organismos, incluindo a de *E. coli* resolvida neste grupo. Para fins de padronização, sempre que um resíduo for referenciado no texto, a posição desse resíduo será correspondente à L-asparaginase de *E. coli*, a menos que expressamente indicado.

6.1 Comparações das Amidohidrolases

Até o momento, estruturas de amidohidrolases selvagens de cinco organismos distintos foram resolvidas, incluindo as estruturas de *Escherichia coli* [Swain *et al.*(1993), Sanches *et al.*(2003)], *Wolinella succinogenes* [Lubkowski *et al.*(1996)], *Erwinia chrysanthemi* [Miller *et al.*(1993), Jaskólski *et al.*(2001)], *Acinetobacter glutaminasificans* [Lubkowski *et al.*(1994a)] e *Pseudomonas 7A* [Lubkowski *et al.*(1994b), Jakob *et al.*(1997)]. Alguns dados relativos à essas estruturas estão sumarizados na Tabela 6.1. Algumas dessas enzimas têm sido classificadas como asparaginases, outras como glutaminase-asparaginases, de acordo com a especificidade da enzima com relação aos substratos asparagina/glutamina. A enzima de *Erwinia chrysanthemi* têm sido descrita como uma L-asparaginase na literatura [Miller *et al.*(1993)], embora sua atividade contra L-Gln seja significativamente mais alta que aquela observada para outras asparaginases, como EcA e WsA. Esta atividade, no entanto, não é suficiente para considerá-la como uma

asparaginase-glutaminase.

Tabela 6.1: Estruturas resolvidas de amidohidrolases utilizadas nas comparações estruturais. A classificação da enzima como asparaginase ou asparaginase-glutaminase está representada pelo substrato de cada uma delas na primeira coluna da tabela. A tabela também discrimina qual o ligante modelado no sítio ativo de cada uma das estruturas, no caso de haver algum.

Substrato	Organismo		N. de aa	Código PDB	Ligante
Asp	<i>E. coli</i>	(EcA)	326	1NNS/4ECA	Asp/Asp
Asp	<i>W. succinogenes</i>	(WsA)	330	1WSA	—
Asp(/Glu)	<i>E. chrysanthemi</i>	(ErA)	327	1HFJ/1HFK/1HFW	SO ₄ ⁻² /SO ₄ ⁻² /L-Glu
Asp/Glu	<i>A. glutaminasificans</i>	(AGA)	331	1AGX	—
Asp/Glu	<i>Pseudomonas 7A</i>	(PGA)	337	3PGA/4PGA	—/NH ₄ ⁺ , SO ₄ ⁻²

Dentre essas estruturas, mostradas na Tabela 6.1, as de *W. succinogenes* e *E. chrysanthemi* foram refinadas sem o *loop* flexível no domínio N-terminal, comprometendo o seu uso nas análises a seguir.

Todas as amidohidrolases possuem um sítio ativo bem conservado. O alinhamento das seqüências das cinco amidohidrolases utilizadas nessa comparação mostra 26.39% de identidade entre elas (90 resíduos), como pode ser visto na Figura 6.1. Os resíduos de Thr12, Tyr25, Thr89, Asp90, Lys162 e Glu283 são absolutamente conservados entre todas as estruturas.

O mecanismo atualmente mais aceito para a transformação de L-asparagina (ou L-glutamina) em L-aspartato (ou L-glutamato) passa por um intermediário covalentemente ligado ao substrato, envolvendo uma β -aspartil enzima, em oposição ao modelo de um anidrido aspartico como intermediário [Ehrman *et al.*(1971), Röhm & van Etten(1986)]. Isso significa que o ataque nucleofílico ocorre na cadeia lateral do substrato eletrofílico (asparagina ou glutamina) seguido da eliminação de amônia, formando um intermediario enzima-substrato covalentemente ligados.

O bolsão catalítico das amidohidrolases é composto por resíduos de ambas as subunidades formando um dímero íntimo, por exemplo, subunidades A e C na EcA (Figura 6.2). Mutações nos resíduos Thr12, Thr89 e Tyr25 de EcA levaram à um

Figura 6.1: Alinhamento de sequência de aminoácidos das cinco amidohidrolases analisadas. Dos resíduos mostrados, 26.39% são idênticos entre as cinco estruturas (*, vermelho), 19.94% apresentam alta similaridade (:, verde) e 12.61% apresentam baixa similaridade (., azul). As duas tríades catalíticas Thr12-Tyr25-Glu283 e Thr89-Lys162-Asp90 estão destacadas por quadrados em azul e vermelho, respectivamente. Os resíduos equivalentes do N248_{EcA} estão destacados por um quadrado verde.

Figura 6.2: Sítio ativo da L-asparaginase de *E. coli* (EcA). As linhas tracejadas representam pontes de hidrogênio entre os resíduos. Os resíduos marcados com “*” fazem parte de uma subunidade distinta daqueles não marcados, ambas no dímero íntimo. Parte do *loop* flexível contendo os resíduos catalíticos Thr12 e Tyr25 também está representado em forma tubular. Note que a linha formada entre os dois oxigênios das cadeias laterais das treoninas 12 e 89 é quase perpendicular ao plano carbonílico da cadeia lateral do produto L-Asp.

decréscimo considerável na atividade catalítica da enzima [Palm *et al.*(1996), Harms *et al.*(1991), Derst *et al.*(1994)], sugerindo que esses resíduos estão diretamente envolvidos na catálise. Uma das duas treoninas é o nucleófilo mais favorável. De fato, estas duas treoninas estão em lados opostos do bolsão catalítico em todas as estruturas de amidohidrolases cristalizadas com um ligante no sítio ativo, sendo que o plano carbonílico do produto é quase perpendicular à linha formada entre os dois oxigênios das cadeias laterais das treoninas. A mutação T12A suprimiu a atividade catalítica da enzima, embora a afinidade pelo ligante aspartato tenha sido mantida, sugerindo que Thr12 não está envolvida no processo de ligação do substrato [Harms *et al.*(1991)]. De fato, em nosso modelo da EcA existe uma densidade eletrônica contínua entre o produto aspartato e os resíduos Thr89 e Ser58, sugerindo que esses resíduos participam do ancoramento do produto no sítio ativo e, possivelmente, fazem o mesmo com o substrato. Até onde pudemos verificar, não existem, até o momento, experimentos

cinéticos com mutações na Ser58 que comprovem nossa hipótese de que esse resíduo está envolvido no processo de ancoramento do substrato.

Um trabalho bastante elucidativo, publicado por Palm *et al.*(1996), demonstra o papel da Thr89 no processo de catálise. Um mutante T89V foi construído e a estrutura cristalina com o substrato covalentemente ligado à Thr12 foi refinada (Figura 6.3), confirmando cristalograficamente as hipóteses de uma β -aspartil enzima e da ação da Thr12 como nucleófilo no passo de acilação. Os autores apontam ainda a falta de uma molécula de água nas vizinhanças da Thr89 e do substrato, provavelmente devido à baixa polaridade do grupo metil da valina. Esta água agiria na reação de desacilação da enzima, justificando a ineficiência do mutante T89V e a presença do substrato ligado covalentemente à enzima. Embora evidências cinéticas suportem esse comportamento, no passado houve uma relutância em propôr uma treonina como nucleófilo primário de uma enzima. No entanto, as estruturas cristalográficas das amidohidrolases resolvidas mostraram que não existe outro resíduo nas vizinhanças do sítio ativo capaz de atuar como nucleófilo além das duas treoninas. Mutações nessas treoninas mostraram que a Thr12 é provavelmente o nucleófilo primário [Harms *et al.*(1991), Palm *et al.*(1996)]. Atualmente o papel das treoninas como nucleófilos é aceito, e as L-asparaginases foram as primeiras enzimas a serem descritas como treonino proteases [Dodson & Wlodawer(1998)]. Embora o mutante T12A inative a enzima, o T12S restaura parcialmente a atividade (20% contra L-Asn, comparado com a enzima selvagem) [Derst *et al.*(1992)], mostrando que um grupo hidroxila é necessário na cadeia lateral do substrato. No entanto o papel da metila na cadeia lateral do nucleófilo não é claro.

O papel da Thr12 como nucleófilo primário também foi comprovado pelo uso dos inibidores suicidas DON e DONV, análogos diazo da glutamina e da asparagina, que se ligam covalentemente à enzima [Ortlund *et al.*(2000)]. À partir das estruturas cristalográficas dos complexos PGA-DON e PGA-DONV os autores sugerem a existência de uma tríade catalítica Thr-Tyr-Glu, com a tirosina agindo como uma base geral na ativação da treonina catalítica.

Figura 6.3: Alinhamento estrutural da região do sítio ativo contendo o *loop* flexível das amidohidrolases de *E. coli*, *E. coli* T89V, *A. glutaminasificans*, *Pseudomonas* 7A e *Thermoplasma* 1. O *loop* possui três regiões de dobradiças, mostrando que o movimento desse *loop* é dividido em dois durante a exposição do sítio ativo, compreendendo um movimento menor, entre as dobradiças 1 e 3, de uma parte do *loop* contendo o resíduo de treonina catalítico, e um maior, entre as dobradiças 2 e 3, contendo o resíduo de tirosina. Nas estruturas de WsA e ErA o *loop* flexível encontra-se desordenado e, portanto, não foram incluídas nessa análise. Note também que o sulfato e a amônia, refinadas no sítio ativo da estrutura PGA-SO₄²⁻ com o *loop* fechado, ocupam posições equivalentes dos grupos carboxilato e amina, respectivamente, no substrato L-Asp da estrutura EcA. Também está representada a distância entre os átomos OH da tirosina e Oγ1 da treonina nas estruturas PGA com o *loop* aberto e fechado, para comparação.

Foi demonstrado por RMN, usando a Tyr25 como sonda [Derst *et al.*(1994)], e por dados cristalográficos [Lubkowski *et al.*(1994a), Jakob *et al.*(1997)] que um dos *loops* que formam o bolsão catalítico e, conseqüentemente, contém alguns dos resíduos importantes para a catálise, particularmente Thr12 e Tyr25, é bastante flexível (Figura 6.3). Este *loop* isola o sítio ativo do solvente nas estruturas onde este encontra-se fechado. Das nove estruturas analisadas, somente uma foi refinada na conformação aberta (PGA [Lubkowski *et al.*(1994a)]), com 3 na conformação fechada (EcA, PGA-SO₄⁻², AGA) e o restante sem densidade para o *loop*. A mobilidade deste *loop* sugere que ele atua como uma tampa do sítio ativo. Ao analisar o rearranjo deste *loop* observa-se que a sua abertura ocorre através do movimento em torno de três pontos de dobradiças, e não duas, como mostra a Figura 6.3. Essas dobradiças são mantidas por pontes de hidrogênio entre as cadeias principais nessas regiões e cadeias laterais de resíduos em regiões de estrutura secundário bem definida. As dobradiças 1 e 3 definem um *loop* menor e mais rígido, contendo o resíduo catalítico Thr12, enquanto que a porção maior do *loop* flexível encontra-se definida entre as dobradiças 2 e 3.

Mutações no resíduo Tyr25 incluído neste *loop*, embora reduzam drasticamente a atividade da enzima, mostraram que este resíduo não é crucial para a catálise [Derst *et al.*(1994)], dado que a reação ainda ocorre na enzima mutada e que este resíduo não está em contato direto com o substrato (baseado na conformação das enzimas com o produto). Tentativas foram feitas no sentido de se analisar cristalograficamente o mutante Y25F, mas essas tentativas foram, infelizmente, frustradas, desde que os autores não encontraram densidade eletrônica para o *loop* em questão [Jaskólski *et al.*(2001)]. Esses dados indicam, no entanto, que a Tyr25 deve ter um papel importante no fechamento do *loop*.

Resíduos em condições de realizar pontes de hidrogênio com o substrato incluem Thr12, Ser58, Gln59, Thr89, Asp90 e Ala114 numa subunidade e Asn248 e Glu283 na outra, como mostra a Figura 6.2. Dentre esses resíduos, Gln59 e N248 não são conservados entre todas as seqüências, como mostra a Figura 6.1. Note-se que variações

nessas posições ocorrem somente naquelas enzimas apresentando atividade glutaminase significativa e deve estar de alguma maneira relacionada à esta inespecificidade.

Baseado nesses dados apresentados acima, propusemos o mecanismo esquematicamente representado na Figura 6.4, com a Thr12 como nucleófilo primário. Nesse mecanismo os passos *a* e *b* (1–3) ilustram a etapa de acilação, enquanto que os passos *c* e *d* (3–5) representam a etapa de desacilação. A acilação pode ser descrita como o ataque nucleofílico do oxigênio do nucleófilo ao carbono carbonílico do substrato, formando um intermediário acil tetraédrico (2). A desacilação, por sua vez, representa o segundo ataque nucleofílico por uma molécula de água, que regenera a enzima e libera o produto.

No passo *a*, a nucleofelicidade da Thr12 é aumentada pela formação de uma tríade catalítica entre os resíduos Thr12, Tyr25 e Glu283 (Figuras 6.2 e 6.5-a). Uma ponte de hidrogênio é formada entre Thr12 e Tyr25, esta última atuando como uma base transitória, abstraindo o próton da Thr12. Esta tirosina é o único resíduo da enzima em condição de realizar uma ponte de hidrogênio com a hidroxila da Thr12. O resíduo de Glu283, na outra subunidade do dímero íntimo, completa a tríade catalítica abstraindo o próton da Tyr25. Na estrutura T89V_{EcA} a Tyr25 apresenta-se igualmente como o único resíduo capaz de fazer uma ponte de hidrogênio com o oxigênio nucleofílico da Thr12, colaborado com essa hipótese. Acreditamos ainda, que o resíduo de tirosina ajude a orientar a cadeia lateral da Thr12 de maneira apropriada à reação de adição. Ao mesmo tempo, a coordenação entre Thr89 e o substrato asparagina faz da cadeia lateral do substrato um eletrófilo mais potente (1) e também estabiliza o oxiânion formado no intermediário tetraédrico (2). No passo *b* a β -aspartil enzima é formada pela liberação de uma molécula de amônia e regeneração da carbonila (3). É provável que a Tyr25 ajude na liberação da amônia formada, intermediada pela constelação de moléculas de água em torno da hidroxila da Tyr25, presente em todas as estruturas analisadas onde o *loop* flexível encontra-se fechado.

A etapa de desacilação começa com o ataque nucleofílico de uma molécula de água

Figura 6.4: Representação esquemática do mecanismo de reação da transformação de asparagina em aspartato (em azul) pelas L-asparaginases, mostrando Thr12 (em vermelho) como nucleófilo primário. De acordo com a norma adotada neste trabalho, a numeração e seqüência indicadas referem-se àquelas da EcA. As linhas pontilhadas representam pontes de hidrogênio, enquanto que as flechas retas indicam ataques nucleofílicos e as flechas curvas indicam movimentos eletrônicos. Os passos *a* e *b* compreendem a etapa de acilação, com o ataque nucleofílico da Thr12 ao substrato L-Asn, gerando o intermediário β -aspartil enzima. Os passos *c* e *d* subsequentes representam a etapa de desacilação, com o ataque de uma molécula de água catalítica, orientada pela Thr89, ao intermediário β -aspartil enzima, regenerando a enzima e liberando o produto L-Asp.

catalítica ao carbono carbonílico da β -aspartil enzima (3) gerando o intermediário tetraédrico oxiânion (4), que lembra o intermediário (2). Essa molécula de água é orientada pela Thr89, cuja nucleofelicidade é aumentada pela formação da tríade catalítica Thr89, Lys162 e Asp90 (Figuras 6.2 e 6.5-b). A regeneração da ligação dupla da carbonila elimina o aspartato recém formado, regenerando a enzima (5). Acreditamos que durante os dois ataques nucleofílicos os intermediários oxiânion 2 e 4 sejam estabilizados pela Thr89. O mutante T89V_{ECA} mostra que o ligante covalentemente ligado à enzima move-se em direção oposta à Lys162 e, portanto, esse resíduo não deve participar diretamente da formação do bolsão do oxiânion.

Figura 6.5: Tríades catalíticas envolvidas na reação das amidohidrolases. Representações utilizando as coordenadas 1NNS da L-asparaginase de *E. coli* resolvida em nosso grupo. (a) tríade proposta por Ortlund *et al.* (2000), com a Thr12 como nucleófilo primário. O resíduo Glu283, na subunidade vizinha, polariza a Tyr25, que age como uma base transitória, facilitando a desprotonação da Thr12 e sua adição ao substrato. Nossa proposta é de que a formação dessa tríade seja importante na etapa de acilação; (b) a tríade proposta por Dodson & Wlodawer (1998), com a Thr89 como nucleófilo. Inicialmente foi proposta como a tríade catalítica para a reação de acilação. Nossa proposta é de que a formação dessa tríade é importante na orientação e ativação da água catalítica na reação de desacilação.

6.1.1 Interações com o Substrato

O reconhecimento do substrato e seu ancoramento no sítio ativo, assim como a liberação dos produtos formados, são pontos importantes a serem considerados nos mecanismos de reação das amidohidrolases. O bolsão de ligação é composto pelos resíduos Ser58, Gln59, Thr89 e Asp90 em uma subunidade e Asn248 e Glu283 na outra subunidade (veja a equivalência para as outras amidohidrolases na Figura 6.1), de acordo com o padrão de pontes de hidrogênio entre o substrato e a cavidade catalítica (Figura 6.2).

Ao analisar a estrutura PGA com um íon sulfato e um amônio ligados ao sítio catalítico, observa-se que esses dois íons estão espacialmente alocados em posições equivalentes do grupo carboxílico e amina da cadeia principal do aspartato no sítio ativo da estrutura EcA (Figura 6.3). Essas duas moléculas pequenas possuem um padrão de pontes de hidrogênio com os mesmos resíduos que acabamos de propor para a composição do bolsão de ligação da EcA, reforçando essa atribuição. Desde que a estrutura PGA foi refinada com o *loop* flexível na conformação fechada, isso sugere que as interações entre os átomos da cadeia principal do ligante e os átomos dos resíduos compondo o bolsão de ligação são responsáveis pelo fechamento desse *loop*. Nós acreditamos que a Thr12 e a Tyr25 nesse *loop* desempenhem um papel importante para esse fechamento, *via* interações do tipo (T,Y)–(Substrato)–(Bolsão de Ligação). Comparando as distâncias entre a treonina e a tirosina nesse *loop* nas conformações fechadas e abertas da estrutura PGA, fica claro que esses dois resíduos, importantes no processo de catalise, são colocados em contato pelo fechamento da cavidade catalítica, como pode ser visto na Figura 6.3. É provável que esses resíduos estejam envolvidos no reconhecimento do substrato com o *loop* na conformação aberta, e que a ligação do substrato no bolsão catalítico medie o fechamento do *loop* *via* interações que acabamos de propor.

6.1.2 Especificidade das Amidohidrolases

Como já foi colocado anteriormente, algumas amidohidrolases são específicas para asparagina, sendo chamadas asparaginases, enquanto que outras são igualmente ativas contra asparagina ou glutamina, sendo designadas glutaminase-asparaginases. Embora a enzima de *E. chrysanthemi* seja considerada uma asparaginase, sua atividade contra glutamina não é desprezível e ela deve ser considerada como uma enzima de transição.

O cálculo do volume da cavidade catalítica fornecesse alguns dados interessantes sobre a especificidade das amidohidrolases. A estrutura EcA possui o menor volume ($1.954 \times 10^2 \text{ \AA}^3$), seguida pela AGA ($4.691 \times 10^2 \text{ \AA}^3$) e PGA ($8.026 \times 10^2 \text{ \AA}^3$). Lubkowski *et al.*(1994) propõem que a AGA deveria possuir um volume de cavidade catalítica menor que a EcA, dado que o resíduo Tyr26_{AGA} encontra-se mais enterrado no sítio ativo, como mostra a Figura 6.3. Na verdade, embora o resíduo Tyr26_{AGA} bloqueie a abertura do sítio ativo, a cavidade nessa estrutura estende-se mais profundamente na direção da interface entre as subunidades que a cavidade na estrutura EcA. Na PGA com o *loop* fechado essa cavidade é ainda mais ampla, desde que ela se estende na direção da interface e na direção do *loop* flexível. Infelizmente não foi possível calcular o volume das cavidades das estruturas ErA e WsA, desde que o *loop* flexível não pode ser modelado nessas enzimas.

No entanto, os dados obtidos com EcA, AGA e PGA indicam que proteínas com atividade glutaminase-asparaginase (AGA e PGA) devem possuir uma cavidade catalítica mais ampla para acomodar o resíduo de glutamina. Dos resíduos que compõem o bolsão de ligação o Asn248 parece ter um papel fundamental no reconhecimento do substrato, desde que as mutações N248A/G/Q/E decrescem a atividade da EcA contra asparagina e glutamina [Derst *et al.*(2000)]. As duas asparaginases analisadas (EcA e WsA) possuem um resíduo de asparagina nesta posição, enquanto que nas glutaminase-asparaginases esta posição é ocupada por um resíduo de serina. Essa substituição está de acordo com nossa proposta de uma cavidade mais ampla nas glutaminase-asparaginases, desde que a serina é um resíduo menor que asparagina e

encontra-se mais afastada do centro do sítio ativo que a asparagina nas asparaginases.

Parte III

Conclusões Gerais

Capítulo 7

Conclusões e Perspectivas

Esta é a primeira vez que proteases de HIV são produzidas, cristalizadas e estruturalmente estudadas no Brasil. A produção das quatro proteases foi feita de maneira reprodutiva e com pureza adequada aos experimentos de cristalização e também para experimentos bioquímicos e biofísicos que venham a ser realizados neste grupo. A inclusão da mutação Q7K e a escolha do vetor de expressão (pET11a) e cepas de *E. coli* (BL21(DE3)-RIL) adequados foram essenciais na obtenção de níveis de expressão alcançados. Essa metodologia nos permitiu obter ≈ 2 mg de proteína purificada, reovelada e concentrada por litro de cultura de células. Além disso a não utilização de plasmídeos pLysS, a adição de glucose para evitar vazamento do promotor e a troca do meio de cultura antes da adição do pré-inóculo foram importantes para o sucesso do procedimento.

Também foram bem sucedidas as cristalizações das proteases brasileiras com o inibidor TL-3. Esses ensaios nos levaram a desenvolver um ensaio preliminar de cristalização de proteases de HIV, que deve ser utilizado em ensaios de cristalização futuros em nosso grupo. Com esses ensaios obtivemos indícios da possibilidade de se cristalizar essas proteases com inibidores comerciais, dado o cristal obtido com o complexo *Bmut*:amprenavir e outros microcristais de *Bmut*:nelfinavir. De maneira geral o ensaio preliminar com os inibidores comerciais foi promissor. É interessante notar que nenhuma das mutações da protease *Bmut* confere alto grau de resistência ao amprenavir (veja Figura 1.9) o que provavelmente viabiliza a sua cristalização com essa protease.

A protease *Fwt* não possui nenhuma mutação primária para qualquer dos inibidores e acreditamos que sua cristalização seja uma questão de tempo. Nenhum ensaio foi realizado com a protease *Fwt* e os inibidores comerciais. Esses estudos de proteases de subtipos não-B com inibidores comerciais farão parte de projetos futuros deste grupo de pesquisa. Atualmente está sendo iniciado um projeto de mestrado com a protease do subtipo C.

A viabilidade dos procedimentos de *annealing* foi observada com todas as proteases coletadas e esse procedimento deve ser utilizado em coletas futuras. Todos os dados foram inicialmente processados no grupo espacial $P6_122$, baseado principalmente nos valores de R_{sym} , nas extinções sistemáticas e nos resultados de substituição molecular. No entanto, o distanciamento exagerado entre os valores de R_{factor} e R_{free} e, mais do que isso, o aumento inaceitável do valor de R_{free} durante a colocação das moléculas de solvente no refinamento, nos mostrou a possibilidade de erro no grupo espacial, que poderia estar sendo gerado por uma simetria não cristalográfica muito próxima da cristalográfica ou dados *twinning*. O abaixamento da simetria para $P6_1$ nos permitiu refinar as estruturas até o final e verificar que, de fato, as duas subunidades da protease são relacionadas por uma simetria não cristalográfica alinhada à possível simetria cristalográfica do grupo $P6_122$. Ainda, a estrutura da protease *Bwt*:TL-3, antes resolvida em $P6_122$ pelo grupo do Dr Alexander Wlodawer, foi reprocessada e re-refinada em $P6_1$, corrigindo alguns erros na estrutura, como a dupla conformação do ligante e a orientação da ligação peptídica entre P3/P3'-P4/P4'.

Os dados cristalográficos aqui apresentados mostraram que, embora o inibidor TL-3 seja simétrico, seu modo de ligação na cavidade catalítica da protease é assimétrico. Esse não é um comportamento novo, tendo sido observado em inúmeros exemplos de inibidores simétricos, inclusive dióis (veja, por exemplo, [Baldwin *et al.*(1995)]). Isso mostra também que o refinamento previamente publicado do complexo *Bwt*:TL-3 com meio ligante construído é essencialmente equivocado e teria que ser refeito, independentemente do grupo espacial.

A comparação das quatro estruturas resolvidas mostrou que as diferenças entre elas é, como era de se esperar, muito sutil, dado que essa é uma enzima pequena e mudanças radicais no seu enovelamento tenderiam a comprometer a atividade enzimática. De todas as mutações e polimorfismos entre elas, a V82A no sítio ativo é, certamente, a que leva à um maior rearranjo. Essa mutação leva à um re-empacotamento do bolsão S1 que, por consequência, obriga à um rearranjo na cadeia lateral do resíduo de fenilalanina em P1. Esse rearranjo no inibidor é muito mais pronunciado nas estruturas aqui apresentadas, *Bmut* e *Fmut*, que naquelas estruturas previamente publicadas com a mutação pontual V82A. Acreditamos que essa mudança conformacional considerável se deva à múltiplas mutações nesses mutantes brasileiros, e não à uma particularidade do inibidor.

Em geral os resíduos em P1'-P4' possuem posições mais divergentes entre as quatro estruturas que aqueles em P1-P4. É possível que essa divergência se dê em decorrência do modo de ligação assimétrico do ligante, que é deslocado na direção de P1', muito embora essa seja uma hipótese difícil de confirmar.

Nossas análises estruturais levam a crer que algumas substituições polimórficas entre os subtipos B e F podem levar à uma maior estabilização de regiões naturalmente flexíveis e importantes para a ligação do inibidor/substrato no subtipo F, originando uma enzima intrinsecamente menos ativa, mas ao mesmo tempo naturalmente resistente à alguns inibidores. Essas observações estão de acordo com a hipótese de Perno *et al.*(2001) de que as proteases não-B, carregando substituições acessório polimórficas, são mais suscetíveis a desenvolver resistência a inibidores comerciais, levando os pacientes infectados por esses subtipos à falha terapêutica mais rapidamente que aqueles com o subtipo B. Essas, diferenças, que poderiam levar à uma enzima menos ativa, podem vir acompanhadas de modificações também nos substratos naturais da protease desse subtipo, aumentando a complementaridade enzima-substrato e regenerando a atividade da enzima.

Duas mutações na protease *Fwt* com relação à *Bwt*, M36I e L89M, estão sendo

diretamente relacionadas à esse tipo de comportamento nesse trabalho. A mutação M36I têm sido classificada como uma mutação acessório (secundária) para inibidores comerciais, incluindo ritonavir e nelfinavir, que são os dois inibidores ministrados aos pacientes das quais as amostras de cDNA viral deste trabalho foram retiradas. Essa substituição conferiria uma maior rigidez às dobradiças dos *flaps* e ao *loop* contendo o resíduo Pro81 que forma o bolsão S1/S1', regiões envolvidas em grandes mudanças conformacionais no processo de ligação do inibidor/substrato. Essa modificação também aparece como consequência de tratamento anti-retroviral em mutantes do subtipo B.

L89M, uma mutação polimórfica encontrada em quase que a totalidade das proteases subtipo F, pode ser um equivalente estrutural da mutação L90M, uma mutação primária para alguns inibidores, e que leva à resistência cruzada a todos os inibidores comercialmente disponíveis. Esses dados ainda devem ser confirmados por ensaios enzimáticos ou ensaios celulares.

É interessante que, até a conclusão desta tese, não nos foi possível encontrar nenhuma publicação de uma estrutura de uma protease de HIV com o inibidor atazanavir. Esse inibidor, o último a ser aprovado pelo FDA, parece selecionar a mutação L89M ao invés da L90M em proteases de algumas cepas do subtipo B [Gong *et al.*(2000)]. Dessa maneira, o atazanavir parece ser um bom inibidor para se estudar o papel da substituição L89M nas proteases do subtipo F comparada com as subtipo B, tanto em ensaios enzimáticos como cristalográficos.

Essas constatações sobre as diferenças entre os subtipos F e B inspiram cuidado na escolha dos inibidores utilizados no tratamento de pacientes infectados com subtipos não-B, devendo ser priorizados aqueles menos suscetíveis às substituições polimórficas desses subtipos, e cujo padrão de mutações de resistência gerado por tratamento em subtipos B não coincidam com essas modificações. Quando esses requerimentos não puderem ser alcançados por um único inibidor, uma combinação de inibidores deveria ser utilizada.

Em geral, dos dados obtidos das várias estruturas de proteases de HIV-1 com ini-

bidores simétricos, incluindo as quatro aqui apresentadas, pode-se concluir que seria mais vantajoso investir na concepção de inibidores assimétricos, que satisfaçam as necessidades específicas de cada um dos bolsões, levando-se conta, inclusive, as possíveis mutações de resistência no sítio ativo, como a V82A aqui mostrada. Baseado nesses preceitos, nós propusemos algumas modificações no inibidor TL-3 que poderiam levar à um aumento de sua potência, tanto frente à proteases selvagens como aquelas contendo a mutação V82A. Essas modificações levariam à uma maior complementaridade no bolsão P1, independentemente da inclusão da mutação V82A. Além disso, algumas modificações foram propostas baseadas na substituição de moléculas de água presentes nos bolsões por átomos doadores de ponte de hidrogênio do inibidor, que levaria à uma contribuição entrópica ao $\Delta\Delta G$ de ligação, metodologia utilizada com sucesso no desenvolvimento dos inibidores contendo um motivo uréia cíclica [Lam *et al.*(1994), Kempf(1994)].

Gostaríamos de ressaltar que neste trabalho estão apresentados os primeiros dados estruturais sobre proteases de HIV-1 de um subtipo não-B. O desenvolvimento de inibidores de protease com ação ampla requer um conhecimento estrutural detalhado de proteases mutantes e de subtipos não-B. A obtenção dessas informações, no entanto, é complicada pelo fato de mais de 20% dos resíduos da protease estarem associados com resistência à vários inibidores [Shafer(2004), Parikh *et al.*(2002)], e de que muitos padrões de resistência consistem de mutações multiplas nos resíduos que compõem a cavidade catalítica, e daqueles que a rodeiam. A deconvolução das relações estrutural-atividade em mutantes resistentes e de subtipos não-B é complicada pela presença de substituições multiplas, tanto no sítio ativo como compensatórias. No entanto, acreditamos que as estruturas apresentadas neste trabalho, e análise das diferenças entre elas, representem o passo inicial na direção do desenvolvimento de inibidores de amplo espectro que levem em conta as diferenças entre os diversos subtipos de HIV-1.

Com relação ao projeto sobre amidohidrolases, acreditamos que as análises apresentadas no artigo em anexo forneceram explicações importantes sobre o estado de

oligomerização e estabilidade da L-asparaginase de *E. coli*. Não temos motivos para crer que essas conclusões não sejam extensivas às outras amidohidrolases bacterianas.

Baseados em dados publicados principalmente para a L-asparaginase de *E. coli*, propusemos um mecanismo de reação geral das amidohidrolases. Todos os resíduos envolvidos nas etapas reacionais são conservados em todas as amidohidrolases bacterianas analisadas, o que favorece nossa generalização. No entanto, as diferenças entre as enzimas dessa classe, que possibilita com que algumas delas utilizem L-glutamina como substrato, deve residir principalmente nas substituições de alguns resíduos que compõem o bolsão de ligação, aumentando o volume dessa cavidade. Dentre essas substituições, a N248S parece ter um papel importante nessa especificidade, segundo demonstram dados cinéticos [Derst *et al.*(2000)] e nossas análises estruturais. Dessa maneira, acreditamos que mutações em alguns resíduos que formam o bolsão de ligação que venham a diminuir o volume da cavidade, ao mesmo tempo em que aumentam a complementaridade desse bolsão com o resíduo de L-asparagina, deveriam aumentar a especificidade das L-asparaginases, diminuindo o efeito colateral no tratamento com essas enzimas devido à atividade glutaminase.

Um experimento interessante, que poderia levar à comprovação dessa hipótese, seria a inclusão da mutação S252N_{AGA} na enzima de *A. glutaminasificans* e S254N_{ErA} na de *E. chrysanthemi* (posições equivalentes da posição 248 na EcA). Na AGA, o *loop* flexível possui um inserção que faz com que a tirosina catalítica se enterre mais profundamente na cavidade catalítica e diminua o volume na entrada do bolsão (Figura 6.3). Essa diminuição de volume deveria levar à uma enzima mais específica para L-asparagina, no entanto, a substituição de um resíduo de asparagina por serina no bolsão de ligação (S252_{AGA}) aumenta o volume da cavidade e, como consequência, essa enzima é classificada como uma asparaginase-glutaminase. Se nossa hipótese está correta, a inclusão da mutação S252N_{AGA} levaria à uma diminuição do volume da cavidade, o que poderia diminuir a atividade glutaminase dessa enzima à níveis menores que aqueles observados para as outras asparaginases. Já a ErA pode ser considerada

uma enzima de transição, com atividade glutaminase mais alta que o esperado, provavelmente por causa do resíduo de serina no bolsão de ligação do substrato, e a mutação S254N_{ErA} poderia, também nesse caso, aumentar a especificidade dessa enzima para L-asparagina.

É de se esperar, numa enzima cujo nucleófilo primário seja uma treonina, que ocorra a formação de uma tríade catalítica similar àquelas observadas nas serino proteases, dada a semelhança entre os resíduos de serina e treonina. Mutações em cada uma das três histidinas da L-asparaginase de *E. coli* mostrou que nenhuma delas é necessária na catálise [Wehner *et al.*(1992)]. A análise das estruturas das amidohidrolases bacterianas mostrou que nenhum desses resíduos de histidina estão nas vizinhanças do sítio ativo, como ocorre nas serino proteases. Por algum tempo, especulou-se que a tríade catalítica da enzima fosse composta por Thr89-Lys162-Asp90 [Dodson & Wlodawer(1998)], dado que não existe nenhum resíduo básico clássico nas vizinhanças do resíduo Thr12. No entanto, estudos de mutagênese posteriores mostraram que a mutação T12A levou à perda de atividade da enzima, e o mutante T89V, resolvido cristalograficamente, mostrou inequivocamente uma β -aspartil enzima covalentemente ligada à treonina 12, o que coloca a Thr12 como o nucleófilo primário mais provável. Ortlund *et al.*(2000) propuseram a existência da tríade Thr12-Tyr25-Glu283 através da análise das estruturas PGA-DON e PGA-DONV.

Esses dados nos levaram a propôr a existência de duas tríades catalíticas diferentes nessas enzimas, atuando em etapas distintas do processo reacional. A primeira delas, Thr12-Tyr25-Glu283 (Figura 6.5-a), atuaria na etapa de acilação, utilizando uma tirosina como base transitória. Isso explicaria a ineficiência dos mutantes substituídos na Thr12, desde que ela atua no primeiro passo da reação. A segunda tríade, Thr89-Lys162-Asp90 (Figura 6.5-b), agiria na etapa de desacilação subsequente, ativando e orientando uma molécula de água catalítica, que regeneraria a enzima e liberaria o produto. Essa atribuição explicaria porque o mutante T89V apresenta o substrato ainda ligado à enzima. Embora esse mecanismo esteja em concordância com os dados

cinéticos e estruturais disponíveis, experimentos adicionais devem ser realizados para a comprovação dessas propostas em outras enzimas dessa família.

Informações Complementares

Informações adicionais, assim como dados atualizados, sobre HIV/AIDS podem ser encontrados nos endereços da web listados a seguir:

1. HIV Protease Database, arquivo até 10/23/2004 (<http://mc11.ncifcrf.gov/hivdb/>)
2. HIV Structural Database (<http://xpdb.nist.gov/hivsdb/hivsdb.html>)
3. HIV Drug Resistance Database - Stanford (<http://hivdb.stanford.edu/>)
4. Los Alamos HIV Sequence Database (<http://hiv-web.lanl.gov>)
5. UNAIDS - Joint United Nations Programme on HIV/AIDS (<http://www.unaids.org>)
6. HIV InSite (<http://hivinsite.ucsf.edu/InSite>)
7. Informações sobre distribuição geográfica dos subtipos:
 "HIV Database, Los Alamos National Laboratory"
 (http://www.hiv.lanl.gov/content/hiv-db/geography/map_right.html)

Bibliografia

[wlo(1991)] (1991). *Inhibitor complexes of the HIV protease: a target for drug design*.
The Robert A. Welch Foundation Conference on Chemical Research.

[una(2003)] (2003). AIDS epidemic update. Publicação conjunta UNAIDS/WHO
(<http://www.unaids.org>).

[una(2004a)] (2004a). 2004 report on the global HIV/AIDS epidemic: 4th global report.
UNAIDS (<http://www.unaids.org>).

[una(2004b)] (2004b). UNAIDS at country level: progress report. UNAIDS Country
and Regional Support Department (<http://www.unaids.org>).

[Abdel-Meguid(1993)] Abdel-Meguid, S. S. (1993). Inhibitors of Aspartyl Proteinases.
Medicinal Research Reviews, **13**(6), 731–778.

[Abergel(2004)] Abergel, C. (2004). Biological crystallography spectacular improvement of X-ray diffraction through fast desiccation of protein crystals. *Acta Crystallographica Section D*, **60**, 1413–1416.

[Bailey(1994)] Bailey, S. (1994). The CCP4 suite—Programs for protein crystallography. *Acta Crystallographica Section D*, **50**, 760–763.

[Baldwin *et al.*(1995)] Baldwin, E. T., Bhat, T. N., Liu, B., Pattabiraman, N., & Erickson, J. W. (1995). Structural basis of drug resistance for the V82A mutant of HIV-1 proteinase. *Nature Structural Biology*, **2**(3), 244–249.

- [Bergfors(2003)] Bergfors, T. (2003). Seeds to Crystals. *Journal of Sctructural Biology*, **142**, 66–76.
- [Bradford(1976)] Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-due binding. *Analytical Biochemistry*, **72**, 248–254.
- [Brindeiro *et al.*(2002)] Brindeiro, P. A., Brindeiro, R. M., Mortensen, C., Hertogs, K., Vroey, V. D., Rubini, N. P. M., Sion, F. S., Sá, C. A. M., Machado, D. M., Succi, R. C. M., & Tarnuri, A. (2002). Testing genotypic and phenotypic resistance in Human Immunodeficiency Virus type 1 isolates of clade B and oher clades from children failing antiretroviral therapy. *Journal of Clinical Microbiology*, **40**(12), 4512–4519.
- [Brünger *et al.*(1998)] Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-X., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., & Warren, G. L. (1998). Crystallography & NMR System: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D*, **54**, 905–921.
- [Brünger(1993)] Brünger, A. (1993). Assessment of phase accuracy by cross validation: the free R value. *Acta Crystallographica Section D*, **49**, 24–36.
- [Caride *et al.*(2001)] Caride, E., Hertogs, K., Larder, B., Dehertogh, P., Brindeiro, R., Machado, E., de Sá, C. A. M., Eyer-Silva, W. A., Sion, F. S., Passioni, L. F. C., Menezes, J. A., Calazans, A. R., & Tanuri, A. (2001). Genotyping and phenotyping analysis of B and non-B Human Immunodeficiency Virus type 1 subtypes from patients under HAART. *Virus Genes*, **23**(2), 193–202.
- [Cedar & Schwartz(1967)] Cedar, H. & Schwartz, J. H. (1967). Localization of 2 L-asparaginases in anaerobically grown *Escherichia coli*. *Journal of Biological Chemistry*, **242**(16), 3753–3755.

- [Chakrabarti & Scuster(1997)] Chakrabarti, R. & Scuster, S. M. (1997). L-asparaginase—Perspectives on the mechanisms of action and resistance. *International Journal of Pediatric Hematology/Oncology*, **4**(6), 597–611.
- [Chen *et al.*(1999)] Chen, Z., Blanc, E., & Chapman, M. S. (1999). Improved free R factors for cross-validation of macromolecular structure — importance for real-space refinement. *Acta Crystallographica Section D*, **55**, 219–224.
- [Crowther(1971)] Crowther, D. (1971). L-asparaginase and human malignant disease. *Nature*, **229**, 168–171.
- [Davis *et al.*(1996)] Davis, D. A., Dorsey, K., Wingfield, P. T., Stahl, S. J., Kaufman, J., Fales, H. M., & Levine, R. L. (1996). Regulation of HIV-1 protease activity through cysteine modification. *Biochemistry*, **35**, 2482–2488.
- [Davis *et al.*(1999)] Davis, D. A., Yusa, K., Gillim, L. A., Newcomb, F. M., Mitsuya, H., & Yarchoan, R. (1999). Conserved cysteines of the Human Immunodeficiency Virus type 1 protease are involved in regulation of polyprotein processing and viral maturation of immature virions. *Journal of Virology*, **73**(2), 1156–1164.
- [Debouck *et al.*(1987)] Debouck, C., Gorniak, J. G., Strickler, J. E., Meek, T. D., Mercalf, B. W., & Rosenberg, M. (1987). Human immunodeficiency virus protease expressed in *Escherichia coli* exhibits autoprocessing and specific maturation of the Gag precursor. *PNAS*, **84**(24), 8903–8906.
- [DeLano(2002)] DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*. DeLano Scientific LLC, San Carlos, CA, USA (<http://www.pymol.org>).
- [Derst *et al.*(1992)] Derst, C., Henseling, J., & Röhm, K. H. (1992). Probing the role of threonine and serine residues of *E. coli* asparaginase II by site-specific mutagenesis. *Protein Engineering*, **5**(8), 785–789.

- [Derst *et al.*(1994)] Derst, C., Wehner, A., Specht, V., & Röhm, K. H. (1994). State and functions of tyrosine residues in *Escherichia coli* asparaginase II. *European Journal of Biochemistry*, **224**, 533–540.
- [Derst *et al.*(2000)] Derst, C., Henseling, J., & Röhm, K.-H. (2000). Engineering the substrate specificity of *Escherichia coli* asparaginase II. Selective reduction of glutaminase activity by amino acid replacements at position 248. *Protein Science*, **9**, 2009–2017.
- [Dodson *et al.*(1996)] Dodson, E., Kleywegt, G. J., & Wilson, K. (1996). Report of a workshop on the use of statistical validators in protein X-ray crystallography. *Acta Crystallographica Section D*, **52**, 228–234.
- [Dodson & Wlodawer(1998)] Dodson, G. & Wlodawer, A. (1998). Catalytic triads and their relatives. *Trends in Biological Science*, **23**, 347–352.
- [Ehrman *et al.*(1971)] Ehrman, M., Cedar, H., & Schwartz, J. H. (1971). L-asparaginase II of *Escherichia coli*. *The Journal of Biological Chemistry*, **246**(1), 88–94.
- [Evans(1997)] Evans, P. R. (1997). Scala. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, **33**, 22–24.
- [Fitzgerald *et al.*(1990)] Fitzgerald, P. M. D., McKeever, B. M., VanMiddlesworth, J. F., Apringer, J. P., Heinbach, J. C., Leu, C.-T., Herber, W. K., Dixon, R. A. F., & Darke, P. L. (1990). Crystallographic analysis of a complex between Human Immunodeficiency Virus type 1 protease and acetyl-pepstatin at 2.0Å resolution. *The Journal of Biological Chemistry*, **265**(24), 14209–14219.
- [Frankel & Young(1998)] Frankel, A. D. & Young, J. A. T. (1998). HIV-1: Fifteen Proteins and an RNA. *Annual Review of Biochemistry*, **67**, 1–25.

- [Gao *et al.*(1999)] Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., Cumminsk, L. B., Arthur, L. O., Peeters, M., Shaw, G. M., Sharp, P. M., & Hahn, B. H. (1999). Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature*, **397**, 436–441.
- [Gong *et al.*(2000)] Gong, Y. F., Robinson, B. S., Rose, R. E., Deminie, C., Spicer, T. P., Stock, D., Colonna, R. J., & Lin, P. F. (2000). In vitro resistance profile of the human immunodeficiency virus type 1 protease inhibitor BMS-232632. *Antimicrobial Agents and Chemotherapy*, **44**(9), 2319–2326.
- [Gulnik *et al.*(1995)] Gulnik, S. V., Suvorov, L. I., Liu, B., Yu, B., Anderson, B., Mitsuya, H., & Erickson, J. W. (1995). Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochemistry*, **34**, 9282–9287.
- [Harms *et al.*(1991)] Harms, E., Wehner, A., Aung, H.-P., & Röhm, K. H. (1991). A catalytic role for threonine-12 of *E. coli* asparaginase II as established by site-directed mutagenesis. *FEBS Letters*, **285**(1), 55–58.
- [Hong *et al.*(2000)] Hong, L., Zhang, X. C., Hartsuck, J. A., & Tang, J. (2000). Crystal structure of an in vivo HIV-1 protease mutant in complex with saquinavir: Insights into the mechanisms of drug resistance. *Protein Science*, **9**, 1898–1904.
- [Ido *et al.*(1991)] Ido, E., Han, H.-P., Kezdy, F. J., & Tang, J. (1991). Kinetic studies of human immunodeficiency virus type 1 protease and its active-site hydrogen bond mutant A28S. *The Journal of Biological Chemistry*, **266**(36), 24359–24366.
- [Jakob *et al.*(1997)] Jakob, C. G., Lewinski, K., LaCount, M. W., Roberts, J., & Lebioda, L. (1997). Ion binding induces closed conformation in *Pseudomonas* 7A glutaminase-asparaginase (PGA): Crystal structure of the PGA-SO₄²⁻-NH₄⁺ complex at 1.7 Å resolution. *Biochemistry*, **36**, 923–931.

- [Jaskólski *et al.*(2001)] Jaskólski, M., Kozak, M., Lubkowski, J., Palm, G., & Wlodawer, A. (2001). Structures of two highly homologous bacterial L-asparaginases: a case of enantiomorphic space groups. *Acta Crystallographica Section D*, **57**, 369–377.
- [Johnson *et al.*(2004)] Johnson, V. A., Brun-Bézinet, F., Clotet, B., Conway, B., Aquila, R. T. D., Demeter, L. M., Kuritzkes, D. R., Pillay, D., Schapiro, J. M., Telenti, A., & Richman, D. D. (2004). Update of the drug resistance mutations in HIV-1: 2004. *Special Contribution - Drug Resistance Mutations in HIV-1*, **12**(4).
- [Jones *et al.*(1991)] Jones, T. A., Zou, J. Y., Cowan, S. W., & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallographica Section D*, **47**, 110–119.
- [Kempf(1994)] Kempf, D. J. (1994). *Retroviral Proteases*, volume 241 de *Methods in Enzimology*, capítulo 18: Design of simmetry-based peptidomimetic inhibitors of Human Immunodeficiency Virus protease, páginas 334–354. Academic Press.
- [Kent *et al.*(2000)] Kent, S., Marshall, G. R., & Wlodawer, A. (2000). Determining the 3D structure of HIV-1 protease. *Science*, **288**, 1590.
- [Kervinen *et al.*(1996)] Kervinen, J., Thanki, N., Zdanov, A., Tino, J., Barrish, J., Lin, P. F., Colonno, R., Riccardi, K., Samanta, H., & Wlodawer, A. (1996). Structural analysis of the native and drug-resistant HIV-1 proteinases complexed with an aminodiol inhibitor. *Protein and Peptide Letters*, **3**(6), 399–406.
- [Kervinen *et al.*(1998)] Kervinen, J., Lubkowski, J., Zdanov, A., Bhatt, B., Dunn, B. M., Hui, K. Y., Powell, D. J., Kay, J., Wlodawer, A., & Gustchina, A. (1998). Toward a universal inhibitor of retroviral proteases: Comparative analysis of the interactions of LP-130 complexed with proteases from HIV-1, FIV, and EIAV. *Protein Science*, **7**, 2314–2323.
- [Kleywegt(1996)] Kleywegt, G. J. (1996). Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallographica Section D*, **52**, 842–857.

- [Kleywegt & Jones(1994)] Kleywegt, G. J. & Jones, T. A. (1994). Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallographica Section D*, **50**, 178–185. Voidoo program.
- [Kleywegt & Jones(1997)] Kleywegt, G. J. & Jones, T. A. (1997). *Methods in Enzymology*, volume 277, capítulo Good Model-building and Refinement Practice, páginas 208–230. Academic Press.
- [Kummerle & Earlywine(1996)] Kummerle, D. R. & Earlywine, V. (1996). An Overview of HIV Infection. *Pharmacy Times*, páginas 68–73.
- [Laco *et al.*(1997)] Laco, G. S., Schalk-Hihi, C., Lubkowski, J., Morris, G., Zdanov, A., Olson, A., Elder, J. H., Wlodawer, A., & Gustchina, A. (1997). Crystal structure of the inactive D30N mutant of Feline Immunodeficiency Virus protease complexed with a substrate and an inhibitor. *Biochemistry*, **36**, 10696–10708.
- [Lam *et al.*(1994)] Lam, P. Y. S., Jadhav, P. K., Eyermann, C. J., Hodge, C. N., Ru, Y., Bacheler, L. T., Meek, J. L., Otto, M. J., Rayner, M. M., Wong, Y. N., Chang, C. H., Weber, P. C., Jackson, D. A., Sharpe, T. R., & Erickson-Viitanen, S. (1994). Rational design of potent, bioavailable, nonpeptide, cyclic ureas as HIV protease inhibitors. *Science*, **263**, 380–384.
- [Lapatto *et al.*(1989)] Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspi, A., Wood, S., Merson, J. R., Whittle, P. J., Danley, D. E., Geoghegan, K. F., Hawrylik, S. J., Lee, S. E., Scheld, K. G., & Hobart, P. M. (1989). X-ray analysis of HIV-1 proteinase at 2.7Å resolution confirms structural homology among retroviral enzymes. *Nature*, **342**, 299–302.
- [Laskowski *et al.*(1993)] Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK - A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, **26**, 283–291.

- [Leslie(1992)] Leslie, A. G. W. (1992). Recent changes to MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, **26**.
- [Li *et al.*(2000)] Li, M., Morris, G. M., Lee, T., Laco, G. S., Wong, C. H., Olson, J., Elder, J. H., Wlodawer, A., & Gustchina, A. (2000). Structural studies of FIV and HIV-1 proteases complexed with an efficient inhibitor of FIV protease. *PROTEINS: Structure, Function, and Genetics*, **38**, 29–40.
- [Lubkowski *et al.*(1994a)] Lubkowski, J., Wlodawer, A., Housset, D., & Weber, I. T. (1994a). Refined crystal structure of *Acinetobacter glutaminasificans* glutaminase-asparaginase. *Acta Crystallographica Section D*, **50**, 826–832.
- [Lubkowski *et al.*(1994b)] Lubkowski, J., Wlodawer, A., Ammon, H. L., Copeland, T. D., & Swain, A. L. (1994b). Structure characterization of *Pseudomonas* 7A glutaminase-asparaginase. *Biochemistry*, **33**, 10257–10265.
- [Lubkowski *et al.*(1996)] Lubkowski, J., Palm, G. J., Gilliland, G. L., Derst, C., & Röhm, K. H. (1996). Crystal structure and amino acid sequence of *Wolinella succinogenes* L-asparaginase. *European Journal of Biochemistry*, **241**, 201–207.
- [Luft & DeTitta(1999)] Luft, J. R. & DeTitta, G. T. (1999). A Method to produce microseed stock for use in the crystallization of biological macromolecules. *Acta Crystallographica Section D*, **55**, 988–993.
- [Majeed *et al.*(2003)] Majeed, S., Ofek, G., Belachew, A., Huang, C. C., Zhou, T., & Kwong, P. D. (2003). Enhancing protein crystallization through precipitant synergy. *Structure*, **11**, 1061–1070.
- [Margolin *et al.*(1991)] Margolin, N., Dee, A., Lai, M., & Vlahos, C. J. (1991). Purification of recombinant HIV-1 protease. *Preparative Biochemistry*, **21**, 163–173.

- [Matthews(1968)] Matthews, B. W. (1968). Solvent content of protein crystals. *Journal of Molecular Biology*, **33**(2), 491–497.
- [Mauclere(2000)] Mauclere, P. (2000). HIV-1 group N in Cameroon and apparent viruses in the chimpanzee. *Bulletin de la Societe de Pathologie Exotique*, **93**(3), 162.
- [Meek(1992)] Meek, T. D. (1992). Inhibitors of HIV-1 protease. *Journal of Enzyme Inhibition*, **6**, 65–98.
- [Mierendorf *et al.*(1994)] Mierendorf, R., Yeager, K., & Novy, R. (1994). The pET system: Your choice for expression. *Newsletter of Novagen* (<http://www4.web.binc.net/html/vectfram.html>), **1**(1).
- [Mildner *et al.*(1994)] Mildner, A. M., Rothrock, D. J., Leone, J. W., Bannow, C. A., Lull, J. M., Reardon, I. M., Sarcich, J. L., Howe, W. J., Tomich, C. S. C., Smith, C. W., Heinrikson, R. L., & Tomasselli, A. G. (1994). The HIV-1 protease as enzyme and substrate: Mutagenesis of autolysis sites and generation of a stable mutant with retained kinetic properties. *Biochemistry*, **33**, 9405–9413.
- [Miller *et al.*(1989a)] Miller, M., Jaskólski, M., Rao, J. K. M., Leis, J., & Wlodawer, A. (1989a). Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature*, **337**(6207), 576–579.
- [Miller *et al.*(1989b)] Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B. H., & Wlodawer, A. (1989b). Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3Å resolution. *Science*, **246**, 1149–1152.
- [Miller *et al.*(1993)] Miller, M., Rao, J. K. M., Wlodawer, A., & Gribskov, M. R. (1993). Left-handed crossover involved in amidohydrolase catalysis—Crystal structure of *Erwinia chrysanthemi* L-asparaginase with bound L-aspartate. *FEBS Letters*, **328**, 275–279.

- [Navaza(1994)] Navaza, J. (1994). AMORE—An automated package for molecular replacement. *Acta Crystallographica A*, **50**, 157–163.
- [Navia *et al.*(1989)] Navia, M. A., Fitzgerald, P. M. D., McKeever, B. M., Leu, C. T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., & Springer, J. P. (1989). Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature*, **337**, 615–620.
- [Ortlund *et al.*(2000)] Ortlund, E., Lacount, M. W., Lewinski, K., & Lebioda, L. (2000). Reactions of *Pseudomonas* 7A glutaminase-asparaginase with diazo analogues of glutamine and asparagine result in unexpected covalent inhibitions and suggests an unusual catalytic triad Thr-Tyr-Glu. *Biochemistry*, **39**, 1199–1204.
- [Palella *et al.*(1998)] Palella, F. J., Delaney, K. M., Moorman, A. C., Loveless, M. O., Fuhrer, J., Satten, G. A., Aschman, D. J., & Holmberg, S. D. (1998). Declining morbidity and mortality among patients with advanced Human Immunodeficiency Virus infection. *The New England Journal of Medicine*, **338**(13), 853–860.
- [Palm *et al.*(1996)] Palm, G. J., Lubkowski, J., Derst, C., Schleper, S., Röhm, K. H., & Wlodawer, A. (1996). A covalently bound catalytic intermediate in *Escherichia coli* asparaginase: Crystal structure of a Thr-89-Val mutant. *FEBS Letters*, **390**, 211–216.
- [Parikh *et al.*(2002)] Parikh, U., Calef, C., Larder, B., Schinazi, R., & Mellors, J. W. (2002). Mutations in retroviral genes associated with drug resistance. HIV Sequence Database (<http://hiv-web.lanl.gov>) e (http://resdb.lanl.gov/Resist_DB).
- [Perno *et al.*(2001)] Perno, C. A., Cozzi-Lepri, A., Balotta, C., Forbici, F., Violin, M., Bertoli, A., Facchi, G., Pezzotti, P., Cadeo, G., Tositti, G., Pasquinucci, S., Pauluzzi, S., Scalzini, A., Salassa, B., Vincenti, A., Phillips, A. N., Dianzani, F., Appice, A., Angarano, G., Monno, L., Ippolito, G., Moroni, M., Monforte, A. A., & Italian Cohort Naive Antiretroviral Study Group (2001). Secondary mutations in the

- protease region of human immunodeficiency virus and virologic failure in drug-naïve patients treated with protease inhibitor based therap. *The Journal of Infectious Diseases*, **184**, 983–991.
- [Polikarpov *et al.*(1998a)] Polikarpov, I., Oliva, G., Castellano, E. E., Garratt, R. C., Arruda, P., Leite, A., & Craievich, A. (1998a). The protein crystallography beamline at LNLS, the Brazilian National Synchrotron Light Source. *Nuclear Instruments & Methods in Physics Research A*, **405**(1), 159–164.
- [Polikarpov *et al.*(1998b)] Polikarpov, I., Perles, L. A., de Oliveira, R. T., Oliva, G., Castellano, E. E., Garrat, R. C., & Craievich, A. (1998b). Set-up and experimental parameters of the protein crystallography beamline at the Brazilian National Synchrotron Laboratory. *Journal of Synchrotron Radiation*, **5**, 72–76.
- [Polikarpov *et al.*(1999)] Polikarpov, I., Oliveira, R. T., & ao Neto, J. A. (1999). Preparation and preliminary X-ray diffraction studies of a new crystal of L-asparaginase from *Escherichia coli*. *Acta Crystallographica Section D*, **55**, 1616–1617.
- [Reeves & Doms(2002)] Reeves, J. D. & Doms, R. W. (2002). Human Immunodeficiency Virus type 2. *Journal of General Virology*, **83**(6), 1253–1265.
- [Reiling *et al.*(2002)] Reiling, K. K., Endres, N. F., Dauber, D. S., Craik, C. S., & Stroud, R. M. (2002). Anisotropic dynamics of the JE-2147-HIV protease complex: drug resistance and thermodynamic binding mode examined in a 1.09Å structure. *Biochemistry*, **41**(14), 4582–4594.
- [Rizzo & Korant(1994)] Rizzo, C. J. & Korant, D. (1994). *Retroviral Proteases*, volume 241 de *Methods in Enzimology*, capítulo 2: Genetic approaches designed to minimize cytotoxicity of retroviral protease, páginas 16–29. Academic Press.
- [Roberts *et al.*(1990)] Roberts, N. A., Martin, J. A., Kinchington, D., Broadhurst, A. V., Craig, J. C., Duncan, I. B., Galpin, S. A., Handa, B. K., Kay, J., Kröhn, A.,

- Lambert, R. W., Merrett, J. H., Mills, J. S., Parkes, K. E. B., Redshaw, S., Ritchie, A. J., Taylor, D. L., Thomas, G. J., & Machin, P. J. (1990). Rational design of peptide-based HIV proteinase inhibitors. *Science*, **248**, 358–361.
- [Robertson *et al.*(1999)] Robertson, D. L., Anderson, J. P., Brada, J. A., Carr, J. K., Foley, B., Funkhouse, R. K., Gao, F., Hahn, B. H., Kalish, M. L., Kuiken, C., Learn, G. H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P. M., Wolinsky, S., & Korber, B. (1999). HIV-1 Nomenclature Proposal. HIV Sequence Database (<http://hiv-web.lanl.gov>).
- [Röhm & van Etten(1986)] Röhm, K. H. & van Etten, R. L. (1986). The ^{18}O isotope effect in ^{13}C nuclear magnetic resonance spectroscopy: mechanism studies on asparaginase from *Escherichia coli*. *Archives of Biochemistry and Biophysics*, **244**(1), 128–136.
- [Rose *et al.*(1993)] Rose, J. R., Salto, R., & Craik, C. S. (1993). Regulation of autoproteolysis of the HIV-1 and HIV-2 proteases with engineered amino acid substitution. *The Journal of Biological Chemistry*, **268**(16), 11939–11945.
- [Sanches *et al.*(2003)] Sanches, M., Barbosa, A. R. G., Oliveira, R. T., Abrahão-Neto, J., & Polikarpov, I. (2003). Structural comparison of *Escherichia coli* L-asparaginase in two monoclinic space groups. *Acta Crystallographica Section D*, **59**, 416–422.
- [Schechter & Berger(1967)] Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochemical and Biophysical Research Communications*, **27**(2), 157–162.
- [Shafer(2004)] Shafer, R. W. (2004). Genotypic testing for HIV-1 drug resistance. In *HIV InSite Knowledge Base Chapter*.
- [Storoni *et al.*(2004)] Storoni, L. C., McCoy, A. J., & Read, R. J. (2004). Likelihood-enhanced fast rotation functions. *Acta Crystallographica Section D*, **60**, 432–438.

- [Stura & Wilson(1991)] Stura, E. A. & Wilson, I. A. (1991). Applications of the streak seeding technique in protein crystallization. *Journal of Crystal Growth*, **110**, 270–282.
- [Swain *et al.*(1993)] Swain, A. L., Jaskólski, M., Housset, D., Rao, J. K. M., & Wlodawer, A. (1993). Crystal structure of *Escherichia coli* L-asparaginase, an enzyme used in cancer therapy. *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 1474–1478.
- [Tsumoto *et al.*(2003)] Tsumoto, K., Ejima, D., Kumagai, I., & Arakawac, T. (2003). Practical considerations in refolding proteins from inclusion bodies. *Protein Expression & Purification*, **28**, 1–8.
- [Turner & Summers(1999)] Turner, B. G. & Summers, M. F. (1999). Structural Biology of HIV. *Journal of Molecular Biology*, **285**, 1–32.
- [Vaishnav & Wong-Staal(1991)] Vaishnav, Y. N. & Wong-Staal, F. (1991). The Biochemistry of AIDS. *Annual Review of Biochemistry*, **60**, 577–630.
- [Wehner *et al.*(1992)] Wehner, A., Harms, E., Jennings, M. P., Beacham, I. R., Derst, C., Bast, P., & Röhm, K. H. (1992). Site-specific mutagenesis of *Escherichia coli* asparaginase II. *European Journal of Biochemistry*, **208**, 475–480.
- [Winn *et al.*(2001)] Winn, M. D., Isupov, M. N., & Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallographica Section D*, **57**, 122–133.
- [Wlodawer & Erickson(1993)] Wlodawer, A. & Erickson, J. W. (1993). Structure-based inhibitors of HIV-1 protease. *Annual Review of Biochemistry*, **62**, 543–585.
- [Wlodawer & Vondrasek(1998)] Wlodawer, A. & Vondrasek, J. (1998). Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. *Annual Reviews of Biophysics and Biomolecular Structure*, **27**, 249–284.

[Wlodawer *et al.*(1989)] Wlodawer, A., Miller, M., Jaskólski, M., K Sathyanarayana, B., Baldwin, E., T Weber, I., M Selk, L., Clawson, L., Schneider, J., & B H Kent, S. (1989). Conserved folding in retroviral proteases: Crystal structure of a synthetic HIV-1 protease. *Science*, **245**, 616–621.

[Wlodawer *et al.*(1995)] Wlodawer, A., Gustchina, A., Reshetnikova, L., Lubkowski, J., Zdanov, A., Hui, K. Y., Angleton, E. L., Farmerie, W. G., Goodenow, M. M., Bhatt, D., Zhang, L., & Dunn, B. M. (1995). Structure of an inhibitor complex of the proteinase from feline immunodeficiency virus. *Nature Structural Biology*, **2**(6), 480–487.

Apêndice A

Artigo Publicado - HIV PR

Sanches, M., Martins, N. H., Calazans, A., Brindeiro, R. M., Tanuri, A., Antunes, O. A. C., & Polikarpov, I.

Crystallization of a non-B and a B mutant HIV protease

Acta Crystallographica (2004). D60, 1625–1627.

Crystallization of a non-B and a B mutant HIV protease

Mario Sanches,^a Nádia Helena Martins,^a Alexandre Calazans,^b Rodrigo de Moraes Brindeiro,^b Amilcar Tanuri,^b Octavio Augusto Ceva Antunes^c and Igor Polikarpov^{a*}

^aCrystallography Group, Physics Department, USP, São Carlos, SP, Brazil, ^bLaboratory of Molecular Virology, Genetic Department, UFRJ, Rio de Janeiro, RJ, Brazil, and ^cInorganic Chemistry Department, Chemistry Institute, UFRJ, Rio de Janeiro, RJ, Brazil

Correspondence e-mail:
 ipolikarpov@ifsc.usp.br

HIV polymorphism is responsible for the selection of variant viruses resistant to inhibitors used in AIDS treatment. Knowledge of the mechanism of resistance of those viruses is determinant to the development of new inhibitors able to stop, or at least slow down, the disease's progress caused by new mutations. In this paper, the crystallization and preliminary crystallographic structure solution for two multi-resistant 99 amino acid HIV proteases, both isolated from Brazilian patients failing intensive anti-AIDS therapy are presented, *viz.* the subtype B mutant, with mutations Q7K, S37N, R41K, K45R, I54V, L63P, A71V, V82A and L90M, and the subtype F (wild type), naturally carrying mutations Q7K, I15V, E35D, M36I, S37N, R41K, R57K, D60E, Q61N, I62V, L63S, I64L and L89M, with respect to the B consensus sequence. Both proteins crystallized as a complex with the inhibitor TL-3 in space group *P*6₂2. X-ray diffraction data were collected from these crystals to resolutions of 2.1 and 2.6 Å for the subtype B mutant and subtype F wild type, respectively, and the enzyme structures were solved by molecular replacement. The crystals of subtype F HIV protease are, to the best of the authors' knowledge, the first protein crystals obtained for a non-B HIV protease.

Received 3 May 2004
 Accepted 22 June 2004

1. Introduction

Acquired immunodeficiency syndrome (AIDS) is a complex of symptoms and diseases resulting from human immunodeficiency virus (HIV) infection. HIV is a member of the *Lentivirus* genus, that includes simian immunodeficiency virus (SIV) and feline immunodeficiency virus (FIV), among others. The HIV genome encodes for two large polypeptides, *gag* and *gag-pol*, that are subsequently processed by its 99-amino-acid homodimeric aspartyl protease (HIV PR) to render a mature HIV particle. The inhibition of this HIV PR leads to the production of immature non-infectious virus. However, the high polymorphism of HIV results in a rapid selection of viral variants resistant towards a specific inhibitor. Continuous *in vivo* HIV replication and the error prone nature of the viral reverse transcriptase are the driving forces for the generation of drug resistance (Caride *et al.*, 2001).

Almost all studies on HIV-1 drug susceptibility have been performed in developed countries, where subtype B still dominates the epidemic, but on a worldwide scale this is by far not a predominant HIV subtype. The Brazilian epidemic is characterized by multiple HIV-1 group M subtypes, primarily subtype B and subtype F, but also subtypes C and D. From the genes deposited in the HIV Sequence Database (<http://www.hiv.lanl.gov/content/>

hiv-db/mainpage.html), 9.8% of Brazilian infections are of subtype F and 5.0% are of subtype C. Though these distributions represent the frequency in the HIV Database and not the population, it can give an idea of the distribution of subtypes in a given country.

In this work, we present the crystallization of two HIV-1 proteases isolated from Brazilian patients failing treatment. A subtype F wild type (*Fwt*) and a multi-resistant subtype B mutant (*Bmut*) HIV PR have calculated molecular weights of 10.6956 kDa and 10.8078 kDa, respectively, and both were crystallized as complexes with the C₂-symmetric inhibitor TL-3 (Li *et al.*, 2000). This is, to our knowledge, the first report of crystals of a non-B subtype HIV-1 protease.

2. Material and methods

2.1. Expression and purification

The protease sequence of both viruses *Bmut* and *Fwt* carrying the point mutation Q7K were subcloned in pET11a and transformed into *E. coli* BL21-CodonPlus (DE3)-RIL cells (Stratagene). The mutations, according to HXB2 consensus sequence, were Q7K, S37N, R41K, K45R, L63P, I54V, A71V, V82A and L90M (nine point mutations) for *Bmut*, and Q7K, I15V, E35D, M36I, S37N, R41K, R57K, D60E, Q61N, I62V, L63S, I64L and L89M (13 point mutations) for *Fwt*. The transformed

cells were grown at 310 K to an OD₆₀₀ of 0.7 in LB medium supplemented with 1% glucose, 0.01% antifoam 204 (Sigma), 35 µg l⁻¹ of chloramphenicol and 100 µg l⁻¹ of ampicillin. IPTG was added to a final concentration of 1 mM. The cells were harvested after 4 h, re-suspended in lysis buffer (50 mM Tris-HCl pH 8.0, 5 mM EDTA, 5 mM DTT) and lysed using sonication. The inclusion bodies containing the protease were washed in lysis buffer containing 2 M urea and 1% Triton X-100, and then again in lysis buffer. The protease *Fwt* was solubilized during the washing procedure and was purified out of the urea/detergent solution. The washed inclusion bodies of *Bmut* were solubilized in buffer A (50 mM Tris-HCl pH 8.0, 8 M urea, 5 mM EDTA, 0.1% β-mercaptoethanol) and clarified by centrifugation (28 222g) before loading onto a Q Sepharose column (Amersham-Pharmacia), equilibrated with buffer A. The flow-through containing the protease was pooled out and was adjusted to 50 mM NaOAc pH 5.0 and loaded onto a HiTrap SP Sepharose column (Amersham-Pharmacia) equilibrated with buffer B (50 mM NaOAc, pH 5.0, 8 M urea, 2 mM EDTA, 0.1% β-mercaptoethanol). The bound proteins were eluted using a buffer C (buffer B + 1 M NaCl) gradient to a final concentration of 20% of this buffer.

Refoldings of pure proteins were carried out using three changes of refolding buffer (20 mM Na₂HPO₄ pH 7.4, 100 mM NaCl, 2 mM EDTA, 0.1% β-mercapto ethanol) with the protein solution in a 6–8 kDa MWCO dialysis bag. After dialysis, the solution was clarified by centrifugation (28 222g), concentrated by ultrafiltration and stored at 353 K. The protein concentration was determined by the Bradford method.

2.2. Crystallization

The buffer of the *Bmut* protease was changed to 50 mM NaOAc pH 5.0, 2 mM

Table 1
Crystallographic data and preliminary statistics for *Bmut* and *Fwt* HIV proteases.

Values relative to the last resolution shell are shown in parentheses when appropriate. The values for the second-best solution of molecular replacement are also shown in parentheses.

	<i>Bmut</i>	<i>Fwt</i>
Space group	<i>P</i> 6 ₁ 22	<i>P</i> 6 ₁ 22
Cell parameters (Å)	<i>a</i> = <i>b</i> = 62.124, <i>c</i> = 84.309	<i>a</i> = <i>b</i> = 62.750, <i>c</i> = 82.410
Resolution (Å)	53.45–2.10 (2.21–2.10)	54.23–2.60 (2.74–2.60)
No. of subunits in a.u.	1	1
<i>V_M</i> (Å ³ Da ⁻¹)	2.1	2.5
Completeness (%)	100 (100)	100 (100)
<i>I</i> /σ(<i>I</i>)	11.0 (2.0)	8.7 (2.0)
<i>R_{merge}</i> (%)	4.6 (38.1)	7.5 (39.0)
No. of unique reflections	6035	3258
Redundancy (%)	22	17.4
Molecular replacement†		
<i>R</i> factor (%)	40.6 (56.0)	41.5 (55.8)
<i>F</i> correlation (%)	63.8 (29.6)	62.5 (32.3)
<i>I</i> correlation (%)	68.7 (33.6)	67.2 (33.2)

† The resolution range 10–3 Å was used in the molecular replacement procedure for both *Bmut* and *Fwt*.

EDTA and 2 mM DTT in two steps directly in the concentrator, and the protein was finally concentrated to 4.3 mg ml⁻¹ (~0.4 mM). It was then mixed with two molar excesses of TL3 inhibitor (22 mM in DMF/DMSO 1:1) for 1 h. Protease (1 µl) was mixed with an equal volume of well solution in a hanging-drop crystallization experiment at 291 K. The initial screen was based on a variation of precipitants: ammonium sulfate (0.48–1.6 M), MPD [0–6%(v/v)] and PEG 3350 [0–15%(w/v)] or 8000 [0–25%(w/v)] containing 0.1 M sodium cacodylate at pH 5.6 and 6.2. Crystals appeared after 2 d and continued to grow for about two weeks in two conditions [0.1 M sodium cacodylate, pH 6.2, 0.26 M ammonium sulfate, 6%(v/v) MPD, 5.1%(w/v) PEG 3350 (condition 1), and 0.1 M sodium cacodylate pH 6.2, 0.32 M ammonium sulfate, 6%(v/v) MPD (condition 2)] rendering just one crystal per drop (see Fig. 1). The crystal sizes were 0.3 × 0.1 ×

0.1 mm for condition 1 and 0.5 × 0.1 × 0.1 mm for condition 2.

In a similar fashion, *Fwt* at a concentration of 4.9 mg ml⁻¹ (~0.4 mM), in refolding buffer, was mixed with two molar excesses of TL3 inhibitor. Sitting drops with a 4 µl final volume were set up by mixing protein and well solution in equal proportions at 291 K. A shower of small rods (< 0.1 mm) grew using ammonium sulfate (0.64 M and 0.8 M) as precipitant. The number of single crystals per drop was decreased by lowering the temperature to 277 K (0.8 M ammonium sulfate), changing the protein-to-well-solution proportions to 2:1 and increasing the final volume of a sitting drop to 9 µl. Only two crystal rods appeared after approximately one week at 277 K with no further enlargement of the crystals. The crystallization box was then transferred to 291 K, which led to a further growth of crystals to a maximum size of 0.4 × 0.1 × 0.1 mm (Fig. 1).

2.3. Data collection and processing

The diffraction patterns for the *Bmut* crystal, grown in condition 1, and the one of the larger *Fwt* crystal were obtained using Cu Kα radiation generated by a Rigaku ultraX 18 rotating-anode generator and recorded on a MAR 345 dtb image plate. Initial diffraction of the *Bmut* protein was significantly improved by three successive rounds of annealing, during which the crystal was transferred to a room-temperature cryosolution and then flash-cooled again (Fig. 2). The data were processed and scaled using MOSFLM 6.2.3 (Leslie, 1992) and SCALA 3.1.20 (Evans, 1997), respectively (Table 1). The solvent contents of the *Bmut* and *Fwt* crystals are close to 42%, with one protein subunit per asymmetric unit.

2.4. Molecular replacement

An HIV PR subtype B (wild type) X-ray structure solved at 1.09 Å resolution (PDB code 1kzk; Reiling *et al.*, 2002) was used as

Figure 1
Crystals of *Bmut* and *Fwt* proteases. (a) *Bmut* grown in condition 1 (used in data collection), (b) *Bmut* grown in condition 2, (c) *Fwt* initial screen obtained at 291 K using 0.8 M ammonium sulfate as precipitant, and (d) *Fwt* final crystals (used in data collection).

Figure 2

Diffraction patterns of the two proteases. (a) A *Bmut* crystal before annealing, (b) the same *Bmut* crystal after three cycles of annealing, and (c) an *Fwt* crystal.

the search model in the molecular replacement of both *Bmut* and *Fwt* proteases. The amino-acid sequence of the model is 92 and 88% identical to *Bmut* and *Fwt*, respectively. In both cases, an *AMoRe* (Navaza, 1994) translation function gave a clear solution in space group $P6_122$ (Table 1), whereas the same procedure with the data processed in $P6_522$ presented no contrast (data not shown).

3. Discussion

The production of purified protein in the amount required for crystallization screens is one of the key points in its structure determination. For that, the inclusion of the Q7K mutation, which is known to enhance protease stability (Rosé *et al.*, 1993), was crucial in attempts to obtain considerable expression levels of both *Bmut* and *Fwt* proteases. Concerning the crystallization experiments, it is important to notice the role of a multi-precipitant procedure in obtaining *Bmut* protease crystals (see, for example, Majeed *et al.*, 2003). Even though

the initial screen included conditions with a single precipitant, the crystals appeared in the drops containing at least two precipitants of different classes: a salt, an organic solvent and a polymeric precipitant in condition 1, and a salt and an organic solvent in condition 2.

The molecular replacement approach for both *Bmut* and *Fwt* yielded clear solutions, which is not surprising given the high amino-acid sequence identity between the search model and multi-resistant B mutant and non-B proteases crystallized by us. Refinement of the crystallographic models is currently in progress. The structures, when complete, will help to explain the molecular basis of resistance of these protease variants to commercial inhibitors.

We thank Alexander Wlodawer and Alla Gustchina (NCI-Frederick) for valuable help and friendly support, and gratefully acknowledge financial support by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brazil, via grant Nos.

99/03387-4, 99/53087-4, 00/03674-2 and 03/09461-9, and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

References

- Caride, E., Hertogs, K., Larder, B., Dehertogh, P., Brindeiro, R., Machado, E., de Sá, C. A. M., Eyer-Silva, W. A., Sion, F. S., Passioni, L. F. C., Menezes, J. A., Calazans, A. R. & Tanuri, A. (2001). *Virus Genes*, **23**, 193–202.
- Evans, P. R. (1997). *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **33**, 22–24.
- Leslie, A. G. W. (1992). *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **26**.
- Li, M., Morris, G. M., Lee, T., Laco, G. S., Wong, C. H., Olson, J., Elder, J. H., Wlodawer, A. & Gustchina, A. (2000). *Proteins*, **38**, 29–40.
- Majeed, S., Ofek, G., Belachew, A., Huang, C., Zhou, T. & Kwong, P. D. (2003). *Structure*, **11**, 1061–1070.
- Navaza, J. (1994). *Acta Cryst. A* **50**, 157–163.
- Reiling, K. K., Endres, N. F., Dauber, D. S., Craik, C. S. & Stroud, R. M. (2002). *Biochemistry*, **41**, 4582–4594.
- Rosé, J. R., Salto, R. & Craik, C. S. (1993). *J. Biol. Chem.* **268**, 11939–11945.

Apêndice B

Artigo Publicado - L-asparaginase

Sanches, M., Barbosa, J. A. R. G., Oliveira, R. T., Abrahão-Neto, J., & Polikarpov, I.

Structural comparison of *Escherichia coli* L-asparaginase in two monoclinic space groups

Acta Crystallographica (2003). **D59**, 416–422.

Structural comparison of *Escherichia coli*
L-asparaginase in two monoclinic space groupsMario Sanches,^a João Alexandre
R. G. Barbosa,^b Ricardo Toledo
de Oliveira,^b José Abrahão Neto^c
and Igor Polikarpov^{a,b*}^aGrupo de Cristalografia, Departamento de
Física em São Carlos, USP, Av. Trabalhador
São-Carlense 400, CEP 13560-970, São Carlos/
SP, Brazil, ^bLaboratório Nacional de Luz
Síncrotron (LNLS), CP 6192, CEP 13084-971,
Campinas/SP, Brazil, and ^cDepartamento de
Bioquímica e Tecnologia Farmacêutica, Escola
de Ciências Farmacêuticas, Universidade de São
Paulo, SP, BrazilCorrespondence e-mail:
ipolikarpov@ifsc.usp.brReceived 29 October 2002
Accepted 19 November 2002PDB Reference: L-aspara-
ginase, 1nns, r1nnsf.

The functional L-asparaginase from *Escherichia coli* is a homotetramer with a molecular weight of about 142 kDa. The X-ray structure of the enzyme, crystallized in a new form (space group *C2*) and refined to 1.95 Å resolution, is compared with that of the previously determined crystal form (space group *P2*₁). The asymmetric unit of the new crystal form contains an L-asparaginase dimer instead of the tetramer found in the previous crystal form. It is found that crystal contacts practically do not affect the conformation of the protein. It is shown that subunit C of the tetrameric form is in a conformation which is systematically different from that of all other subunits in both crystal forms. Major conformational differences are confined to the lid loop (residues 14–27). In addition, the stability of this globular protein is analyzed in terms of the interactions between hydrophobic parts of the subunits.

1. Introduction

L-Asparaginase is an enzyme that catalyses the hydrolysis of L-asparagine to L-aspartate and ammonia. Asparaginases are expressed in many bacterial organisms and the crystallographic structures of asparaginases from *Erwinia chrysanthemi* (Miller *et al.*, 1993), *Pseudomonas* 7a (Lubkowski, Wlodawer, Ammon *et al.*, 1994), *Wolinella succinogenes* (Lubkowski *et al.*, 1996) and *Escherichia coli* (Swain *et al.*, 1993; Polikarpov *et al.*, 1999) have been described. Two L-asparaginases, namely asparaginases I and II, have been isolated from *E. coli*. Type II L-asparaginase is located near the cell surface and is highly active against asparagine-deficient tumours in mammals, whereas type I L-asparaginase is constitutive and does not inhibit tumour growth (Cedar & Schwartz, 1967). Type II L-asparaginase is a chemotherapeutic agent used in the treatment of paediatric acute lymphoblastic leukaemia. Its effectiveness comes from the rapid and complete depletion of circulating asparaginase. However, the treatment is limited by its toxicity, multiple side effects and spontaneous resistance (Chakrabarti & Scuster, 1997).

The active *E. coli* L-asparaginase acts as a homotetramer of 142 kDa, with 326 amino-acid residues per subunit (Epp *et al.*, 1971). Each subunit of the tetramer makes two types of contact with the neighbouring subunits. One could be called an intimate contact, leading to formation of the intimate dimer (Swain *et al.*, 1993), and the other could be termed a distant contact, forming the distant dimer. The intimate dimer accommodates two separate active sites in its interface. All but one of the L-asparaginases structurally determined to date crystallize with a homotetramer or a homodimer of the enzyme in the asymmetric unit cell, with L-asparaginase-glutaminase from *Acinetobacter glutaminasificans* being the

only exception. This enzyme crystallized in the orthorhombic space group $I222$, with a monomer in the asymmetric unit (Lubkowski, Wlodawer, Housset *et al.*, 1994). In this case, the functional tetramer could be generated by crystal symmetry operations. L-asparaginase from *E. coli* was originally crystallized in the monoclinic space group $P2_1$, with a homotetramer in the asymmetric unit. The structure was refined to 2.4 Å resolution (PDB code 3eca; Swain *et al.*, 1993). We will refer to this crystal form as crystal form *A*. Structures of mutants T98V and Y25F in the space groups $P2_12_12_1$ and $P6_522$, respectively, have also been determined (Palm *et al.*, 1996; Jaskólski *et al.*, 2001).

Native L-asparaginase from *E. coli* was crystallized in our laboratory in a different monoclinic space group ($C2$; Polikarpov *et al.*, 1999), which we will subsequently refer to crystal form *B*. Crystal form *B* contains two subunits of the enzyme (a dimer) in the asymmetric unit. The purpose of this paper is to report the structure of the native L-asparaginase in crystal form *B*, refined to 1.95 Å resolution, and to compare it with the crystal form *A* structure.

2. Materials and methods

2.1. Crystallization and data-collection

Crystallization of *E. coli* L-asparaginase in crystal form *B* has been described in Polikarpov *et al.* (1999). Briefly, L-asparaginase purchased in a lyophilized form from Merck, Sharp and Dohme was crystallized using a sparse-matrix screen at 291 K. The crystals were grown using the hanging-drop vapour-diffusion technique by mixing equal volumes of 20 mg ml⁻¹ protein solution and of reservoir solution containing 30% 2-methyl-2,4-pentanediol, 4% PEG 3350, 0.1 M 2-(*N*-morpholino)ethanesulfonic acid buffer pH 6. Crystals of dimensions 0.5 × 0.3 × 0.2 mm appeared after one week.

2.2. Data collection and refinement

X-ray diffraction data were collected at the dedicated protein crystallography beamline at the Brazilian National Synchrotron Light Laboratory (Polikarpov, Oliva *et al.*, 1998; Polikarpov, Perles *et al.*, 1998). The data were processed using *DENZO* and *SCALEPACK* (Otwinowski, 1993). The structure was solved by the molecular-replacement method with *AMoRe* (Navaza, 1994), using a distant dimer of crystal form *A* L-asparaginase as the search model. The rotation function was calculated using diffraction data in the resolution range 10.0–4.0 Å and a Patterson radius of 36 Å. The translation search was performed using the Crowther and Blow translation function (Crowther & Blow, 1967). Both rotation and translation searches resulted in clear solutions well above the noise level (Polikarpov *et al.*, 1999). Further positional and *B*-factor refinement was performed with *REFMAC* and *REFMAC5* (Murshudov *et al.*, 1997). Water molecules were inserted with *ARP/wARP* (Lamzin & Wilson, 1993). Details of the data-collection statistics and refinement are given in Table 1. The final calculated electron density was of excellent

Table 1

Crystallographic data and statistics for L-asparaginase (crystal form *B*).

Values in parentheses are for the last resolution shell (2.0–1.95 Å).

Crystal dimensions (mm)	0.5 × 0.3 × 0.2
Space group	$C2$
Unit-cell parameters (Å, °)	$a = 76.3$, $b = 134.6$, $c = 64.8$, $\beta = 110.5$
Z (No. of chains in a.u.)	2
V_M (Matthews coefficient)	2.19 Å ³ Da ⁻¹
Resolution (Å)	13.0–1.95
$I/\sigma(I)$	12.7 (2.4)
Completeness (%)	91.8 (93.2)
R_{merge} (%)	4.7 (32.5)
R factor (%)	13.4
R_{free} (%)	17.4
Unique reflections	40 782
Observed reflections	63 394
Ramachandran plot	
Most favoured (%)	92.8
Allowed (%)	6.5
Generously allowed (%)	0.4
Disallowed (%)	0.4
R.m.s.d. bond lengths (Å)	0.014
R.m.s.d. bond-angle distances (Å)	0.032

Figure 1

Representative part of the electron density together with the respective part of the crystallographic model. Asterisks mark residues that belong to a neighbouring subunit.

quality for all the present crystallographic model, except perhaps for the flexible loop (residues 14–27) (see Fig. 1).

2.3. Structure comparison and analyses

In crystal form *B*, the functional tetramer of L-asparaginase is located on the twofold crystallographic axis. Therefore, the contents of the asymmetric unit comprise a dimer. For the purpose of comparison with crystal form *A*, a tetramer was generated using the program *O* (Jones & Kjeldgaard, 1993) by applying C2 symmetry operations to the crystallographically independent dimer. Subunits *A* and *B* are crystallographically related to subunits *C* and *D*, respectively. The root-mean-square deviations (r.m.s.d.s) were calculated using the program *LSQKAB* (Collaborative Computational Project, Number 4, 1994). Intermolecular and intersubunit contacts were calculated for both crystal forms using the program *CONTACT* (Collaborative Computational Project, Number 4, 1994). Only contacts with some probability of forming hydrogen bonds, according to the program, were considered. The program uses a simple bricking algorithm, where every atom is placed in a $6 \times 6 \times 6$ Å box and the contacts are limited to neighbouring boxes. Hydrogen bonds are assigned to every contacting donor/acceptor pair. Interfacial hydrogen bonds were calculated using the program *DIMPLLOT* (Wallace *et al.*, 1995), which employs the program *HBPLUS* with a somewhat more elaborate algorithm. The program first determines the theoretical positions of the H atoms relative to each possible donor, using the ideal hybridization angle to the given atom. Each donor/acceptor pair is then analyzed with respect to its distances and angles. Those that match the limits stipulated by the program are considered to be high-probability hydrogen bonds. Secondary-structure elements were assigned by *PROCHECK* (Laskowski *et al.*, 1993). The buried area (BA) was obtained using the accessible surface area (ASA) calculated using the program *NACCESS* (Hubbard & Thornton, 1993). ASA was calculated for each subunit, each dimer (distant and intimate) and the whole molecule. The buried area was calculated using the expression

$BA_{xy} = (ASA_x + ASA_y) - ASA_{xy}$, where ASA_x and ASA_y were calculated for each separate interacting chain and ASA_{xy} was calculated for both. The theoretical number of hydrogen bonds per Å² is given by $n = 5.34s \times 10^{-3}$ Å², where s is the buried area of the interface (Xu *et al.*, 1997).

Planarity was calculated using the program *SURFNET* (Laskowski, 1995). The r.m.s.d.s of all the interface atoms from the least-squares plane through all these atoms was calculated. The gap index (GI) was calculated using the expression $GI = GV - ASA$, where GV is the gap volume between the interfaces, which was also calculated using the program *SURFNET* (Laskowski, 1995).

3. Results and discussion

3.1. Structure description

The structure of an L-asparaginase subunit consists of two α/β domains (Swain *et al.*, 1993). The N-terminal flavodoxin-like domain contains nine α -helices and ten β -strands and the C-terminal domain consists of five α -helices and four β -strands. In each domain there is a parallel β -strand core which forms a twisted plane involving α -helices. The connection between the N- and C-terminal domains is achieved by a random-coil sequence encompassing residues 201–213. There is a single disulfide bond in each subunit, connecting Cys77, which is located in a random coil between N α 5 and N α 6, and Cys105, which is located in a random coil between N α 7 and N α 4. A structural cartoon of the subunit is given in Fig. 2.

Table 2
Main-chain, side-chain and all-atom r.m.s.d.s (in Å), respectively, between all subunits of the two structures. The crystal form is indicated in parentheses.

Chains	<i>B</i> (form <i>B</i>)	<i>A</i> (form <i>A</i>)	<i>B</i> (form <i>A</i>)	<i>C</i> (form <i>A</i>)	<i>D</i> (form <i>A</i>)
<i>A</i> (form <i>B</i>)	0.095, 0.797, 0.548	0.317, 0.910, 0.663	0.360, 1.138, 0.821	0.493, 1.163, 0.871	0.366, 1.228, 0.880
<i>B</i> (form <i>B</i>)		0.322, 1.121, 0.800	0.361, 0.903, 0.671	0.499, 1.194, 0.892	0.374, 1.237, 0.887
<i>A</i> (form <i>A</i>)			0.287, 1.118, 0.791	0.432, 1.178, 0.864	0.284, 1.225, 0.861
<i>B</i> (form <i>A</i>)				0.373, 1.180, 0.851	0.304, 1.222, 0.863
<i>C</i> (form <i>A</i>)					0.426, 1.252, 0.910

Table 3
R.m.s.d.s (in Å) for the loop between residues 14 and 27, comparing the same subunit of the two crystal forms.

Chain loop	Main chain	Side chain	All atoms
<i>A</i>	0.368	1.447	0.992
<i>B</i>	0.701	1.991	1.400
<i>C</i>	0.985	2.109	1.562
<i>D</i>	0.616	1.854	1.293

Figure 4
R.m.s.d.s between each related subunit of both crystal forms. Below each plot there is a schematic representation of the crystalline contacts. Each amino acid engaged in a crystalline contact is represented by an arrow: up arrows indicate form *A* and down arrows indicate form *B*.

Each L-asparaginase tetramer contains four active sites. The active sites are located at the interface between two subunits forming an intimate dimer (Swain *et al.*, 1993); there are two active sites per interface. The distant dimer interface does not contain active sites. The molecule assembles in a tetramer which may therefore be described as a dimer of dimers (Fig. 3). Crystal form *B* contains one distant dimer (*A*–*B*) in the asymmetric unit. Another distant dimer (*C*–*D*) in crystal form *B* resides in a neighbouring asymmetric unit. The intimate dimers *A*–*C* and *B*–*D* are related by a twofold crystallographic axis in this particular crystal form. The only two residues in disallowed regions of the Ramachandran plot are Thr198 of both subunits, both of which display a perfect fit to the electron density. The reason for the deviation from the allowed conformation seems to be the hydrogen-bonded interaction of Thr198 OG1 with a water molecule coordinated by His197 O and Ser199 OG. The form *B* structure has a slightly higher number of water molecules compared with form *A* (658 per tetramer in form *B* compared with 410 in the form *A* tetramer). This difference may be attributed to the higher resolution of form *B*. The r.m.s.d. between the subunits in the two structures are given in Table 2. As can be seen, one chain (subunit *C*) systematically presents r.m.s.d. values that are somewhat higher than those of the other chains. The r.m.s.d.s for the subunit *C* main-chain atom positions compared with all other subunits of both crystal forms are 0.452 ± 0.051 Å, whereas the r.m.s.d.s of the main-chain atoms of all other subunits superimposed pairwise amongst themselves are 0.336 ± 0.097 Å. The r.m.s.d.s for the whole tetramer are 0.46 Å (main chain), 1.09 Å (side chains) and 0.81 Å (all atoms) between crystal forms *A* and *B*. The highest r.m.s.d.s correspond to the loop between residues 14 and 27 (Table 3). Electron density in this region is poor in both crystal forms (Swain *et al.*, 1993). Here, chain *C* clearly stands out as having the highest r.m.s.d. values. Graphical representations of the r.m.s.d.s between each related subunit of both crystal forms are given in Fig. 4. Disorder in the region of the loop (residues 14–27) can also be seen from the *B*-factor plot (Fig. 5). The *B* factors of the polypeptide chains are particularly high in this region and are quite different for all four chains in crystal form *A*, with chain *A* having the lowest *B* factors and chain *C* presenting the highest *B* factors. Remarkably, the thermal parameters of the ligand L-asparagine vary from subunit to subunit in a similar way, being most disordered in subunit *C*. *B*-factor plots for each of the crystallographically independent subunits in the crystal form *B* are very similar, although no *B*-factor NCS restraints were applied during the refinement.

3.2. Conformation and stability

Even though the intimate dimer theoretically possesses all the necessary conditions for activity, L-asparaginases act as

Table 4
Surface–surface interaction parameter calculations.

The parameters were calculated for each possible interaction between two subunits of the tetramer of crystalline form *B*.

Interaction	BA† (Å ²)	%‡	n _T §	n _D §	P¶ (Å)	GI†† (Å)	Type
AC	4667.73	16.28	25	28	6.22	1.72	Intimate
BD	4689.65	16.09	25	30	6.12	1.82	Intimate
AB = CD	1877.95	6.55	10	7	2.70	2.76	Distant
AD	2049.83	7.12	11	16	2.84	2.86	Distant
BC	2049.76	7.06	11	16	2.84	2.83	Distant

† Buried area. ‡ The percentage refers to the amount of area buried between the two interactive parts compared with the sum of the total accessible surface area of each subunit alone. § The theoretical number of hydrogen bonds by interface (*n_T*, hydrogen bonds per Å²) and this number determined using the program *HBPLUS* (McDonald & Thornton, 1994) (*n_D*, hydrogen bonds per Å²). ¶ Planarity. †† Gap index.

tetramers. Why then does the enzyme assemble in tetramers under physiological conditions? The tetrameric assembly of the protein under physiological conditions may be explained, at least partially, by the nature of the dimeric interface interactions. The interface between intimate dimers (*i.e.* the interface between dimer AC and a dimer BD) can be seen in Figs. 6(*a*) and 6(*b*). It is clear that the interaction surface is mainly hydrophobic, with the majority of the charged and polar residues being in contact with solvent on the external surface of the enzyme. In line with the results of Jones & Thornton (1996), the interaction between the dimers bears the characteristics of physiologically relevant complexes, being more hydrophobic and presenting low intersubunit hydrogen-bond content, as can be seen in Table 4. The tetramer keeps the most hydrophobic part of the protein buried inside of its core, away from contact with solvent, providing stabilization and rendering a globular-shaped molecule, with the polarizable surface oriented towards the external medium (Figs. 6*b* and 6*c*).

L-Asparaginases are very stable enzymes which preserve their activity over a wide pH (4.5–11.5) and temperature range (Stecher *et al.*, 1999; Cammack *et al.*, 1972). Moreover, on being exposed to extreme pH and brought back to physiological pH, the enzyme activity is restored (Stecher *et al.*, 1999; Cammack *et al.*, 1972). The very robust active sites are buried in the interface between the intimate dimers (see BA in Table 4) and are not directly affected by pH changes. Above and below the range of pH stability, the side chains of charged residues are considerably affected (deprotonated or protonated) and therefore presumably interfere with the intramolecular electrostatic interactions, protein folding and subunit assembly.

The buried area between subunits forming intimate dimers is composed of about 56% non-polar residues and represents about 16% of the total area of the subunits of an intimate dimer (Table 4). Comparison of the two crystal forms, *A* and *B*, shows that there is no significant difference in buried area between them, indicating that they are not affected by crystal packing. The theoretical accessible surface area is given by the expression $ASA = 6.3m^{0.73} \text{ \AA}^2$, where *m* is the molecular mass of the subunit or oligomer (Miller *et al.*, 1987). The area buried upon formation of the oligomer can be calculated using this

expression as the difference between the total surface of the exposed areas of subunits composing the oligomer and that of the oligomer itself. Theoretically, the area buried upon formation of a dimer calculated in this manner from two subunits with a molecular mass of 35.5 kDa is equal to 4510 Å², which is very close to the buried area for the intimate dimers calculated on the basis of the crystallographic model (4660 Å²).

The interacting surface area in an intimate dimer shows a relatively high deviation from planarity and a low gap index (Table 4), indicating a strong interaction between the molecules. The number of hydrogen bonds between subunits forming the intimate dimers calculated using *HBPLUS* is slightly higher than the expected theoretical value (Table 4). The same is observed for distant interaction between subunits AD and BC, respectively. The other two interactions, AB and CD, present fewer hydrogen bonds than expected, owing to the apolar nature of this interaction. This pattern suggests that

Figure 5
B factors of C^α positions. (*a*) Form *A*, (*b*) form *B*. The separated plot, marked Asp, at the very right of the figure represents *B* factors of all atoms of the ligand aspartate.

the tetrameric assembly can be viewed as an association of two intimate dimers maintained mainly by *AD* and *BC* interactions. Additionally, according to the study of Xu *et al.* (1997), there are an average of two salt bridges per interface. The intimate dimer assembly matches this value, with the strong interactions on interfaces *AD* and *BC* including six possible salt bridges compared with none on the other two interfaces.

3.3. Crystal contacts

Crystal contacts are quite different in crystal forms *A* and *B*. The amino-acid residues involved in these contacts are flagged at the bottom of Fig. 4. Monomer *C* of crystal form *A* participates in 14 crystal contacts, compared with the eight, 11 and nine contacts in which subunits *A*, *B* and *D* are involved, respectively. Of these 14 contacts, seven are specific to this subunit (not observed in other subunits) compared with one, four and three specific contacts of subunits *A*, *B* and *D*, respectively. The crystal form *B* subunits (chains *A* and *B*) are involved in nine crystalline contacts each. All of these contacts are specific. In contrast, the non-crystallographic contacts are very similar in all subunits of both crystal forms.

The r.m.s.d.s calculated only for residues involved in crystallographic contacts amount to 0.46 Å for the main-chain

atoms, 1.79 Å for the side-chain atoms and 1.30 Å for all atoms. Interestingly, the r.m.s.d. for the main-chain atoms of residues involved in the crystallographic contacts is exactly the same as for all main-chain atoms of the protein. Also, significant differences in r.m.s.d. are only observed for the side-chain atoms (Table 2 and §3.1). This indicates that the conformational differences of the side chains induced by the crystalline lattice do not affect the conformation of the protein to any great extent.

4. Conclusions

The detailed crystallographic comparison of the two crystal forms of native *E. coli* L-asparaginase presented in this paper demonstrates that the enzyme conformation is not significantly affected by crystallographic contacts. No significant conformational changes in protein structure are observed. The only exception seems to be subunit *C* of the tetramer in crystalline form *A* (space group *P2*₁). The 14–27 loop of this subunit systematically presents higher r.m.s.d.s compared with the subunits in the other crystal form as well as the other subunits of the same tetramer (Table 3). Moreover, the 14–27 loop is more flexible or disordered in this particular subunit compared with the same loop of the rest of the tetramer and the dimer of crystal form *B*. This loop is proposed to be involved in the reaction process (Swain *et al.*, 1993; Palm *et al.*, 1996), probably acting as a lid to the active site. Moreover, the crystal contacts of the subunit *C* are very specific, indicating a systematic difference of this subunit compared with the other three subunits of the same tetramer. There could be two explanations for this phenomenon. First of all, the subunit *C* conformation may be affected by the crystalline contacts. The second explanation is that subunit *C*, even in solution before crystallization, was (and continues to be) in a particular conformation different from that of all the other subunits. This conferred the specific orientation of the L-asparaginase tetramer selected during the molecular crystal assembly and growth and therefore the particular position of subunit *C*. We strongly favour the second hypothesis. There are several arguments in support of this hypothesis. First of all, the r.m.s.d.s of the subunit *C* polypeptide chain are not correlated with the crystalline contact positions (Fig. 4 and Tables 2, 3 and 5). Secondly, the temperature factors of the lid loop 14–27 and the ligand for this particular subunit are significantly higher than the *B* factors of the same loop and the ligands of the rest of subunits in both crystal forms (Fig. 5). Since the occupancy of the ligand was set to 1, the *B* factors represent both its disorder and its low occupancy. Thirdly, the *E. coli* enzyme contains an average of three molecules of L-aspartic acid per molecule of tetramer (*i.e.* per four molecules of subunits; Jayaram *et al.*, 1986) and it is tempting to propose that the occupancy of

Figure 6
GRASP (Nicholls *et al.*, 1991) representation of the molecular surface of the intimate dimer (*A–C*). (*a*) The blue area shows the contact surface between two intimate dimers in the range 2.5 Å. (*b*) The same view as in (*a*), but showing hydrophobic residues in white, polar residues in yellow and charged residues in red. (*c*) External part of the dimer. Hydrophobic areas are white, polar areas are yellow and charged niches are red. (*d*) The same point of view as in (*c*), but showing an electrostatic potential surface.

the ligand in the active site of one particular subunit (subunit C) is very low. Lastly, this explains the high factors of the lid loop (residues 14–27), which is more disordered in the apo enzyme (Lubkowski *et al.*, 1996). This loop serves as a flexible lid that mediates the access of the ligand to the active site, as observed previously in the L-asparaginase from *W. succinogenes* (Lubkowski *et al.*, 1996).

The stability of the tetramer can be explained in terms of the interface interactions of the subunits. The assembly of the intimate dimers is mainly hydrophobic in nature (Table 4), with a number of characteristics of permanent physiologically relevant complexes (Jones & Thornton, 1996). This indicates that the entropic contribution is important for the assembly of intimate dimers, whereas significant enthalpic contribution seems to arise from the ion-pair interactions on the distant dimer interfaces A–D and B–C.

We gratefully acknowledge financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brazil via grants Nos. 99/03387-4, 98/12043-4 and 0/03674-2.

References

- Cammack, K. A., Marlborough, D. I. & Miller, D. S. (1972). *Biochem. J.* **126**, 361–379.
- Cedar, H. & Schwartz, J. H. (1967). *J. Biol. Chem.* **242**, 3753–3755.
- Chakrabarti, R. & Scuster, S. M. (1997). *Int. J. Pediatr. Hematol./Oncol.* **4**, 597–611.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst. D50*, 760–763.
- Crowther, R. A. & Blow, D. M. (1967). *Acta Cryst.* **23**, 544–548.
- Epp, O., Steigemann, W., Formanek, H. & Huber, R. (1971). *Eur. J. Biochem.* **20**, 432–437.
- Hubbard, J. S. & Thornton, J. M. (1993). *NACCESS*. Department of Biochemistry and Molecular Biology, University College London.
- Jaskólski, M., Kozak, M., Lubkowski, J., Palm, G. & Wlodawer, A. (2001). *Acta Cryst. D57*, 369–377.
- Jayaram, H. N., Cooney, D. A. & Huang, C. Y. (1986). *J. Enzyme Inhib.* **1**, 151–161.
- Jones, S. & Thornton, J. M. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 13–20.
- Jones, T. A. & Kjeldgaard, M. (1993). *O Version 5.9, The Manual*. Uppsala University, Uppsala, Sweden.
- Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
- Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst. D49*, 129–147.
- Laskowski, R. A. (1995). *J. Mol. Graph. Model.* **13**, 323–330.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
- Lubkowski, J., Palm, G. J., Gilliland, G. L., Derst, C. & Rohm, K.-H. (1996). *Eur. J. Biochem.* **241**, 201–207.
- Lubkowski, J., Wlodawer, A., Ammon, H. L., Copeland, T. D. & Swain, A. L. (1994). *Biochemistry*, **33**, 10257–10265.
- Lubkowski, J., Wlodawer, A., Housset, D., Weber, I. T., Ammon, H. L., Murphy, K. C. & Swain, M. L. (1994). *Acta Cryst. D50*, 826–832.
- McDonald, I. K. & Thornton, J. M. (1994). *J. Mol. Biol.* **238**, 777–793.
- Miller, S., Lesk, A. M., Janin, J. & Chothia, C. (1987). *Nature (London)*, **328**, 834–836.
- Miller, M., Rao, J. K. M., Wlodawer, A. & Gribskov, M. R. (1993). *FEBS Lett.* **328**, 275–279.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst. D53*, 240–255.
- Navaza, J. (1994). *Acta Cryst. A50*, 157–163.
- Nicholls, A., Sharp, K. & Honig, B. (1991). *Proteins Struct. Funct. Genet.* **11**, 281–296.
- Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
- Palm, G. J., Derst, C., Lubkowski, J., Schleper, S., Rohm, K. H. & Wlodawer, A. (1996). *FEBS Lett.* **390**, 211–216.
- Polikarpov, I., Oliva, G., Castellano, E. E., Garratt, R. C., Arruda, P., Leite, A. & Craievich, A. (1998). *Nucl. Instrum. Methods Phys. Res. A*, **405**, 159–164.
- Polikarpov, I., Oliveira, R. T. & Abrahão-Neto, J. (1999). *Acta Cryst. D55*, 1616–1617.
- Polikarpov, I., Perles, L. A., de Oliveira, R. T., Oliva, G., Castellano, E. E., Garratt, R. C. & Craievich, A. (1998). *J. Synchrotron. Rad.* **5**, 72–76.
- Stecher, A. L., Deus, P. M., Polikarpov, I. & Abrahão-Neto, J. (1999). *Pharm. Acta Helv.* **290**, 1–9.
- Swain, A. L., Jaskólski, M., Housset, D. & Rao, J. K. M. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 1474–1478.
- Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). *Protein Eng.* **8**, 177–134.
- Xu, D., Tsai, C.-J. & Nussinov, R. (1997). *Protein Eng.* **10**, 999–1012.