

Fernanda Canduri

Estudo Estrutural de Proteinases Aspárticas
por Cristalografia de Difração de Raios X



1910001027



Fernanda Canduri

Estudo Estrutural de Proteinases Aspárticas por Cristalografia de Difração de Raios X



T 1.027/02

Tese para obtenção do título de Doutor em Biofísica Molecular, área de concentração em Biofísica Molecular apresentada ao Departamento de Física do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho" – UNESP.

Orientador: Prof. Dr. Walter Filgueira de Azevedo Jr.

Co-orientador: Prof. Dr. João Ruggiero Neto

São José do Rio Preto – SP

2001

577.15
C219e

Enzimas

Assunto CAPES:

2.08.05.00-4 - Enzimologia

Canduri, Fernanda.

Estudo estrutural de proteinases aspárticas por cristalografia de difração de raios X / Fernanda Canduri. – São José do Rio Preto : [s.n.], 2001.

139 f. : il. ; 30 cm.

Orientador: Walter Filgueira de Azevedo Junior.

Tese (doutorado) – Universidade Estadual Paulista. Instituto de Biociências, Letras e Ciências Exatas.

1. Enzimas. 2. Proteinases aspárticas. 3. Cristalografia. 4. Estrutura. 5. Uropepsina. I. Azevedo Junior, Walter Filgueira de. II. Universidade Estadual Paulista. Instituto de Biociências, Letras e Ciências Exatas. III. Título.

CDU – 577.15



Data da Defesa: 26 de junho de 2001.

Banca Examinadora

Prof. Dr. Walter Filgueira de Azevedo Jr. (Orientador)

Profa. Dra. Yvonne Primerano Mascarenhas

Prof. Dr. Mário Sérgio Palma

Profa. Dra. Roseli A. da Silva Gomes

Prof. Dr. Luis Fernando Delboni

Este trabalho foi realizado no Departamento de Física do Instituto de Biociências, Letras e Ciências Exatas de São José do Rio Preto, da Universidade Estadual Paulista “Júlio de Mesquita Filho”, com auxílio da Fundação de Amparo à Pesquisa do Estado de São Paulo (Bolsa de Doutorado, processo n. 98/07155-8).



À Ana
Minha companheira
incondicional
em todas as etapas
deste trabalho
e em todos os dias
da minha vida.

Agradecimentos

Ao término deste trabalho expressei meu agradecimento a todos que colaboraram para a sua concretização.

De forma especial ao Prof. Dr. Walter Filgueira de Azevedo Jr., a quem devo o melhor dos meus agradecimentos, pela confiança, pelo apoio e pela presença amigável em todas as etapas vencidas, e por me proporcionar crescimento e amadurecimento científico.

A todos os professores do Departamento de Física pelos ensinamentos transmitidos, pelo acolhimento e incentivo e, em especial ao Prof. Dr. João Ruggiero Neto pela co-orientação.

A todos os colegas do curso de Pós-graduação pelo companheirismo, em especial aos amigos Fábio Renato Lombardi, Fátima Pereira de Souza, Flávio Augusto Vicente Seixas, Gláucia R. Peglow Borges e Selma Anita Tamashiro pela valiosa colaboração em fases de acúmulo de trabalho e pela amizade.

Aos estagiários José Henrique Pereira e João Carlos Câmera Jr. pela colaboração na realização deste trabalho.

Aos funcionários do Departamento de Física por toda ajuda e atenção, especialmente a secretária Ilva pela amizade e paciência.

Aos funcionários da biblioteca pela presteza no serviço de comutação bibliográfica.

A Direção do Instituto pela disposição das condições materiais.
A Profa. Dra. Roseli A. S. Gomes e ao pessoal do Departamento de
Bioquímica e Biofísica da Faculdade de Medicina do Triângulo
Mineiro de Uberaba/MG pela colaboração na execução deste
trabalho.

A Andressa Salbé de Oliveira pelas aulas de inglês, tão
importantes, e pela amizade.

A Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP
pela ajuda financeira.

Ao Fundo Bunka de Pesquisa da Sociedade Brasileira de Cultura
Japonesa pelo auxílio concedido.

Aos meus pais Ana e Antonio e aos meus irmãos Marcelo e Marco,
pelo carinho e incentivo constantes.

**"Tens contigo os companheiros certos que te auxiliam no
aperfeiçoamento a que aspiras."**

**EMMANUEL
(Francisco Cândido Xavier)**

Índice

Resumo	01
Abstract	02
Introdução	03
Introdução às enzimas	04
A história da enzimologia	04
Biologia Estrutural	07
As enzimas são catalizadores biológicos	09
A eficiência catalítica deriva da ligação do reagente à enzima	10
Classificação e nomenclatura das enzimas	12
Introdução às proteinases aspárticas	16
Precusores	21
Uropepsina	24
Pepsina I e II	25
Interações com o ligante	27
Mecanismo catalítico	29
Transferência de energia	31
Mudança conformacional de enzimas	32
Objetivos	33
Materiais e Métodos	35
Purificação	36
Atividade Proteolítica	37
Cristalização	38
Aplicação do Método da Matriz Esparsa	39
Cristalização da uropepsina	42
Cristalização do complexo binário uropepsina:pepstatina	42
Cristalização da pepsina I	42
Eletroforese em gel de poliacrilamida SDS-PAGE	43

Coleta e processamento dos dados	43
Substituição Molecular	44
Refinamento da estrutura	47
Análise da estrutura	50
Modelagem do complexo binário uropepsina:pestatina	50
Resultados e Discussão	52
Atividade proteolítica	53
Cristalização da uropepsina	53
Cristalização do complexo binário uropepsina:pestatina	54
Cristalização da pepsina I	56
Coleta e processamento dos dados do cristal da uropepsina	57
Substituição Molecular	58
Refinamento da uropepsina	60
Análise da estrutura	64
Parâmetros de vibração térmica	69
Contatos intermoleculares e empacotamento do cristal	70
Comparação com outras proteinases aspárticas	74
Interação com pepstatina e sítio de ligação do substrato	76
Potencial eletrostático	78
Conclusões	82
Referências Bibliográficas	85
Apêndice I	94
Uma visão geral sobre cristalografia de macromoléculas biológicas	95
1. Cristalização de macromoléculas biológicas	95
1.1. Principais diferenças entre cristais de macromoléculas biológicas e cristais minerais	95



1.2. Técnicas de cristalização de macromoléculas biológicas	96
1.3. O método da matriz esparsa	100
1.4. Principais parâmetros a serem controlados num experimento de cristalização	102
1.4.1. Pureza da macromolécula biológica	102
1.4.2. Solubilidade e supersaturação	102
1.4.3. Nucleação, crescimento e cessação de crescimento	104
1.5. Solubilidade de macromoléculas biológicas	105
1.6. “Salting in” e “salting out”	107
1.7. Diagramas de fase de macromoléculas biológicas	110
1.8. Melhoria das condições de cristalização	113
2. Difração de raios X	114
2.1. Espalhamento Thomson	115
2.2. Espalhamento Compton	115
2.3. Espalhamento por elétrons	116
2.4. Fator de espalhamento atômico	117
2.5. Espalhamento de raios X por uma molécula	119
2.6. Espalhamento de raios X por um cristal	120
2.7. Densidade eletrônica	122
2.8. O problema da fase	124
3. Substituição molecular	125
4. Refinamento das coordenadas	130
4.1. “Simulated annealing”	132
5. Construção do modelo e validação	134
Apêndice II	138



Lista de Figuras

Figura 1. Estrutura tridimensional do RNA catalítico	06
Figura 2. Estrutura da mioglobina e da hemoglobina, primeiras proteínas estudadas Por cristalografia de difração de raios X	08
Figura 3. Na presença do catalisador, a reação ocorre através de um novo caminho	10
Figura 4. Mecanismo da hidrólise de um éster catalisada por H^+	11
Figura 5. Estrutura tridimensional da tripla hélice do colágeno	13
Figura 6. Enovelamento típico das proteinases aspárticas	18
Figura 7. Diagrama Ribbon da estrutura dimérica da proteinase do HIV	19
Figura 8: Diagrama esquemático das ligações de hidrogênio da pepstatina ligada à pepsina	20
Figura 9. Estrutura tridimensional do pepsinogênio	22
Figura 10. Sequência de aminoácidos N-terminal do pepsinogênio A5 humano	24
Figura 11. Molécula de pepstatina, inibidor de proteinases aspárticas	28
Figura 12. Visão tridimensional do sítio catalítico da pepsina em complexo com a Pepstatina	29
Figura 13: Mecanismo de ação das proteinases aspárticas	31
Figura 14. Principais passos envolvidos na resolução da estrutura de uma macromolécula biológica	38
Figura 15. Caixa para cultura de células usada para cristalização	40



Figura 16. Sequência de eventos para a montagem de uma gota de cristalização	40
Figura 17. Fotomicrografia do cristal de uropepsina	54
Figura 18. Fotomicrografia dos cristais de uropepsina na presença de pepstatina	55
Figura 19. Foto da eletroforese em gel de poliacrilamida a 12%	55
Figura 20. Fotomicrografia dos cristais de pepsina I	56
Figura 21. “Omit map” para a região em que existe a substituição na posição 291 (isoforma Leu-Val)	61
Figura 22. Gráfico mostrando o valor do R_{factor} e do R_{free} ao longo de todos os passos do refinamento da uropepsina	62
Figura 23. Mapas de densidade eletrônica na região do sítio ativo ao final de cada Etapa de refinamento	63
Figura 24. Mapa de densidade eletrônica do sítio ativo da uropepsina humana, mostrando os dois resíduos de ácido aspártico 32 e 215	64
Figura 25. Porção da cadeia polipeptídica indicando os ângulos ϕ e φ	65
Figura 26. Gráfico de Ramachandran para a uropepsina	66
Figura 27. Elementos de estrutura secundária da uropepsina mostrando átomos da cadeia principal	67
Figura 28. Gráfico mostrando o desvio de r.m.s. entre a uropepsina e a pepsina 3A humana. (a) todos os átomos da proteína; (b) carbonos α, (c) cadeia principal	68
Figura 29. Valores de fator B usando a cadeia principal e a cadeia lateral para os 326 resíduos de aminoácidos da uropepsina	70
Figura 30. Visão estéreo do empacotamento dos cristais de uropepsina e pepsina humana	72

Figura 31. Alinhamento da sequência primária da uropepsina comparada as outras proteinases aspárticas	75
Figura 32. Sítio de ligação da uropepsina com os resíduos de aminoácidos que participam da catálise enzimática	77
Figura 33. Potencial eletrostático da superfície da pepsina e da uropepsina	79
Figura 34. Potencial eletrostático da região do sítio catalítico da superfície da Pepsina e da uropepsina	80
Figura 35. Representação estéreo do sítio de ligação da uropepsina sobreposto ao sítio de ligação da pepsina 3A na presença da pepstatina	81

Lista de Tabelas

Tabela 1. As seis classes de enzimas e as reações que catalisam	14
Tabela 2: Distribuição das famílias de enzimas proteolíticas	17
Tabela 3. Lista das proteinases aspárticas usadas como modelos de busca para substituição molecular	47
Tabela 4. Estatísticas da coleta para ambos conjuntos de dados	57
Tabela 5. Detalhes estatísticos da coleta para os dados de difração de raios X no síncrotron	58
Tabela 6. Ângulos de Euler $\{\alpha, \beta, \gamma\}$ e coordenadas fracionárias $\{x, y, z\}$ após função de translação	59
Tabela 7. Coordenadas fracionárias após função de translação para outros possíveis grupos espaciais	59
Tabela 8. Valores de R_{factor} e R_{free} para cada etapa do refinamento da uropepsina	60
Tabela 9. Estatísticas do refinamento para uropepsina humana	66
Tabela 10. Contatos por ligações de hidrogênio entre uropepsina humana e seus quatorze vizinhos relacionados por simetria	71
Tabela 11. Contatos por ligações de hidrogênio entre pepsina humana e seus Vinte e dois vizinhos relacionados por simetria	73
Tabela 12. Valores dos desvios de r.m.s. para a sobreposição da uropepsina com outras proteinases aspárticas	74
Tabela 13. Ligações de hidrogênio intermoleculares da pepsina e da uropepsina com a pepstatina	77
Tabela 14. Resíduos envolvidos nos sushsios de especificidade da uropepsina humana	80

Lista de Abreviaturas

Abz – ácido benzóico

APS – persulfato de amônio

Bisacrilamida – N, N' - metileno bis-acrilamida

CC – coeficiente de correlação

CCP4 – “Colaborative computational project 4”

DEAE – dietil aminoetil

Eddnp – etilenodiaminodinitrofenol

F_{obs} – fator de estrutura observado

F_{calc} – fator de estrutura calculado

FPLC – “Flow Performance Liquid Chromatography”

Hic-up – “Hetero-compound Information Center – Uppsala”

HIV – vírus da imunodeficiência humana

K_{cat} – constante catalítica

LNLS – Laboratório Nacional de Luz Síncrotron

PAGE – eletroforese em gel de poli(acrilamida

PDB – “Protein Data Bank”

PEG – Polietileno glicol

PGA – pepsinogênio A

PGC – pepsinogênio C

r.m.s. – “root mean square”

RNA – ácido ribonucleico

SDS – dodecilsulfato de sódio

Sta – ácido 4-amino-3-hidroxi-6-metil heptanóico

TEMED – N, N, N', N' - tetrametiletlenodiamina



Lista de programas utilizados

AMoRe – “Automatic Molecular Replacement”

CNS – “Crystallography & NMR System: A new software suite for macromolecular structure determination”

DENZO – programa para processamento de dados de difração de raios X

Distang – programa que calcula contatos inter e intramoleculares

GRASP – programa para visualização da superfície e do potencial eletrostático e da macromolécula biológica

LSQKAB – programa que faz sobreposições de macromoléculas

O – programa para visualização gráfica

PROCESS – programa para processamento de dados de difração de raios X

PROCHECK – programa que analisa a estrutura secundária, graficando os ângulos ϕ contra os ângulos ψ do modelo entre outros parâmetros estereo-químicos

PROFIT – “Protein Least Square Fitting” – programa para sobreposições de macromoléculas biológicas

SCALEPACK – programa para processamento de dados de difração de raios X

X-PLOR – pacote de programas para cristalografia e RMN

XtalView – programa para visualização gráfica



Prefácio

Aqui está reunido todas as atividades realizadas em meu doutoramento. Este trabalho está dividido em 9 partes, sendo a primeira e segunda partes, o Resumo e *Abstract*, respectivamente. A terceira parte consta da Introdução, onde realizei um levantamento bibliográfico, abrangendo trabalhos de até 50 anos atrás, mas que contém informações importantes e atuais, e muitos trabalhos realizados recentemente nesta área de pesquisa. Portanto, a introdução reúne muitos dados sobre proteinases aspárticas e mais especificamente sobre a pepsina, destacando sua origem, função e importância. Em Materiais e Métodos descrevi todas as etapas do trabalho, com relação às técnicas usadas na elaboração do trabalho. É importante frisar que grande parte da literatura dos métodos utilizados estão reunidos no final do trabalho, numa seção denominada Apêndice. Em Resultados e Discussão estão expostos todos os dados obtidos nestes três anos, incluindo os resultados positivos e negativos do trabalho, e uma discussão sobre os dados apresentados, finalizando com as conclusões e referências utilizadas na pesquisa bibliográfica.



Resumo

As enzimas proteolíticas são importantes ferramentas na análise seqüencial de proteínas e na identificação e isolamento de domínios de outros complexos protéicos multifuncionais. Com o advento da biologia molecular, uma nova dimensão tem sido adicionada ao estudo de proteinases. Esta nova fonte de informações têm contribuído significativamente para o entendimento de domínios organizacionais, funcionais e evolutivos de proteinases reguladoras.

Três proteinases aspárticas foram cristalizadas e a estrutura molecular da uropepsina humana, uma proteinase encontrada na urina e produzida na forma de pepsinogênio A na mucosa gástrica, foi determinada por substituição molecular usando a pepsina humana como modelo de busca. Os cristais de uropepsina pertencem ao grupo espacial $P2_12_12_1$ com parâmetros de cela de $a=50.99$, $b=75.56$ and $c=89.90$ Å. O refinamento cristalográfico resultou em um R_{factor} de 0,161 a 2,45 Å de resolução. As posições dos 2437 átomos da proteína em 326 resíduos foram determinados e o modelo contém 143 moléculas de água. A estrutura é bilobal, consistindo, predominantemente de 2 folhas β as quais, como observado em outras proteinases aspárticas, estão relacionadas por um eixo binário. Tentativas de cristalizar a uropepsina na presença do inibidor pepstatina não resultaram em bons cristais, por esta razão, um modelo da uropepsina complexada foi construído, baseado na estrutura da pepsina humana complexada com pepstatina resolvida à alta resolução.



Abstract

Proteolytic enzymes are important tools in the sequential analysis of proteins and in the identification of domains from others multifunctional complexes of proteins. Molecular biology opens a new dimension about the study of proteinases. This new information source has contributed significantly for the understanding of organizational, functional and evolutive domains from regulative proteinases.

Three aspartic proteinases have been crystallized and the molecular structure of human uropepsin, an aspartic proteinase from the urine, produced in the form of pepsinogen A in the gastric mucosa, has been determined by molecular replacement using human pepsin as the search model. Crystals belong to the space group $P2_12_12_1$ with the cell parameters $a=50.99$, $b=75.56$ and $c=89.90$ Å. Crystallographic refinement led to an R_{factor} of 0.161 at 2.45 Å resolution. The positions of 2437 non-hydrogen protein atoms in 326 residues have been determined and the model contains 143 water molecules. The structure is bilobal, consisting of two predominantly β -sheet lobes which, as observed in other aspartic proteinases, are related by a pseudo 2-fold axis. Attempts to crystallize the uropepsin in presence of the pepstatin inhibitor does not result in good crystals, thus a model of the uropepsin complex has been constructed based on the high resolution crystal structure of pepsin complexed with pepstatin.

Abstract

Introdução

Introdução às enzimas	04
A história da enzimologia	04
Biologia estrutural	07
As enzimas são catalisadores biológicos	09
A eficiência catalítica deriva da ligação do reagente à enzima	10
Classificação e nomenclatura das enzimas	12
Introdução às proteinases aspárticas	16
Precusores	21
Uropepsina	24
Pepsina I e II	25
Interações com o ligante	27
Mecanismo catalítico	29
Transferência de energia	31
Mudança conformacional de enzimas	32

INTRODUÇÃO ÀS ENZIMAS

A história da enzimologia

Grande parte da história da bioquímica é a história da pesquisa sobre as enzimas. O primeiro reconhecimento que alguma forma de catálise toma parte em sistemas biológicos ocorreu nos primeiros anos do século XIX, a partir de estudos sobre a digestão da carne pelas secreções estomacais e da transformação do amido em glicose pela saliva e por vários extratos de plantas. Subseqüentemente, um grande número de exemplos de catálise biológica, hoje reconhecidos como enzimáticos, foram descritos (Lehninger, 1985). Em 1810, pesquisas sobre fermentação foram realizadas por Joseph Gay-Lussac, que descobriu que o etanol e o CO_2 eram os principais produtos da decomposição da glicose, pelo lêvedo (Voet & Voet, 1995).

Louis Pasteur, na década de 1850, concluiu que esta reação é catalisada por “fermentos”. Ele postulou que esses fermentos eram inseparáveis da estrutura das células vivas do lêvedo, uma opinião que prevaleceu por muitos anos. De fato, o nome “enzima” (do grego: *en*, in + *zima*, lêvedo) foi inventado em 1878 por Fredich Wilhelm Kuhne num esforço para enfatizar que ‘existe algo no lêvedo’, em oposição aos lêvedos em si que catalisam as reações de fermentação. Foi, portanto, um grande marco na história da bioquímica quando, em 1897 Eduard Buchner teve sucesso em extrair do lêvedo, na forma solúvel ativa, o conjunto de enzimas que catalisa a fermentação da glicose em álcool. Esta descoberta provou que este

importante conjunto de enzimas que catalisa uma grande via metabólica liberadora de energia, pode continuar funcionando depois de removida da estrutura de células vivas. Isto encorajou os bioquímicos a tentar o isolamento de muitas enzimas diferentes e, subseqüentemente, estudar suas propriedades catalíticas (Voet & Voet, 1995; Lehninger, 1985).

No início do século XX, Emil Fisher realizou os primeiros estudos sistemáticos sobre especificidade enzimática. Outros estudaram a cinética da atividade das enzimas e formularam teorias sobre seu mecanismo de ação. Mas, em relação à composição química das enzimas, nada se sabia até o início do século 20. Foi em 1926, na Universidade Cornell, que James Sumner cristalizou a primeira enzima, a urease, obtida de extratos do feijão de soja. Sumner constatou que os cristais de urease, a qual catalisa a hidrólise da uréia em NH_3 e CO_2 , eram formados inteiramente por proteínas. Isto levou-o a postular que todas as enzimas são proteínas. Suas conclusões sofreram vigorosa oposição do renomado bioquímico alemão Richard Willstätter, que insistiu em ser as enzimas compostos de baixo peso molecular e que as proteínas dos cristais da urease eram apenas contaminantes. Apenas na década de trinta, após John Northrop e Moses Kunitz cristalizarem pepsina e tripsina, e determinarem que elas também eram proteínas, que foi amplamente aceita a natureza protéica das enzimas. Mas as proteínas não são as únicas macromoléculas biológicas capazes de realizar catálise. A descoberta da atividade catalítica do RNA causou um profundo impacto no modo de pensar dos bioquímicos. Os primeiros RNAs catalíticos (as ribozimas) descobertos incluíam os que catalisam seu próprio processamento. Foi demonstrado que os

RNAs também podem catalisar reações envolvidas na síntese de proteínas. A eficiência catalítica dos RNAs catalíticos é menor do que a das enzimas protéicas, e a eficiência catalítica dos sistemas atualmente existentes de RNA é grandemente aumentada em presença de subunidades de proteínas junto com o RNA (Campbell, 2000). A figura 1 mostra a estrutura tridimensional do RNA catalítico em complexo com um ligante.

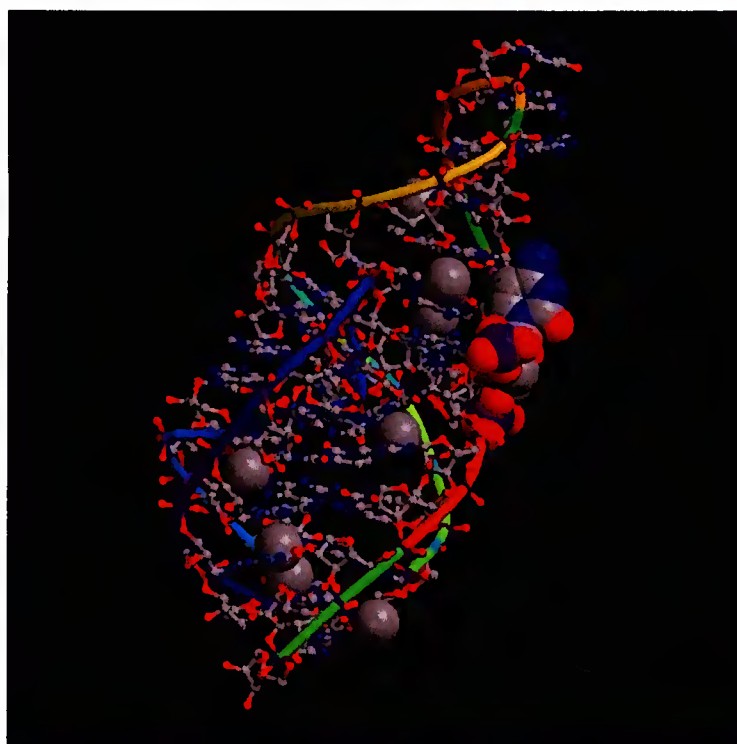


Figura 1. Estrutura tridimensional do RNA catalítico em complexo com um ligante.

Não foi antes de 1963 que a primeira seqüência de aminoácidos de uma enzima, a ribonuclease A pancreática bovina foi reportada em sua totalidade. Seguiu-se então, um período de intensas pesquisas sobre as enzimas catalisadoras das reações do metabolismo celular.

Biologia Estrutural

Nos últimos anos as ciências experimentais, valendo-se de suas técnicas tradicionais, têm se aproximado cada vez mais de seus limites. Esgotadas as possibilidades, pesquisadores tem procurado cada vez mais abordagens interdisciplinares, de forma a extrair novas informações que lhes permitirão entender melhor o fenômeno estudado. Uma das novas frentes interdisciplinares é a biologia molecular estrutural. Lançando-se mão de novas técnicas físico-químicas e computacionais, os biólogos estruturais estão obtendo informações mais precisas sobre a real engenharia da vida, através das proteínas. Estas desempenham papéis de interface com o ambiente, podendo ser tanto ferramentas quanto blocos construtores de diversos sistemas biológicos.

A biologia molecular como conhecemos até recentemente se resumia a descrever, decifrar e entender as informações codificadas nos genes, uma etapa de suma importância, mas que necessitava de continuidade. Fez-se imperativo compreender como os produtos gênicos se interagem com o meio ambiente e os mecanismos que tornam isso possível. Em outras palavras, descrever para proteínas e enzimas as suas funções específicas, e os mecanismos através dos quais determinada função é exercida.

Uma rica fonte de informações acerca da função exata de enzimas e proteínas está na caracterização da posição de todos os átomos que as compõem, através da correlação entre estrutura e função. Cada átomo desempenha um papel importante no mecanismo geral e, ao mapearmos com precisão estes detalhes,

poderemos associar a dados de outras fontes e especular com maior precisão sobre o seu funcionamento.

Uma das mais poderosas técnicas para a obtenção de dados dessa natureza é a cristalografia de proteínas, que começou na década de 50 com os primeiros estudos sobre difração de raios X por cristais de mioglobina e hemoglobina, em experimentos conduzidos por Perutz e Kendrew, laureados com o Prêmio Nobel de Química do ano de 1962. A figura 2 mostra as estruturas cristalográficas da mioglobina e da hemoglobina.

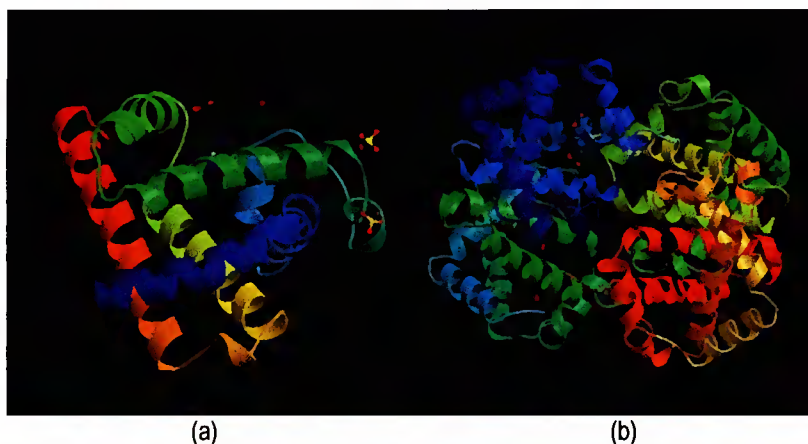


Figura 2. Estruturas da mioglobina (a) e da hemoglobina (b), primeiras proteínas estudadas por cristalografia de difração de raios X.

A partir de então, a cristalografia de proteínas se viu em franca expansão, principalmente nos últimos dez anos. Este crescimento é resultado do desenvolvimento de novas técnicas, tais como fontes síncrotrons, possibilitadas pelo crescimento científico em diversas áreas e também pelo contínuo aumento do poder computacional a preços acessíveis, uma vez que grande parte do trabalho depende de cálculos numéricos exaustivos.

As enzimas são catalisadores biológicos

A grande variedade de reações bioquímicas que compreende a vida são quase todas mediadas por uma série de catalisadores biológicos conhecidos como enzimas. Embora as enzimas estejam sujeitas às mesmas leis da natureza que governam o comportamento de outras substâncias, elas se diferem dos catalisadores químicos em relação a muitos aspectos importantes (Lehninger, 1985):

1. Maior velocidade das reações: reações enzimaticamente catalisadas necessitam de menor energia de ativação e são da ordem de 10^6 a 10^{12} vezes mais rápidas do que reações não-catalisadas, e são muitas ordens de grandeza mais eficientes do que reações quimicamente catalisadas.
2. Reações ocorrem em condições moderadas: reações enzimaticamente catalisadas ocorrem em condições moderadas, ou seja, temperatura abaixo de 100°C , pressão atmosférica e pH's próximos à neutralidade. Em contraste, uma catálise química eficiente muitas vezes requer temperatura e pressão elevadas, tanto quanto valores extremos de pH.
3. Reações de maior especificidade: como as enzimas possuem alto grau de especificidade, ocorrerão em uma célula, dentre todas as reações potencialmente possíveis entre as numerosas substâncias presentes, apenas aquelas para as quais a célula possua enzimas adequadas.
4. Capacidade de regulação: as atividades catalíticas de muitas enzimas variam em resposta à concentração de outras substâncias, diferentes de seus substratos. Os mecanismos destes processos regulatórios incluem controle alostérico, modificação



covalente de enzimas e variação da quantidade de enzimas sintetizadas. Portanto, sua concentração celular e sua atividade podem ser reguladas, permitindo o ajuste a diferentes condições fisiológicas.

Resumindo, as enzimas (a) diminuem a energia de ativação, levando a altas velocidades de reação (Figura 3), (b) são muito específicas, (c) são sintetizadas pelas próprias células e (d) têm concentração e atividade moduláveis, permitindo um ajuste fino do metabolismo ao ambiente celular. O conjunto destes aspectos favoráveis possibilita a manutenção da vida, justificando o alto investimento energético necessário para a síntese de enzimas. Considerando estas propriedades das enzimas, algumas das questões centrais da bioquímica são levantadas: Como as enzimas trabalham? Como os aminoácidos, por si só incapazes de acelerar reações químicas, podem produzir tal força catalítica quando unidos em seqüências específicas? (Lehninger, 1985; Marzzoco & Torres, 1999).

A eficiência da catálise enzimática deriva da ligação do reagente (substrato) à enzima

Os catalisadores criam um novo caminho para a reação, com um novo estado de transição, que requer energia de ativação menor (Figura 3).

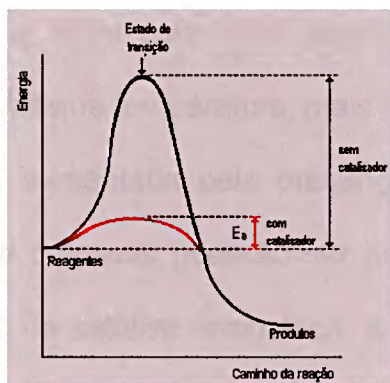


Figura 3. Na presença do catalisador, a reação ocorre através de um novo caminho com energia de ativação E_a menor (Marzzoco & Torres, 1999).

Um exemplo simples deste novo caminho é apresentado na Figura 4, onde a hidrólise de um éster é catalisada por íons H^+ .

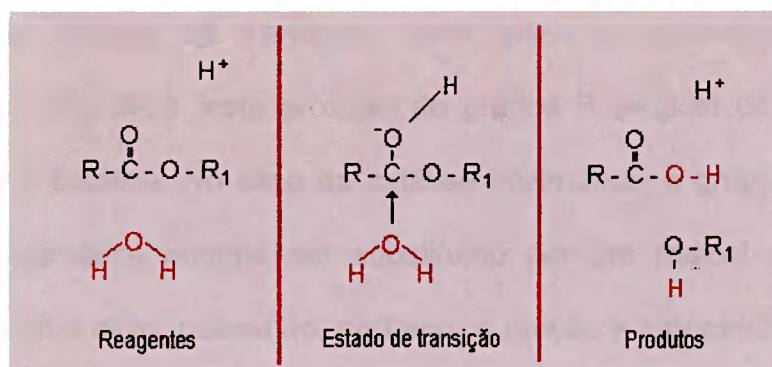


Figura 4. Mecanismo da hidrólise de um éster catalisada por H^+ . A presença do íon altera a distribuição de cargas elétricas do éster, criando um caminho de reação que necessita de energia de ativação menor do que o da reação não catalisada (Marzzoco & Torres, 1999).

A reação consiste no ataque do oxigênio (que tem carga residual negativa), pertencente à molécula de água, ao carbono presente no éster (que tem carga residual positiva, em virtude de sua dupla ligação com o oxigênio). A energia de ativação requerida para atingir o estado de transição é alta. A presença dos íons H^+ cria um caminho alternativo para a reação: o íon H^+ liga-se ao oxigênio presente no éster, aumentando a carga positiva do carbono e tornando-o mais susceptível ao ataque do oxigênio da água. Para este novo caminho, a energia necessária é menor e, portanto, em uma mesma temperatura, mais moléculas poderão reagir e a velocidade da reação será aumentada pela presença de H^+ . Seguindo modelo semelhante, muitas reações químicas poderão ser aceleradas por íons OH^- , por íons de metais etc. No caso da catálise enzimática, a aproximação e a ligação do substrato à enzima altera o delicado balanço de forças responsáveis pela

manutenção da estrutura tridimensional da enzima, “moldando” sua forma à forma do substrato e fazendo-a adquirir uma nova configuração, ideal para a catálise. Os substratos têm sua molécula tensionada e distorcida, aproximando-se da conformação do estado de transição. Além disso, o substrato, corretamente posicionado no sítio ativo, está próximo de grupos R (radical ou cadeia lateral) decisivos para a catálise. No caso da catálise enzimática, o grupo positivo H^+ da catálise não-enzimática poderia ser substituído por um radical positivo de um aminoácido do sítio ativo, passando, portanto, a reação a independe dos choques casuais entre as moléculas dos reagentes. É este conjunto de mecanismos que torna a catálise enzimática tão eficiente, muito mais eficiente do que a catálise inorgânica (Marzzoco & Torres, 1999).

CLASSIFICAÇÃO E NOMENCLATURA DAS ENZIMAS

Pelas regras oficiais de classificação e nomenclatura, as enzimas são divididas em seis grupos, de acordo com o tipo de reação de catalisam (Tabela 1). Cada um destes grupos é ainda subdividido em classes e subclasses, numeradas de tal forma que cada enzima possa ser identificada sem ambigüidade. Assim, por exemplo, a enzima que catalisa a remoção de elétrons do etanol (portanto, uma óxido-redutase) é designada *álcool:NAD⁺:óxido-redutase* e recebe o número de classificação EC 1.1.1.1 (EC de *Enzyme Commission*). Em muitos casos, nomes mais simples foram adotados e se tornaram clássicos. A enzima que catalisa a oxidação do etanol é comumente referida como álcool desidrogenase; a enzima que catalisa a síntese de glicogênio (oficialmente designada *UDPglicose:glicogênio 4- α -D-*

glicosiltransferase) é chamada glicogênio sintase. Portanto, na nomenclatura usual, o nome é dado indicando o substrato, seguido de uma outra palavra terminada em *ase* que especifica o tipo de reação que a enzima catalisa. Mesmo esta forma simplificada de nomenclatura apresenta exceções, como é o caso das enzimas digestivas: pepsina, tripsina etc., cujos nomes triviais tornaram-se clássicos (Marzzoco & Torres, 1999).

As enzimas, na sua grande maioria, são proteínas globulares. Nestas, a cadeia polipeptídica dobra-se várias vezes, resultando em uma estrutura globular compacta. Como classe, elas têm conformação mais complexa do que as proteínas fibrosas (Figura 5), com uma variedade de funções biológicas muito maior e atividades muito mais dinâmicas que estáticas.

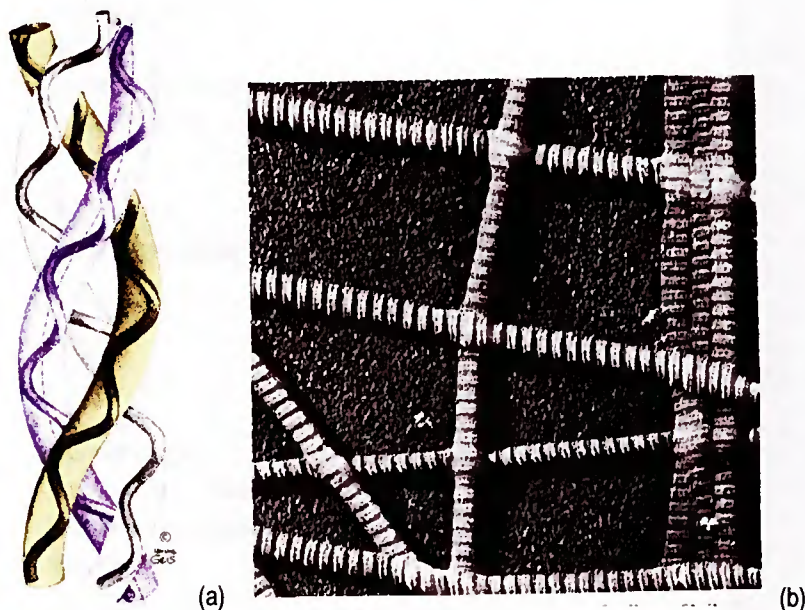
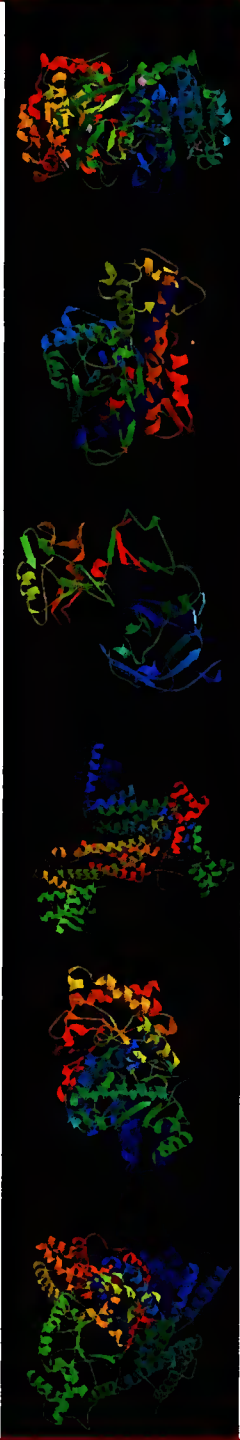


Figura 5. (a) Estrutura tridimensional da tripla hélice do colágeno e (b) micrografia eletrônica das fibras de colágeno da pele (Voet & Voet, 1995).

Tabela 1. As seis classes de enzimas e as reações que catalisam (Marzzoco & Torres, 1999).

Classe	Tipo de reação catalisada	Exemplo	Enovelamento
1. Óxido-redutases	Transferência de elétrons	Álcool desidrogenase	
2. Transferases	Reações de transferência de grupos	Hexoquinase	
3. Hidrolases	Reações de hidrólise	Pepsina	
4. Liases	Adição de grupos a duplas ligações ou remoção de grupos, deixando dupla ligação	Fumarase	
5. Isomerases	Rearranjos intramoleculares	Fosfoglicoisomerase	
6. Ligases	Condensação de duas moléculas, associada à hidrólise de uma ligação de alta energia (em geral, do ATP)	Piruvato carboxilase	

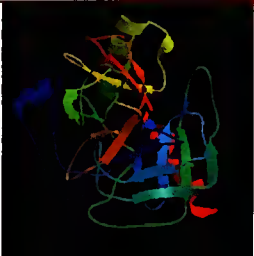
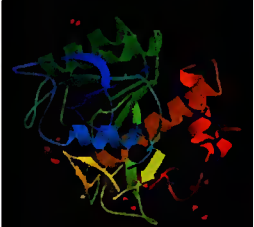
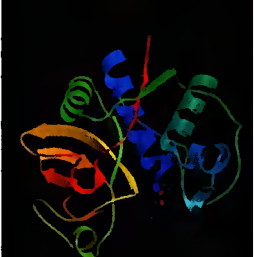
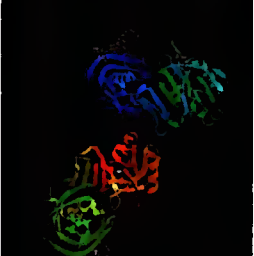
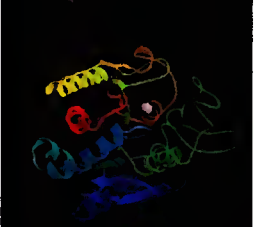
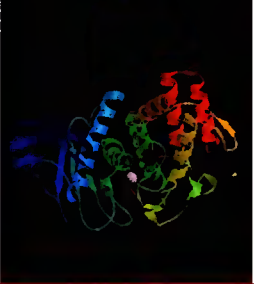
A cadeia polipeptídica de uma proteína é determinada pela seqüência de aminoácidos. Enquanto a estrutura secundária é determinada pela interação de grupos R próximos entre si, a estrutura terciária é conferida por aspectos da interação a longa distância na seqüência de resíduos de aminoácidos. A formação de curvas na cadeia polipeptídica e a direção e angulação destas curvas, são todas determinadas pela localização de resíduos específicos formadores da curvatura, como a prolina, a treonina e a serina. Muitos dos resíduos são invariáveis em proteínas homólogas. Resíduos que estão sempre presentes em posições específicas, independentemente da espécie, ao longo da cadeia polipeptídica. Alguns ocorrem nas curvaturas da cadeia ou próximo a elas; outros, como os resíduos de cisteína, nos pontos de interligação de alças da estrutura terciária. Outros ainda, estão presentes em locais específicos, como o sítio ativo de enzimas (Lehninger, 1985).

Muitas informações importantes sobre a estrutura e mecanismo catalítico das enzimas têm sido obtidas por análises de difração de raios X. As enzimas, em sua forma cristalina, têm sido submetidas à análise estrutural com raios X. Estes estudos têm sido frequentemente correlacionados com estudos químicos de (1) seqüência de aminoácidos, (2) especificidade para o substrato, (3) ação de inibidores específicos e, (4) identificação de grupos funcionais específicos no sítio catalítico. Desta maneira, representantes da maioria das mais importantes classes de enzimas foram examinados, num esforço para determinar as possíveis relações entre a ação catalítica e a estrutura tridimensional das enzimas.

INTRODUÇÃO ÀS PROTEINASES ASPÁRTICAS

As enzimas proteolíticas são classificadas baseando-se em sua função, especificidade, estrutura tridimensional e na comparação dos sítios ativos, de acordo com o seu mecanismo catalítico, sendo separadas em quatro classes de mecanismos e entre estas classes, seis famílias de proteinases: serino-proteinases I e II, metaloproteinases I e II, cisteíno-proteinases e proteinases aspárticas (Kay, 1982). Cada família tem sua característica de resíduos de aminoácidos funcionais arranjados numa configuração específica no sítio ativo (Tabela 2).

Tabela 2. Distribuição das famílias de enzimas proteolíticas (Kay, 1982).

Família	Proteinases Representativas	Resíduos de aminoácidos característicos do centro ativo*	Enovelamento
Serino-Proteinases I	Quimotripsina Tripsina Trombina Elastase Calicreína pancreática	Asp102, Ser195, His57	
Serino-Proteinases II	Subtilisina	Asp32, Ser221, His64	
Cisteíno-Proteinases	Papaina Actinidina Catepsinas B e H	Cys25, His159, Asp158	
Proteinases Aspárticas	Penicilopepsina Proteinases ácidas Renina Catepsina D e E	Asp33, Asp231	
Metaloproteinases I	Carboxipeptidase A bovina	Zn, Glu270, Tyr248	
Metaloproteinases II	Termolisina	Zn, Glu143, His231	

* O número de resíduos correspondem à sequência de aminoácidos das enzimas listadas em negrito na coluna 2.

A família das proteinases aspárticas, alvo de nossos estudos, inclui proteínas de mamíferos, de plantas, de fungos e de vírus. Em mamíferos, estas enzimas executam funções diversas que variam da digestão de proteínas alimentares a papéis regulatórios muito específicos. Estas proteinases contêm três regiões distintas: um domínio terminal N, um domínio terminal C, e um interdomínio composto por folhas β antiparalelas do terminal N. A região de ligação dos dois domínios, terminal N e terminal C, contribuem cada um, com um ácido aspártico para o estabelecimento do sítio ativo (Baldwin *et al.*, 1993). A figura 6 ilustra o enovelamento típico das proteinases aspárticas.

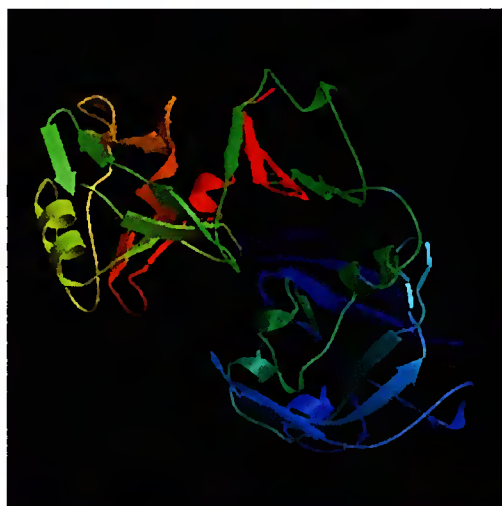


Figura 6. Enovelamento típico das proteinases aspárticas (Cód. PDB: 1PSN). Podemos observar a cavidade que leva ao sítio catalítico na região inferior esquerda da figura.

Estas proteinases participam de uma variedade de processos fisiológicos, e alterações ao nível de atividade expressa podem estar associadas ao início de condições patológicas, tais como hipertensão, úlceras gástricas e distrofia

muscular. Quanto à sua aplicação biotecnológica, as proteinases aspárticas são usadas em duas grandes áreas de processamento de alimentos – na coagulação do leite e na produção de alimentos fermentados a partir de feijão de soja, arroz e cereais. Apresentam, portanto, grande importância econômica nesta área (Green *et al.*, 1984).

Algumas proteinases aspárticas já foram resolvidas por técnicas de difração de raios X. A penicilopepsina (James & Sielecki, 1983), a proteinase de *Endothia* (Blundell *et al.*, 1984), a proteinase de *Rhizopus chinensis* (Bott *et al.*, 1982), a pepsina de porco (Andreeva *et al.*, 1984), a proteinase do vírus HIV (Sham *et al.*, 1996) e a pepsina humana (Fujinaga *et al.*, 1995) foram resolvidas e refinadas a alta resolução (1,8 a 2,1Å). Todas apresentam similaridade na estrutura tridimensional (Blundell *et al.*, 1984; James & Sielecki, 1983), com exceção da proteinase viral, que sofre dimerização para o estabelecimento do sítio ativo, conservando ainda as características do sítio ativo (Figura 7). Portanto, em cada enzima há distinções na estrutura e, conseqüentemente, na função fisiológica específica em seu próprio meio.

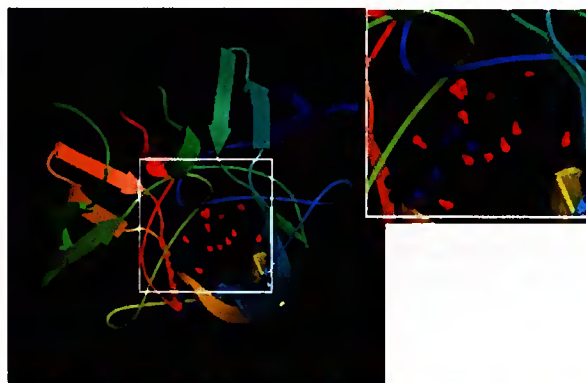


Figura 7. Diagrama ribbon da estrutura dimérica da proteinase aspártica do vírus da imunodeficiência humana (HIV) na presença de um inibidor ligado ao sítio catalítico (em destaque). Os dois resíduos de ácido aspárticos estão localizados nas alças em azul e verde do destaque.

Existem resíduos que contribuem para os sítios de especificidade nas diferentes enzimas da família. Esta especificidade é caracterizada pelos resíduos de aminoácidos adjacentes à ligação peptídica a qual será hidrolisada, mas a ligação do substrato e a especificidade podem envolver resíduos de aminoácidos, os quais estão localizados em posições de 2 a 4 resíduos, distantes da ligação peptídica a ser clivada. Nas proteinases aspárticas com sítios de ligação estendidos, os subsítios de ligação (S) e as posições correspondentes dos resíduos de aminoácidos (P) são designados como mostrado na figura 8 (Foltmann, 1981).

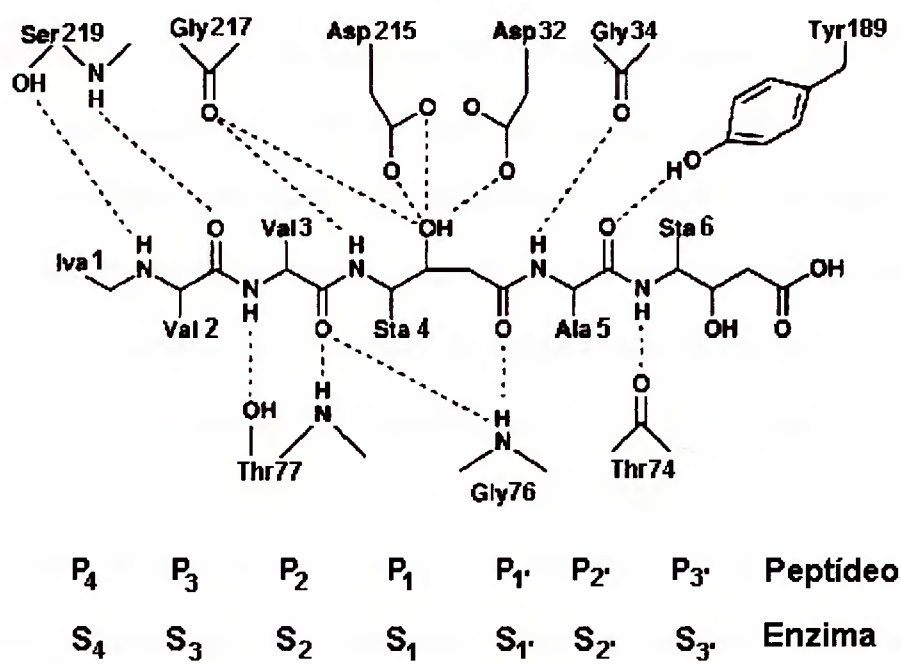


Figura 8. Diagrama esquemático das ligações de hidrogênio entre o inibidor pepstatina e os subsítios da pepsina gástrica. A cadeia lateral do inibidor é numerada de acordo com a nomenclatura de Schechter & Berger (1967), onde P_n designa a cadeia lateral do resíduo do peptídeo inibidor que interage com o correspondente subsítio S_n do sítio ativo da enzima.

Todas as proteinases aspárticas possuem uma larga cavidade no sítio ativo (Tang *et al.*, 1978) a qual pode acomodar até 7 resíduos de aminoácidos de um substrato nos subsítios $S_4 - S_3'$ com uma distinta preferência para clivar a ligação entre resíduos hidrofóbicos (Powers *et al.*, 1977). Diferenças significantes nos tipos de interações são observadas nos subsítios S_4 , S_3 , S_2 , S_1 , S_1' , S_2' e S_3' , os quais podem ser responsáveis pelas diferenças na especificidade. Os subsítios S_1' e S_1 são responsáveis pela hidrólise da ligação peptídica.

Precursores

Algumas enzimas, cujo local de ação é extracelular (plasma, trato digestivo) são sintetizadas na forma de precursores inativos chamados zimogênios. Todas as proteinases aspárticas de mamíferos são produzidas na forma de zimogênio. Para que um zimogênio adquira as propriedades funcionais, é necessário que haja hidrólise de determinadas ligações peptídicas, com a consequente remoção de um segmento da cadeia de aminoácidos. A cadeia polipeptídica remanescente adquire nova estrutura espacial, onde é organizado um centro ativo funcional.

A pepsina humana possui sete diferentes isoformas de zimogênios, denominados pepsinogênios (Figura 9) (Samloff, 1969; Foltmann, 1981). Estas isoformas são enzimas que apresentam a mesma seqüência primária da pepsina, com exceção de um único aminoácido, em determinada posição. Estes foram divididos em dois tipos: pepsinogênio I (pepsinogênio A – PGA), o qual consiste de 5 isoformas (PGA1-5) e pepsinogênio II (pepsinogênio C – PGC), o qual consiste de 2 isoformas (PGC6 e 7).

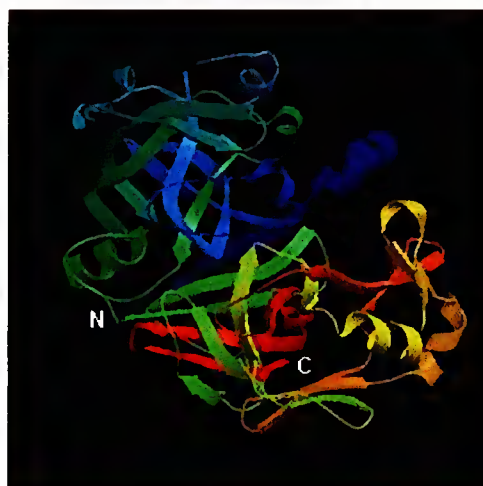


Figura 9. Estrutura tridimensional do pepsinogênio, zimogênio da pepsina, destacando o N-terminal em azul, que é removido para dar origem à forma ativa.

Seijffers *et al.*, (1963), Kushner *et al.* (1964) e Hirsch-Marie, (1968) realizaram experimentos usando cromatografia de troca iônica e eletroforese em gel de agarose e sugeriram que PGA consistiam de dois zimogênios que diferiam na carga elétrica, mas estudos realizados por Samloff, (1971) usando uma técnica eletroforética, identificaram 5 isoformas de PGA, em extratos de mucosa gástrica. A nomenclatura foi designada em ordem da diminuição da mobilidade eletroforética. O pepsinogênio A5 foi subsequentemente identificado na urina (Gedde-Dahl *et al.*, 1978).

Há um mecanismo no qual pepsinogênio é transportado das células pépticas para a circulação. Ambos, pepsinogênio A e C estão presentes no soro (Samloff, 1971; Samloff, 1982, Ichinose *et al.*, 1982). Em contraste, a urina normalmente contém apenas pepsinogênio A (Seijffers *et al.*, 1964, Samloff & Townes, 1970; Hirsch-Marie & Conte, 1969) e o fluido seminal contém apenas pepsinogênio C (Hirsch-Marie, 1968; Seijffers *et al.*, 1965; Samloff & Liebman, 1972).

Estudos realizados por Samloff & Townes, (1970) revelaram que alguns indivíduos não possuem PGA5 em sua urina, e que esta ausência é associada com a ausência deste pepsinogênio na mucosa gástrica. Inversamente, indivíduos com PGA5 em sua urina também o possuem na mucosa gástrica. Existem também diferenças com relação às relativas concentrações das isoformas de PGA em ambos, urina e mucosa gástrica (Samloff & Townes, 1970; Taggart *et al.*, 1979). Esta variação mostrou ser um traço genético (Korsnes & Gedde-Dahl, 1980; Taggart *et al.*, 1978). Taggart *et al.*, (1979) identificaram 9 fenótipos na população dos Estados Unidos. Entretanto, fenótipos adicionais foram encontrados em populações européias (Korsnes & Gedde-Dahl, 1980). Frants *et al.*, (1984) propuseram que PGA3, 4 e 5 são codificados por genes separados e que a concentração de cada um é determinado principalmente pelo número de cópias do gene.

O primeiro passo na ativação do pepsinogênio consiste na sua mudança conformacional, na qual o sítio ativo é descoberto (James & Sielecki, 1986). Pepsinogênios são ativados pela liberação de 42 a 47 resíduos de seu N-terminal. Dois sítios altamente susceptíveis à clivagem estão presentes no segmento de ativação. Um deles localiza-se no final do segmento de ativação (por exemplo, entre os resíduos 47 e 48), denominado sítio P, e o outro localiza-se na região central do segmento de ativação (sítio I). Quando o sítio P é clivado primeiro, o pepsinogênio é diretamente convertido à pepsina, com a liberação do segmento de ativação intacto (Kageyama & Takahashi, 1982). Por outro lado, quando o sítio I é o primeiro a ser clivado, uma forma intermediária é gerada. Neste caso, clivagens adicionais são

necessárias para completar a ativação. A figura 10 ilustra o mecanismo de ativação do pepsinogênio A5.

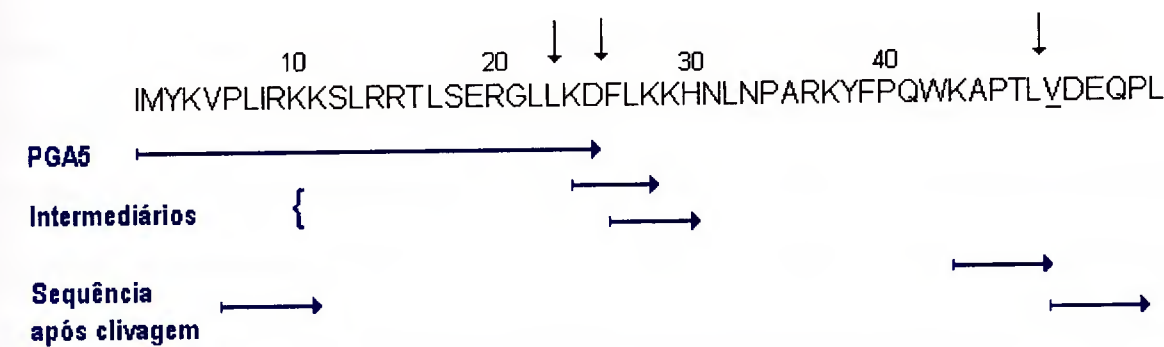


Figura 10. Sequência de aminoácidos N-terminal do pepsinogênio A5 humano. Setas verticais indicam pontos de clivagem durante a ativação (sítios I e P, na sequência). O N-terminal da pepsina é a Val (em grifo) (Foltmann, 1988).

Portanto, reações intramolecular e intermolecular, estão envolvidas nestas clivagens (McPhie, 1972). As intramoleculares são importantes, funcionando como uma reação inicializadora para gerar a molécula ativa, e reações intermoleculares são essenciais para acelerar e completar a ativação, sendo estas reações realizadas pela própria pepsina ativa (Kageyama & Takahashi, 1987).

Uropepsina

Proteínas de baixo peso molecular podem ser filtradas pelos rins. O peso molecular limite é de cerca de 80kDa. Portanto, moléculas de proteínas pequenas (<80kDa) podem fazer parte da urina. Apesar da presença de atividade enzimática na urina ser conhecida a cerca de 160 anos, suas propriedades físico-químicas e enzimáticas não são ainda muito elucidadas. De uma maneira geral, estas enzimas estão presentes na urina em quantidades relativamente pequenas, mas ainda assim

podem ser a única fonte potencial de enzimas humanas (Raab, 1972, Minamiura *et al.*, 1984).

A excreção do uropepsinogênio depende da presença do pepsinogênio no soro, o qual está diretamente relacionado à secreção pelas glândulas gástricas. A remoção de 42 resíduos de aminoácidos do N-terminal catalisa a conversão do uropepsinogênio em uropepsina. A uropepsina trata-se de uma isoforma da pepsina gástrica, apresentando apenas uma modificação na posição 291, onde uma leucina foi substituída por uma valina. A substituição 'provavelmente' afeta a especificidade em S_3' e talvez também em S_1' , devido à sua proximidade no sítio ativo (Fujinaga *et al.*, 1995). Tem peso molecular de aproximadamente 35 kDa, e uma única cadeia polipeptídica. Sua atividade proteolítica máxima é em pH 2,0.

Pepsina I e II

Pepsinas I e II foram descritas como os dois menores componentes proteolíticos (de menor peso molecular) do suco gástrico humano, pois migram mais rapidamente do que outras frações quando submetidas à eletroforese em gel de agarose, em pH 5,0 (Etherington & Taylor, 1967). A pepsina I, primeira fração obtida na purificação da pepsina, denominada gastricsina, é de grande interesse devido à correlação entre sua concentração no suco gástrico e em úlceras pépticas (Walker & Taylor, 1980), e devido à sua alta atividade na digestão de colágeno (Etherington *et al.*, 1980) e de mucoproteínas gástricas (Pearson *et al.*, 1984). A pepsina II, segunda e maior fração, a qual teve sua estrutura cristalográfica de difração de raios X resolvida por Fujinaga (Fujinaga *et al.*, 1995), foi denominada

pepsina 3A. A fração II da pepsina é também o componente que, quando na corrente sanguínea, é filtrada pelos rins, dando origem a uropepsina.

A pepsina consiste de uma cadeia polipeptídica simples de peso molecular de 35 kDa e 326 resíduos de aminoácidos. A Ser 68 é fosforilada, mas este fosfato é removido sem alteração significativa das propriedades da enzima. Como em outras proteinases aspárticas, o sítio ativo está em uma área estendida que pode acomodar de quatro a sete resíduos do substrato (Zinchenko *et al.*, 1976). Tem preferência por aminoácidos hidrofóbicos e, um dado cinético sobre clivagem de ligações peptídicas mostra que existe uma especificidade para leucina, fenilalanina, triptofano e ácido glutâmico no sítio S_1 , e para triptofano, tirosina, isoleucina e fenilalanina no sítio S_1' .

Os vários tipos de proteinases aspárticas diferem consideravelmente na seletividade que elas exibem para substratos particulares. A pepsina hidrolisa a ligação peptídica da porção amino da Leu, Phe, Trp e Tyr, mas não hidrolisa regiões em que há prolina ligada a estes aminoácidos (Voet & Voet, 1995). Raramente catalisa a hidrólise de éster, com exceção de ésteres de ácido L- β -fenilático e alguns sulfito ésteres (Denburg *et al.*, 1968).

Os dois resíduos cataliticamente ativos do sítio ativo são dependentes do pH, os quais tem pH ótimo entre 2 e 3, e dependem da forma ácida de um grupo de $pK_a \sim 4,5$ e da forma básica de um grupo de $pK_a \sim 1,1$ (Denburg *et al.*, 1968). Os valores de pK_a 's foram determinados a partir de reações de inibidores irreversíveis, que são designados para reagir especificamente com grupos carboxil ionizados ou não-ionizados. Componentes diazo – tais como N-diazoacetil-L-fenilalanina metil

éster, o qual reage com grupos carboxil não-ionizados – reagem especificamente com o Asp215 em pH 5,0 (Bayliss & Wybrandt, 1969). A dependência do pH sobre a velocidade de modificação dos aspartatos mostra que o pKa do Asp32 é menor que 3,0 (Hartsuck & Tang, 1972).

Interações com o ligante

Os vários tipos de proteinases aspárticas diferem consideravelmente na seletividade que exibem para substratos particulares. Interações entre todos os subsítios servem para orientar o substrato, de tal modo que o peptídeo coloca-se na proximidade dos resíduos de ácido aspárticos cataliticamente ativos. As cadeias laterais dos aspartil catalíticos estão localizadas numa cavidade central (Andreeva *et al.*, 1984; Bott *et al.*, 1982). Há uma molécula de água equidistante das carboxilas dos ácido aspárticos (James & Sielecki, 1983), a qual deve ter sua nucleofilicidade aumentada no ataque ao átomo de carbono do grupo carbonil da ligação peptídica a ser hidrolisada (Pearl, 1985). Alternativamente, os grupos aspartil devem servir para polarizar a carbonila do substrato (James *et al.*, 1985). É provável que exista um mecanismo capaz de produzir 2 átomos tetraedrais nas posições carboxi e amido, a partir da quebra da ligação, no estado de transição. Isto é confirmado pela estrutura do estado de transição da estatina do inibidor pepstatina (Marciniszyn *et al.*, 1976; Rich *et al.*, 1985) (Figura 11). Sua hidroxila ocupa a posição da molécula de água, que interage com os dois ácido aspárticos do sítio ativo (Baldwin *et al.*, 1993).

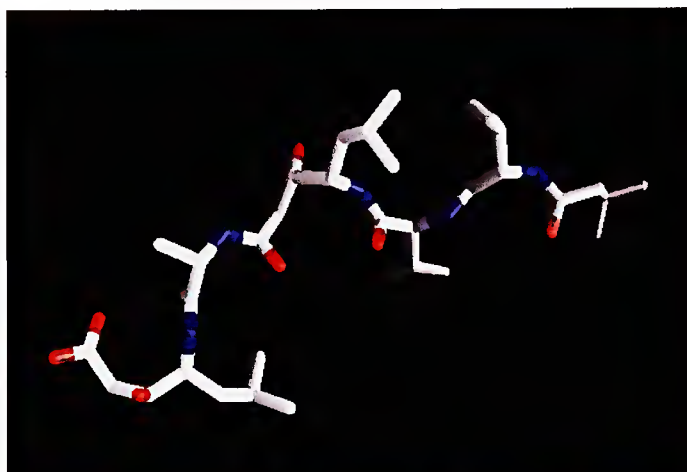


Figura 11. Molécula de pepstatina, inibidor específico de proteinases aspárticas. Em branco estão representados os átomos de carbono, em azul nitrogênio e em vermelho oxigênio.

A pepstatina é um inibidor de proteinase aspártica extremamente eficaz: é não-reativo, efetivo e específico. É um peptídeo de seis resíduos de aminoácidos que se liga especificamente à cadeia lateral de cada aspartil catalítico, inativando a enzima (Shewale *et al.*, 1985). Várias ligações de hidrogênio são necessárias para estabilizar o complexo binário. As ligações são realizadas entre átomos da cadeia principal do inibidor e átomos da cadeia principal e lateral da enzima (Figura 12) (Baldwin *et al.*, 1993).

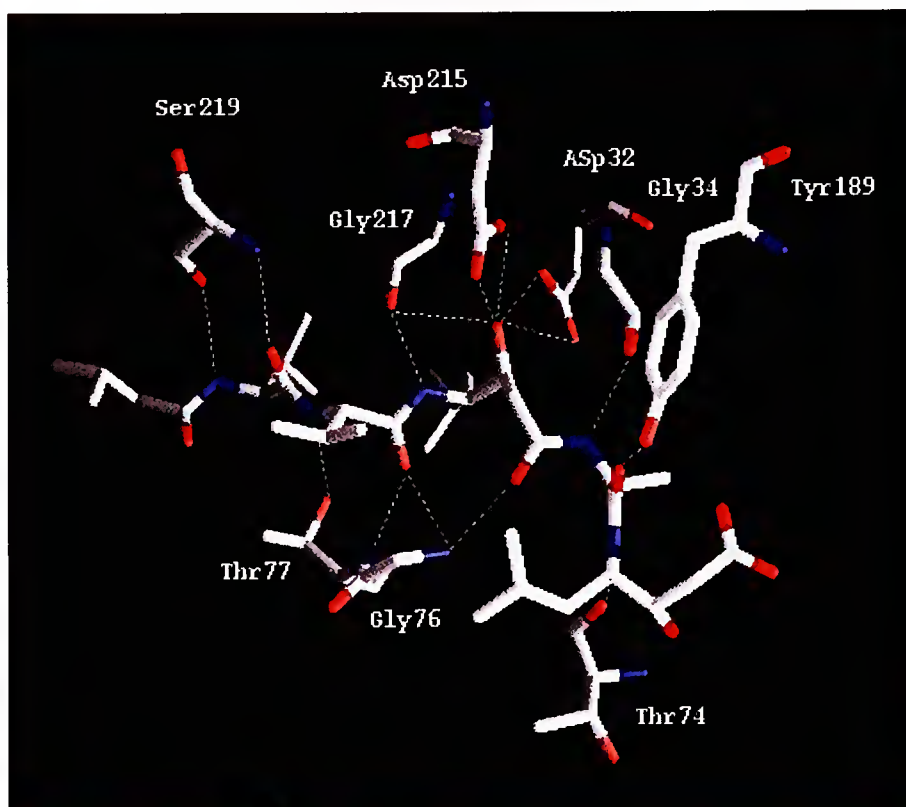


Figura 12. Visão tridimensional do sítio catalítico da pepsina em complexo com a pepstatina. Em branco estão representados os átomos de carbono, em azul nitrogênio, em vermelho oxigênio.

Mecanismo catalítico

Enzimologistas sugerem que a carboxila dos aspartatos funcionem, alternativamente, como um ácido ou uma base. Este modelo requer que a carboxila de um dos aspartatos seja protonada e a outra seja desprotonada, quando o substrato se liga à enzima. Dados de difração de raios X de proteinases aspárticas mostram que a estrutura do sítio ativo, na vizinhança dos dois aspartatos, é altamente simétrica. Os dois aspartatos parecem agir como uma díade catalítica (análogo à tríade catalítica das serino-proteinases). A díade de prótons deve ser

covalentemente ligadas a cada aspartato na enzima livre ou no complexo enzima-substrato (Figura 13) (Garrett & Grisham, 1995).

Com o substrato ligado, segue-se um passo no qual dois prótons facilitam o ataque nucleofílico sobre a carbonila do substrato, pela água. Sendo assim, o Asp32 age como uma base, recebendo um próton de uma molécula de água presente no sítio ativo, enquanto o Asp215 age como um ácido, doando um próton para o oxigênio da carbonila do peptídeo. Em virtude da transferência destes dois prótons, o ataque nucleofílico ocorre sem formação explícita do íon hidróxido no sítio ativo. O intermediário resultante é denominado uma amida diidratada. O estado de protonação dos dois resíduos aspartatos estão agora opostos àqueles na enzima livre (Figura 13) (Garrett & Grisham, 1995).

A quebra da amida diidratada ocorre por um mecanismo similar à sua formação. A carboxila ionizada do Asp32 recebe um próton a partir de uma hidroxila da amida dihidratada, no qual a carboxila protonada do Asp215 simultaneamente doa um próton ao átomo de nitrogênio de um dos produtos (peptídeos) (Garrett & Grisham, 1995).

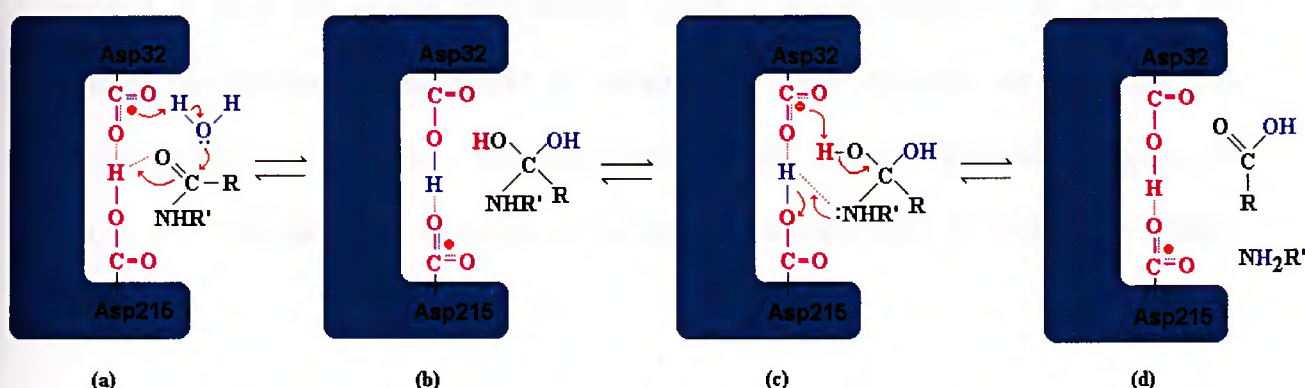


Figura 13: Mecanismo de ação das proteinases aspárticas. (a) e (b) 2 prótons facilitam o ataque nucleofílico da água sobre a carbonila. (c) e (d) o Asp 32 recebe um próton a partir de uma das hidroxilas da amina diidratada, e o outro aspartato (Asp 215) doa um próton ao nitrogênio da amina hidrolisada (Garrett & Grisham, 1995).

Transferência de energia de ligação

A ligação de um substrato ou inibidor à enzima é acompanhado por uma variação de energia suficientemente exotérmica, quanto consideravelmente favorável, ao estado ligado. O estado exotérmico resulta da interação entre grupos que constituem o ligante e aqueles que formam o sítio de ligação. É desvantajoso para as proteinases aspárticas utilizar toda a energia de interação disponível ao complexo ligado. Isto deve aumentar a barreira de ativação para alcançar o estado de transição, desfavorecendo a reação. Entretanto, é altamente vantajoso para a enzima transferir parte da energia de interação, diminuindo a barreira de ativação da reação seguinte. A grande diferença na função de hidrólise entre diferentes substratos pode ser explicada como diferenças na energia de ligação para estes substratos e o grau para o qual eles podem promover tal trabalho (Pearl, 1985).

O substrato sofre uma distorção a qual facilita a reação. Quanto maior o substrato, menor o grau de torção dele. A rotação da ligação C – N requer menos

energia e deve ter efeitos importantes sobre a susceptibilidade do carbonil no ataque nucleofílico. A distorção é estabilizada pela energia de interação do substrato com os subsítios, formando um sítio de ligação estendido. O grau de torção e a hidrólise são funções do tamanho e composição do substrato (Pearl, 1985).

Mudança conformacional de enzimas

Para suprir a energia de ligação e promover a catálise, deve-se exercer uma influência sobre os grupos que catalisam a reação, sobre a ligação a ser hidrolisada, ou ambas. Um modelo induzido sugere que a conformação da enzima muda quando da ligação do substrato. Como deve existir uma energia de interação entre os aspartatos catalíticos e o substrato para que ocorra a hidrólise, deve existir um caminho para a transferência dessa energia. Entretanto, deve-se somar a este efeito, a ligação a diferentes subsítios, o que contribui para a hidrólise. Geralmente, é mais complexo para um substrato de maior peso molecular se ligar ao sítio catalítico, tendo efeito cumulativo (Pearl, 1985).

Objetivos

OBJETIVOS

O objetivo deste trabalho foi a resolução das estruturas cristalográficas de duas proteinases aspárticas, uropepsina, no estado nativo e inibido, e pepsina I, uma fração da pepsina gástrica de baixo peso molecular, também denominada gastricsina. Para tanto, foram usadas técnicas de difração de raios X.

A comparação entre estas e outras proteinases aspárticas foram realizadas.

Materiais e Métodos

Purificação	36
Atividade Proteolítica	37
Cristalização	38
Aplicação do Método da Matriz Esparsa	39
Cristalização da uropepsina	42
Cristalização do complexo binário uropepsina:pepstatina	42
Cristalização da pepsina I	42
Eletroforese em gel de poliacrilamida SDS-PAGE	43
Coleta e processamento dos dados	43
Substituição Molecular	44
Refinamento da estrutura	47
Análise da estrutura	50
Modelagem do complexo binário uropepsina:pepstatina	50

MATERIAL E MÉTODOS

1. Purificação:

As purificações da uropepsina e da pepsina I foram realizadas pelo grupo da Profa. Dra. Roseli A. S. Gomes da faculdade de Medicina do Triângulo Mineiro (FMTM) de Uberaba-MG. As proteínas foram purificadas usando o mesmo procedimento (Gomes *et al.*, 1996).

A uropepsina humana foi extraída da urina de indivíduos jovens sem doença renal. A urina foi estocada em tubos contendo 6,0M de HCl. Cinco litros desta urina foram filtradas em papel de filtro e dialisadas contra água deionizada, as quais foram acrescentados HCl 6,0M para a obtenção de pH4,0. Nesta etapa foi adicionado azida sódica e a solução foi concentrada até um volume final de 50 ml, e em seguida, liofilizada. A amostra foi reconstituída em tampão acetato de sódio 0,01M pH4,9 e submetida a 3 etapas de cromatografia.

A pepsina I foi extraída de mucosa gástrica humana, a partir de tecido estomacal normal. Para a extração foi usado tampão glicina-HCl 0,2M, pH 3,5 contendo inibidor de serino-proteinases. Este extrato foi usado para a purificação da enzima.

As três etapas de cromatografia utilizadas nas purificações estão descritas a seguir:

Etapa 1: cromatografia de troca iônica em resina DEAE Bio Gel, utilizando tampão acetato de sódio 0,01M, pH4,9 para a eluição. Para o gradiente salino foram usados tampão A: acetato de sódio 0,01M e NaCl 0,1M, pH4,9 e tampão B: acetato de



sódio 0,01M e NaCl 0,8M, pH4,9. As frações coletadas foram analisadas quanto ao teor de proteínas em comprimento de onda de 280 nm e feita a determinação da atividade proteolítica utilizando hemoglobina bovina a 8% como substrato, segundo método de Anson (1939). As frações contendo atividade proteolítica foram reunidas, dialisadas contra tampão fosfato de sódio 0,05M, pH6,0 e concentradas.

Etapa 2: A amostra foi submetida à cromatografia de troca iônica em coluna Mono-Q HR 5/5, em sistema FPLC, utilizando tampão fosfato de sódio 0,05M e NaCl 1,0M para o gradiente salino (tampão B). Novamente as frações recolhidas foram analisadas quanto à atividade proteolítica.

Etapa 3: As frações ativas foram submetidas à cromatografia de filtração em gel em coluna Superdex 75 HR10/30, em sistema FPLC. Nesta etapa foi obtido um pico protéico simétrico, que é sobreposto pelo pico de atividade proteolítica.

2. Atividade proteolítica

A atividade proteolítica da uropepsina foi medida e comparada com a atividade da pepsina 3A, usando as mesmas condições experimentais. Estes ensaios foram realizados pelo grupo da Profa. Dra. Roseli A. S. Gomes.

As cinéticas de ambas proteinases foram realizadas em triplicatas usando tampão acetato de sódio 0,2M, pH5,0 e substrato Abz KPIEFFRL Eddnp. As concentrações do substrato variaram de 0,117 μM a 5,66 μM . Os valores da constante catalítica (k_{cat}) foram calculados após a titulação do sítio ativo segundo método de Henderson (1972), na presença de pepstatina. Foram realizados também ensaios enzimáticos usando-se pepsina I, nas mesmas condições experimentais.



3. Cristalização:

O principal objetivo da cristalização de uma macromolécula biológica, tal como proteínas, ácidos nucleicos ou vírus, é o seu estudo estrutural por métodos de difração de raios X. Este estudo é fundamental para o entendimento molecular de vários mecanismos biológicos. E a cristalização é uma das partes mais críticas na resolução da estrutura de uma macromolécula biológica. Os principais passos envolvidos na resolução da estrutura de uma macromolécula biológica, a nível atômico, estão mostrados na Figura 14.

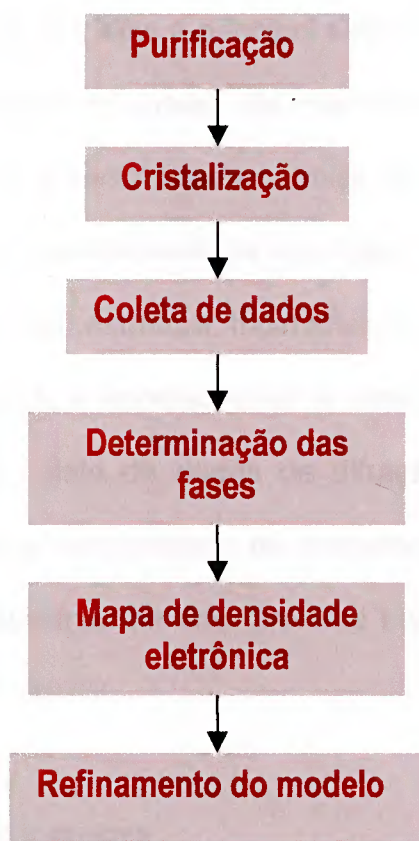


Figura 14. Principais passos envolvidos na resolução da estrutura de uma macromolécula biológica.

Entre todos estes passos, a cristalização é o menos entendido. Biocristalização, como qualquer cristalização, é um processo de múltiplos parâmetros que envolve os três passos clássicos da cristalização: nucleação, crescimento e término de crescimento (“cessation of growth”).

O que torna o processo de cristalização de macromoléculas biológicas diferente é o número de parâmetros envolvidos, muito maior do que aqueles envolvidos no crescimento de cristais de pequenas moléculas, além das propriedades físico-químicas peculiares destes compostos. Por exemplo, a estabilidade ótima das macromoléculas biológicas em um meio aquoso está restrita a uma faixa estreita de pH. Mas a principal diferença entre crescimento de cristais de pequenas moléculas e cristais de macromoléculas biológicas é a flexibilidade conformacional e a versatilidade química dessas macromoléculas e, consequentemente, sua maior sensibilidade às condições externas (Apêndice I).

Para a determinação da estrutura tridimensional de uma proteína por técnicas de difração de raios X, é imprescindível a obtenção de monocristais de tamanho apropriado para a coleta de dados de difração de raios X. Como o número de variáveis envolvidas no processo de cristalização é grande, encontrar as condições ideais de cristalização de uma proteína pode levar dias, meses ou até anos (Blundell & Johnson, 1976).

3.1. Aplicação do método da matriz esparsa

Após o preparo das 50 soluções para cristalização (Apêndice II), nós usamos o método da gota pendurada (“hanging drop”) para testar as condições de cristalização. Colocamos 1 ml de cada uma destas soluções em cada um dos



reservatórios de uma caixa Linbro (Linbro modelo 76-033-05) (Figura 15). Estas caixas foram originalmente desenhadas para cultura de células e possuem 24 reservatórios. É necessário o uso de três caixas Linbro, sendo que a terceira terá só 02 reservatórios usados para o experimento inicial. O restante dos reservatórios da terceira caixa podem ser usados em experimentos futuros. Temos então 50 reservatórios com 1 ml de cada uma das soluções de cristalização.

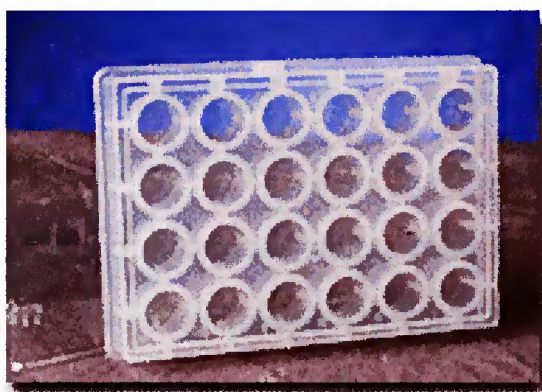


Figura 15. Caixa para cultura de células usada para a cristalização de macromoléculas biológicas.

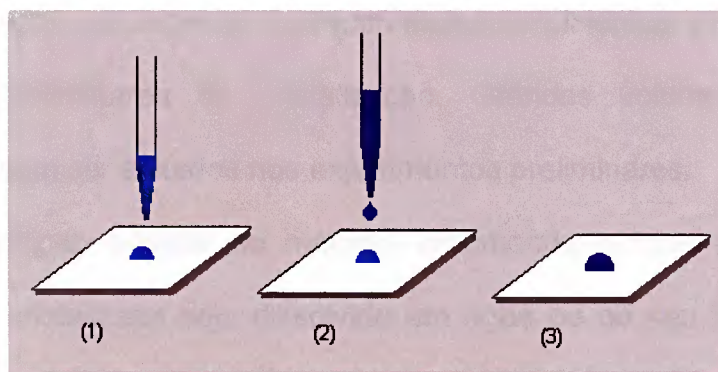


Figura 16. Sequência de eventos para a montagem de uma gota de cristalização: (1) Coloca-se 1 μ l da solução da macromolécula biológica sobre a laminula de vidro. (2) Adiciona-se 1 μ l da solução do reservatório à gota. (3) Ao final temos uma gota (1:1) com a solução de macromolécula biológica mais a solução do reservatório.

Colocamos 1 μl da solução da macromolécula biológica em uma lamínula de vidro de 22 x 22 mm que foram siliniconizadas com 1% Prosil-28 (PCR Inc., Gainseville, Florida, U.S.A.) e 1 μl da solução do reservatório (Figura 16). Cobrimos cada um dos reservatórios das caixas Linbro com as respectivas lamínulas de vidro, usando graxa de vácuo para vedar os reservatórios. Este tipo de composição é o que convencionou chamar-se de gota (1+1). Nesta situação, teremos uma concentração menor de precipitante na gota, do que no reservatório. E a difusão de vapor d'água se procederá da gota para o reservatório. No caso em que as espécies químicas do reservatório possuam pressão de vapor menor que a pressão de vapor d'água, fará com que a concentração da solução de macromolécula biológica, bem como das outras espécies químicas presentes na gota, aumente. Em casos favoráveis, isto levará a macromolécula biológica a uma condição de supersaturação favorável à formação dos primeiros agregados cristalinos (zona de nucleação). Tal composição (1 μl + 1 μl) pode ser mudada, bem como o volume inicial da gota. Contudo, em um experimento de cristalização deve-se ter em mente o uso racional da macromolécula biológica, principalmente nos experimentos preliminares de cristalização. Grandes volumes de gotas de cristalização devem ser evitados nos experimentos preliminares.

Na proposição original do método, recomenda-se que a macromolécula biológica a ser cristalizada seja dissolvida em água ou no seu tampão original e assim, trazida a uma concentração de 10 mg/ml ou mais, dependendo da macromolécula em estudo. Muitas macromoléculas biológicas não conseguem



atingir tal concentração sem precipitar. Neste caso, recomenda-se usar a mais alta concentração possível.

3.2. Cristalização da uropepsina

A uropepsina liofilizada foi dissolvida em água deionizada para uma concentração de 15mg/ml e centrifugada por 10 minutos a 15.000g, a 25°C. O sobrenadante claro foi misturado, na proporção de 1:1 (v/v) com a solução do reservatório (Jancarik & Kim, 1991).

Os cristais foram montados em tubos capilares de borosilicato para a coleta de dados de difração de raios X.

3.3. Cristalização do complexo binário uropepsina:pepstatina

Para a cristalização do complexo binário a uropepsina liofilizada foi dissolvida em 66 µl de água deionizada para uma concentração de 15mg/ml e adicionado 1mM de pepstatina. Esta mistura foi mantida em repouso por aproximadamente 15 minutos, para que o complexo fosse formado.

3.4. Cristalização da pepsina I

A pepsina I (gastricsina) liofilizada foi dissolvida em água deionizada para uma concentração de 15mg/ml e centrifugada por 10 minutos a 15.000g, a 25°C. O sobrenadante claro foi misturado, na proporção de 1:1 (v/v), com a solução do reservatório.



4. Eletroforese em gel de poliacrilamida SDS-PAGE

Realizamos uma eletroforese em gel de poliacrilamida SDS-PAGE a 12% no laboratório de bioquímica da Faculdade de Medicina do Triângulo Mineiro (FMTM) de Uberaba (MG). O gel foi feito segundo o método de Laemmli (1970). Foi usado o kit para realização de eletroforese em gel de poliacrilamida da Bio Rad. Usamos as seguintes concentrações dos géis:

Gel de Corrida: Bisacrilamida 12%; Tampão do gel de corrida (Tris-HCl 0,375M, pH 8,8; SDS 0,1%); APS 0,05%; TEMED 0,464M.

Gel de Empilhamento: Bisacrilamida 10%; Tampão do gel de empilhamento (Tris-HCl 0,303M, pH6,8; SDS 0,1%); APS 0,2%; TEMED 0,162M.

A eletroforese foi realizada na presença de tampão de corrida Tris-Glicina (0,025M Tris pH 8,3; 0,192M Glicina; 0,1% SDS) a 60V por 2 horas.

Para visualização das bandas do gel foi usado a coloração por nitrato de prata, pelo método de Gordon (Gordon, 1975).

5. Coleta e processamento dos dados

Os dados de difração de raios X de um monocristal de uropepsina foram primeiramente coletados a temperatura ambiente usando um sistema de placa de imagem do R-AXIS IV e radiação $\text{CuK}\alpha$ produzida pelo gerador de ânodo rotatório Rigaku RU300 operando a 50kV e 100mA, sendo que a distância cristal-placa de imagem foi de 150mm. Um total de 40 imagens foram coletadas usando uma faixa de oscilação de $2,5^\circ$ por imagem. O tempo de exposição por imagem foi de 20

minutos. Os dados de difração de raios X foram processados a 2,8 Å de resolução, usando o programa PROCESS (Higashi, 1990).

Um segundo conjunto de dados de difração de raios X foi coletado a um comprimento de onda de 1,38 Å usando fonte de radiação síncrotron (estação PCr, Laboratório Nacional de Luz Síncrotron, LNLS, Campinas, Brasil) com placa de imagem MAR Research de 30 cm de diâmetro. A distância entre o cristal e a placa de imagem foi de 175 mm com tempo de exposição de 3 minutos por imagem. Usando uma faixa de oscilação de 1,5°, 65 imagens foram coletadas e novos dados de difração de raios X foram processados a 2,45 Å de resolução, usando os programas DENZO (Gewirth, 1995) e SCALEPACK (Gewirth, 1995). A faixa de resolução desejável está entre 1,5 a 2,8 Å. Procedimentos de auto indexação, combinados com análises do padrão de difração de raios X e a média das intensidades equivalentes, foram usados na caracterização da simetria de Laue.

6. Substituição molecular

Os dados de difração de raios X coletados usando radiação síncrotron foram usados na substituição molecular.

A estrutura do cristal da uropepsina humana foi determinada por métodos de substituição molecular padrões usando os programas AMoRe (Navaza, 1994) e X-PLOR (Brünger, 1992) (Apêndice I). Valores estimados do volume de Mathews e do conteúdo de solvente do cristal foram calculados segundo as equações:

$$V_m = \frac{V_{\text{cela}}}{Z \cdot PM}$$



onde V_m é o volume de Mathews (Mathews encontrou que, para muitos cristais, a razão do volume da cela unitária e o peso molecular estão entre 1,7 e 3,5 Å³/Da, com mais valores a cerca de 2,15 Å³/Da), V_{cela} é o volume da cela unitária, encontrado a partir da equação abaixo:

$$V_{\text{cela}} = a.b.c \sqrt{1 - \cos^2 \alpha - \cos^2 \beta - \cos^2 \gamma + 2 \cos \alpha \cos \beta \cos \gamma}$$

onde para uma cela ortorrômbica, $\alpha=\beta=\gamma=90^\circ$, temos que $V_{\text{cela}} = a.b.c$, dado em Å³. Z é o número de moléculas por cela unitária e PM é o peso molecular da macromolécula biológica em estudo (dado em Daltons).

Para encontrarmos o valor da fração de volume do solvente, fazemos

$$V_{\text{macromolécula}} + V_{\text{solvente}} = 1$$

portanto,

$$V_{\text{solvente}} = 1 - V_{\text{macromolécula}}$$

onde $V_{\text{macromolécula}}$ e V_{solvente} são, fração de volume da macromolécula biológica e fração de volume do solvente na cela unitária, respectivamente. Para encontrarmos $V_{\text{macromolécula}}$, usamos a equação

$$V_{\text{macromolécula}} = \frac{m_H}{\rho_{\text{prot}} \cdot V_M} = \frac{1,23}{V_M}$$

onde m_H é a massa do hidrogênio e ρ_{prot} é a densidade da proteína.

A densidade do cristal também foi determinada segundo a equação

$$\rho = \frac{Z \cdot PM \cdot m_H}{V_{\text{cela}}} + \rho_{\text{solv}} \cdot V_{\text{solv}}$$

onde, m_H é $1,673 \cdot 10^{-24} \text{g}$ e ρ_{solv} é a densidade do solvente (1g/cm^3).

As coordenadas atômicas de dez proteinases aspárticas diferentes depositadas no PDB (Protein Data Bank) foram usadas como modelos de busca. Foi removido todo o solvente e moléculas de ligantes dos modelos de busca e o fator de temperatura (fator B) para todos os átomos foi fixado a $20,00 \text{ \AA}^2$. Os modelos de busca que apresentaram mais do que uma subunidade foram modificados, de forma que só uma subunidade foi usada na substituição molecular. As coordenadas atômicas para todos os modelos de busca foram transladados de forma que o centro de gravidade ficasse na origem. Os modelos também foram girados de forma que os eixos principais de inércia dos modelos de busca ficassem paralelos aos eixos ortogonais. O número de acesso ao PDB e a identificação dos modelos de busca estão listados na Tabela 3. Inicialmente, as funções de auto-rotação, baseados numa faixa de resolução de $10\text{-}4,5 \text{ \AA}$, foram calculados através do programa AMoRe usando um passo simples de 1° (Navaza, 1994). As funções de rotação cruzadas foram calculadas numa faixa de resolução de $10\text{-}4,5 \text{ \AA}$ usando um passo de $2,5^\circ$ para todos os 10 modelos de busca. Estes cálculos foram feitos com um raio de integração de $25 \text{ \AA}$. A rotação que gerou o coeficiente de correlação mais alto (CC) foi aplicado aos modelos de busca e usados nos cálculos da função de translação subseqüentes, baseados em dados na mesma faixa de resolução.

Foram selecionadas as melhores soluções para cada modelo de busca baseado na magnitude do R_{factor} e coeficiente de correlação (Apêndice I). O modelo



de busca que gerou a melhor solução também foi usado para a substituição molecular como implementado no programa X-PLOR (Brünger, 1992), usando a mesma faixa de resolução e o mesmo raio de integração, para confirmar a solução obtida usando o programa AMoRe (Navaza, 1994). Funções de translação para grupos espaciais (P222, P2₁2₁2 e P222₁) também foram calculadas, usando a mesma faixa de resolução, para confirmar o grupo espacial e o melhor modelo de busca.

Tabela 3. Lista das proteinases aspárticas usadas como modelos de busca para substituição molecular.

Número de Acesso ao PDB	Proteína
1eag	Proteinase aspártica de <i>Candida albicans</i> (Candida pepsina)
1lya	Catepsina D de fígado humano
1zap	Proteinase aspártica de <i>Candida albicans</i>
1bbs	Renina humana
1mpp	Pepsina (renina) de <i>Mucor pusillus</i>
2ren	Renina (recombinante da renina humana)
4apr	Proteinase aspártica de <i>Rhizopus chinensis</i>
1psn	Pepsina 3A de <i>Homo sapiens</i>
3app	Penicilopepsina de <i>Penicillium janthinellum</i>
1cms	Quimosina B bovina (<i>Bos taurus</i>)

7. Refinamento da estrutura

O modelo foi refinado usando o programa X-PLOR (Brünger, 1992) (Apêndice I). Primeiramente geramos as informações moleculares da estrutura a ser refinada. Estas informações foram divididas em dois arquivos, sendo que um deles possui a topologia da molécula (massa e tipo dos átomos, ligação e ângulos de ligação, ângulos diédricos, carga elétrica e ligações de hidrogênio) e o outro possui as informações químicas (constantes a serem usadas no cálculo da função

energia e que será minimizado durante o refinamento). Estes arquivos foram gerados a partir das coordenadas do modelo em formato PDB.

O segundo passo foi o refinamento de corpo rígido, em que foram refinadas as posições de grupos rígidos dos átomos contra as intensidades observadas. Foram realizados 20 ciclos de refinamento de corpo rígido. Em seguida, realizamos o refinamento posicional, o qual usa como modelo inicial as coordenadas de saída do refinamento de corpo rígido. Uma vez que o refinamento posicional tenha convergido, o passo seguinte foi o refinamento por dinâmica molecular ou “simulated annealing”. O cálculo da dinâmica molecular começa com a atribuição de velocidades aos átomos da macromolécula. Estas velocidades são derivadas de uma distribuição de Maxwell a uma temperatura na faixa de 2000 a 4000 Kelvin. No tempo $t = 0$ s, os átomos estão em uma configuração inicial com energia potencial. Sobre cada átomo, existe uma força agindo, que é derivada da energia potencial. A partir da mecânica Newtoniana, a aceleração para cada massa do átomo pode ser calculada e aplicada. Após um curto intervalo de tempo (femtosegundos), o processo é repetido com os átomos em novas posições. Foram realizados 900 ciclos de dinâmica molecular no refinamento da uropepsina, o suficiente para que um novo mínimo de energia potencial fosse atingido. Após elevar a temperatura do sistema o suficiente (4000 K), e então resfriá-lo lentamente para que a energia se aproximasse do mínimo, foram feitos novos ciclos de refinamento posicional.

As coordenadas atômicas do modelo refinado foram usadas para gerar mapas de densidade eletrônica. No estudo da uropepsina foram usados dois tipos de mapas de densidade eletrônica. O primeiro foi o $F_{\text{obs}} - F_{\text{calc}}$, onde F_{obs} é o fator de estrutura observado e F_{calc} é o fator de estrutura calculado e as fases são



aquelas calculadas a partir do modelo refinado (Apêndice I). É a partir deste mapa que podemos localizar partes da estrutura que não fazem parte do modelo refinado, tais como ligantes e moléculas de água, que não estão ainda na estrutura do modelo refinado. O outro mapa utilizado foi $2F_{\text{obs}} - F_{\text{calc}}$. Este mapa é útil na localização de pequenos desvios da cadeia lateral e para a observação do aspecto geral da estrutura. Tal visualização foi feita com a superposição das coordenadas do modelo refinado com o mapa de densidade eletrônica no programa XtalView (McRee, 1993). Após os ajustes feitos no modelo, novo ciclo de refinamento foi realizado, até que todas as moléculas de água fossem adicionadas, e o R_{factor} e o R_{free} tivessem convergido. Para a localização das moléculas de água foi utilizada a rotina de procura de moléculas de água do programa XtalView (McRee, 1993), usando os mapas $F_{\text{obs}} - F_{\text{calc}}$.

Os dois índices de concordância, R_{factor} e R_{free} , são calculados durante cada ciclo de refinamento pelo programa X-PLOR. O R_{factor} mostra a concordância entre os fatores de estrutura observados e calculados e é mais baixo para as estruturas mais próximas aos dados experimentais.

Um outro índice de concordância bem mais robusto entre os fatores de estrutura observados e calculados é o R_{free} . Este índice é calculado da mesma forma que o R_{factor} , só que ele usa somente parte das reflexões do conjunto de dados (por exemplo, 10% das reflexões são deixadas fora do refinamento) (Apêndice I). Os valores de R_{free} e R_{factor} podem ser calculados para cada ciclo do refinamento, sendo possível acompanhar se o processo está convergindo para bons índices de concordância. Valores de R_{free} abaixo de 35% são indicativos de que a estrutura está de alguma forma correta, necessitando algum refinamento.

Valores muito acima de 35% após vários ciclos de refinamento, indicam que grande parte da estrutura está errada (de Azevedo, 1997).

8. Análise da estrutura

Diferença de desvios de r.m.s. a partir da geometria ideal para distância entre ligações, ângulos e ângulos diédricos foram calculados com X-PLOR 3.1 (Brunger, 1992).

A qualidade estereoquímica geral do modelo final foi observado pelo programa PROCHECK (Laskowski *et al.*, 1993).

9. Modelagem do complexo binário uropepsina:pepstatina

Várias tentativas de co-cristalizar a uropepsina com pepstatina não produziram cristais de alta qualidade. Um modelo do complexo binário foi construído. Este modelo foi baseado na estrutura da pepsina complexada com pepstatina resolvido a alta resolução (código PDB: 1pso) (Fujinaga *et al.*, 1995). O modelo do complexo binário da proteína foi obtido pela sobreposição das coordenadas dos três domínios (N-, central e C-terminal) da pepsina inibida, mas na ausência do inibidor pepstatina, utilizando o programa ProFit V. 1.8 (Martin, 1992-8). O inibidor foi adicionado posteriormente. As coordenadas da pepstatina foram obtidas em HIC-Up (<http://xray.bmc.uu.se/hicup/>).

As coordenadas da uropepsina e do inibidor foram minimizados usando-se o programa X-PLOR (Brünger, 1992). Foi criado um arquivo de entrada em que as coordenadas dos três domínios (N-terminal, central e C-terminal) da proteína foram

denominados cadeias A, B e C, respectivamente, e o arquivo das coordenadas do inibidor foi denominado cadeia D. Deste arquivo de entrada foi gerado o arquivo a ser refinado. Foram feitos 200 ciclos de refinamento posicional, onde o peso do termo de raios X da função energia foi colocado como zero. Após os 200 ciclos de refinamento posicional, a energia caiu de $2,9 \times 10^6$ para 3165,25 kcal/mol. Este modelo foi usado para comparações com o complexo binário pepsina:pepstatina.

Os modelos atômicos foram sobrepostos usando o programa LSQKAB a partir do CCP4 (CCP4, 1994).



Resultados e Discussão

Atividade Proteolítica	53
Cristalização da uropepsina	53
Cristalização do complexo binário uropepsina:pestatina	54
Cristalização da pepsina I	56
Coleta e processamento dos dados do cristal da uropepsina	57
Substituição Molecular	58
Refinamento da uropepsina	60
Análise da estrutura	64
Parâmetros de vibração térmica	69
Contatos intermoleculares e empacotamento do cristal	70
Comparação com outras proteinases aspárticas	74
Interação com pepstatina e sítio de ligação do substrato	76
Potencial eletrostático	78

RESULTADOS E DISCUSSÃO

1. Atividade proteolítica

As cinéticas realizadas usando-se uropepsina e pepsina 3A nas mesmas condições experimentais resultaram em valores de K_{cat} de $6,01 \pm 0,11 \text{ s}^{-1}$ e $5,92 \pm 0,21 \text{ s}^{-1}$, respectivamente. A constante de Michaelis-Menten (K_M) para a uropepsina foi de $1,76 \pm 0,09 \text{ }\mu\text{M}$, e para a pepsina 3A foi de $1,53 \pm 0,11 \text{ }\mu\text{M}$. A constante catalítica para a pepsina I foi de $0,03 \text{ s}^{-1}$ e o K_M foi de $2,69 \text{ }\mu\text{M}$.

2. Cristalização da uropepsina

Os cristais de uropepsina humana foram obtidos em diferentes condições de cristalização. Os melhores cristais foram obtidos após 1 semana em gotas nas quais $3 \text{ }\mu\text{l}$ de solução de enzima foram misturados com igual volume de $0,1\text{M}$ de tampão Hepes (pH 7,5) contendo 2% de polietileno glicol 400 e $2,0\text{M}$ de sulfato de amônio. A Figura 17 ilustra uma fotomicrografia do cristal de uropepsina apropriado para os experimentos de difração de raios X, com dimensões médias de $0,3 \times 0,4 \times 0,5 \text{ mm}$.



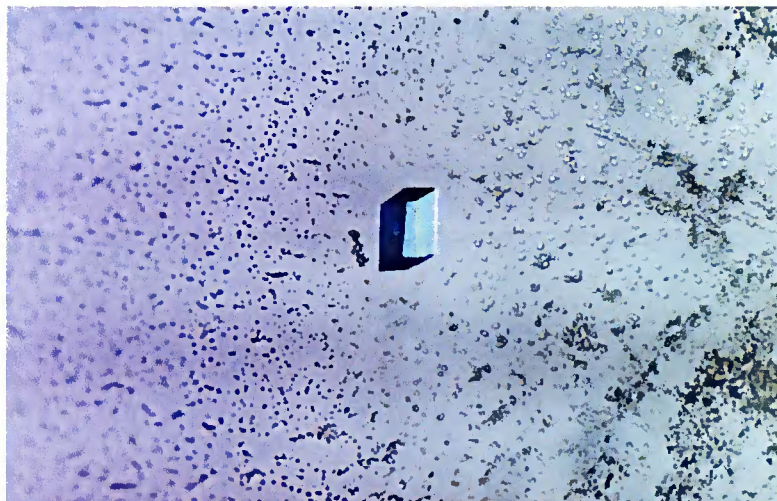


Figura 17. Fotomicrografia do cristal de uropepsina obtido em 0,1M de tampão Hepes, pH7,5, 2% de PEG 400 e 2,0M de sulfato de amônio, a 20°C.

Tentativas de cristalizar a nova amostra de uropepsina nativa não reproduziram cristais usando as condições anteriores (0,1 M de Hepes pH 7,5, 2% de polietilenoglicol 400 e 2,0 M de sulfato de amônio).

3. Cristalização do complexo binário uropepsina:pepstatina

Não foi possível a obtenção de cristais de boa qualidade da uropepsina na presença de pepstatina. Ou seja, os cristais não apresentaram um empacotamento ordenado das moléculas de proteína dentro da cela unitária. Os cristais obtidos apareceram nas seguintes condições:

- a) 0,1 M de Hepes pH 7,5 e 1,5 M de sulfato de lítio
- b) 0,2 M de acetato de magnésio, 0,1 M de cacodilato de sódio pH 6,5 e 20% polietilenoglicol 8000.

Variações na concentração do precipitante foram feitas, além de mudanças nos valores do pH. Os cristais estão ilustrados na Figura 18.

A eletroforese em gel de poliacrilamida (Figura 19) mostra a pureza da proteína purificada.

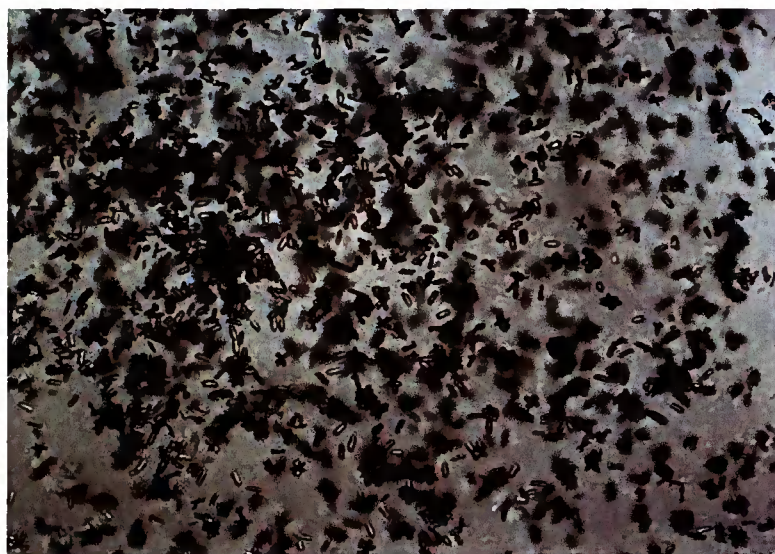


Figura 18. Fotomicrografia dos cristais de uropepsina na presença de pepstatina obtidos em 0,1 M tampão de Hepes, pH 7,5 e 1,5 M de sulfato de lítio, a 20°C.

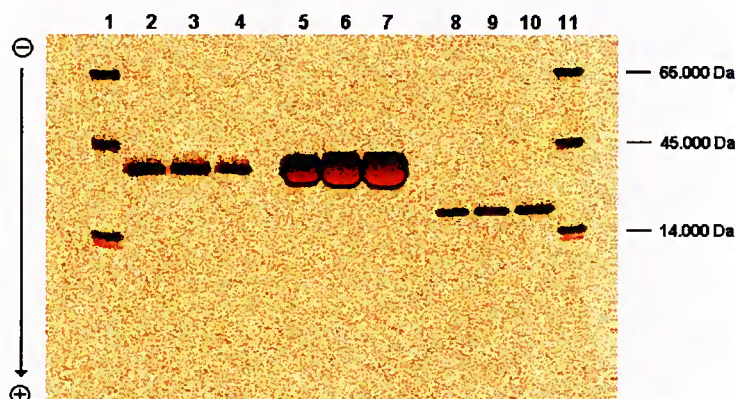


Figura 19. Foto da eletroforese em gel de poliacrilamida desnaturante a 12%. (1,11) marcador: peso molecular: soroalbumina bovina: 65.000 Da, ovoalbumina: 45.000 Da e lisozima: 14.000 Da; (2,3,4) uropepsina; (5,6,7) pepsina II; (8,9,10) pepsina I.

Pelo gel de eletroforese observamos que a diferença entre a pepsina I e a II é o peso molecular, sendo que a pepsina II possui mesmo peso molecular que a uropepsina, já que são isoenzimas.

4. Cristalização da pepsina I humana

Tentativas de cristalizar a pepsina I também não resultaram em bons cristais. Os microcristais obtidos cresceram em:

- (a) 0,2 M de sulfato de amônio, 0,1 M de cacodilato de sódio pH 6,5 e 30% de polietilenoglicol 8000
- (b) 0,1M de Hepes pH 7,5 e 1,5M de sulfato de lítio.

Os microcristais estão ilustrados na Figura 20 e as bandas da pepsina I estão nas canaletas 8, 9 e 10 do gel de eletroforese (Figura 19).

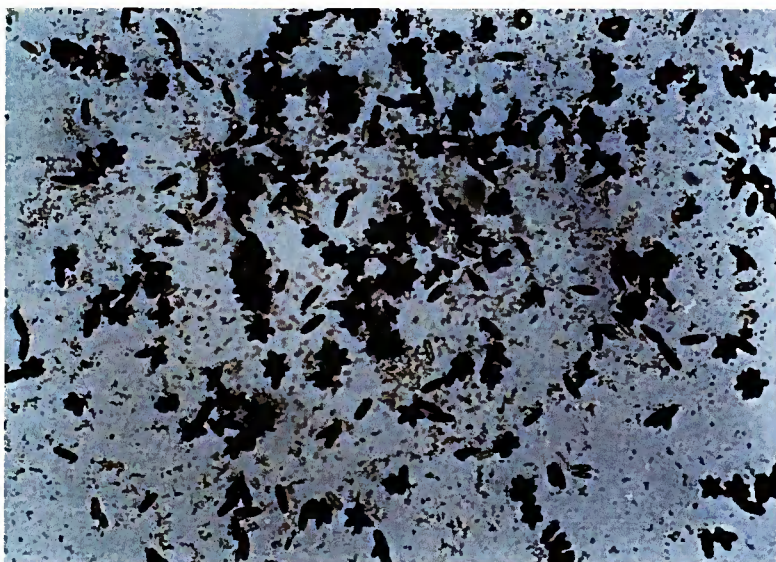


Figura 20. Fotomicrografia dos cristais de pepsina I obtidos em 0,1M tampão de Hepes, pH 7,5 e 1,5 M de sulfato de lítio, a 20°C.

5. Coleta e processamento dos dados do cristal da uropepsina

O cristal da uropepsina nativa é primitivo ortorrômbico e tem grupo espacial $P2_12_12_1$. O volume da cela unitária é $346 \times 10^3 \text{ \AA}^3$, compatível com um monômero na unidade assimétrica, o valor do volume de Mathews (V_m) é $2,17 \text{ \AA}^3/\text{Da}$. Assumindo o valor de $0,74 \text{ cm}^3 \text{ g}^{-1}$ para o volume parcial específico da proteína, o conteúdo do solvente calculado no cristal é de 43,3% e a densidade do cristal é de $1,21 \text{ g cm}^{-3}$ (Canduri *et al.*, 1998).

As estatísticas da difração de raios X para os dois conjuntos de dados estão resumidos na Tabela 4. Detalhes estatísticos da coleta de dados, usando luz síncrotron, estão na Tabela 5.

Tabela 4. Estatísticas da coleta para ambos conjuntos de dados.

	R-AXIS IV	LNLS
a (Å)	51,08	50,99
b (Å)	75,91	75,56
c (Å)	89,99	89,90
Número de medidas com $I > 2 \sigma(I)$	17686	24189
Número de reflexões independentes	10232	13134
R^*_{sym} (%)	10,4	7,7
Camada de mais alta resolução (Å)	3,5 – 2,8	2,51 – 2,45
Completeza à mais alta resolução (%)	79,3	99,1
R^*_{sym} à mais alta resolução (%)	18,5	30,3

* $R_{\text{sym}} = 100 \sum \sum |I(h) - \langle I(h) \rangle| / \sum I(h)$ com $I(h)$, intensidade observada e $\langle I(h) \rangle$, intensidade de reflexões h sobre todas as medidas de $I(h)$.

Tabela 5. Detalhes estatísticos da coleta para os dados de difração de raios X usando luz síncrotron.

Faixa de Resolução			Número de reflexões independentes	R_{sym} (%)	Completeza (%)
13,00	-	5,86	916	4,2	94,4
5,86	-	4,73	888	6,1	98,1
4,73	-	4,15	895	6,1	99,0
4,15	-	3,78	894	7,0	99,8
3,78	-	3,52	892	7,7	99,9
3,52	-	3,31	881	8,3	99,7
3,31	-	3,15	864	9,8	99,5
3,15	-	3,02	862	12,1	99,4
3,02	-	2,90	880	15,1	99,5
2,90	-	2,80	861	17,4	99,8
2,80	-	2,71	865	20,2	99,2
2,71	-	2,64	886	22,4	99,0
2,64	-	2,57	836	26,4	98,8
2,57	-	2,51	846	28,5	98,9
2,51	-	2,45	868	30,3	99,1
Total de Reflexões			13134	7,7	98,9

6. Substituição molecular

A análise do pico da função de rotação não revelou a presença de nenhum eixo de simetria local significativa, sugerindo que uma subunidade simples está presente na unidade assimétrica. Isto está de acordo com os valores estimados do conteúdo de solvente do cristal e do valor de V_m (Mathews, 1968). Os resultados da substituição molecular usando os 10 modelos de busca diferentes estão listados na Tabela 6. Os coeficientes de correlação calculados após função de translação estão entre 14,4 e 67,8% e o R_{factor} varia de 35,0 a 54,1%. O pico mais alto calculado para a função translação usando AMoRe (Navaza, 1994) foi 42,8% acima do próximo pico mais alto. O modelo de busca que apresentou o melhor coeficiente de correlação e R_{factor} foi a pepsina 3A de *Homo sapiens* (número de acesso: 1psn). Este modelo de busca foi submetido à substituição molecular usando o programa X-

PLOR, e a solução obtida após a busca de translação foi ($\alpha=23,6^\circ$, $\beta=47,0^\circ$, $\gamma=59,8^\circ$, $x=0,314$, $y=0,029$, $z=0,443$, $R_{\text{factor}}=34,6\%$) próximo ao obtido no programa AMoRe (Navaza, 1994).

Tabela 6. Ângulos de Euler (α , β , γ) e coordenadas fracionárias (x , y , z) após função de translação.

Proteína	$\alpha (^\circ)$	$\beta (^\circ)$	$\gamma (^\circ)$	x	y	z	CC (%)	$R^*_{\text{factor}} (\%)$
1eag	33,83	65,23	76,65	0,2231	0,1861	0,2065	14,4	54,1
1lya	156,00	46,61	121,10	0,3169	0,4727	0,0587	39,4	46,7
1zap	122,98	39,94	38,05	0,2905	0,2007	0,3434	14,4	55,2
1bbs	152,67	50,03	121,40	0,3250	0,4819	0,0588	26,5	50,8
1mpp	21,80	90,00	4,34	0,2000	0,1747	0,0514	16,2	53,5
2ren	30,54	40,01	235,18	0,3283	0,0154	0,4332	33,1	48,4
4apr	149,04	44,86	125,18	0,3176	0,4861	0,0618	22,1	52,5
1psn	24,44	46,13	59,44	0,3224	0,0247	0,4385	67,8	35,0
3app	146,53	46,99	307,78	0,3058	0,4739	0,0524	20,2	53,2
1cms	152,27	45,15	122,87	0,3209	0,4742	0,0610	39,3	48,9

* $R_{\text{factor}} = 100 \times \Sigma |F_{\text{obs}} - F_{\text{calc}}| / \Sigma (F_{\text{obs}})$, a soma deve ser maior em todas as reflexões com $F/\sigma(F) > 2$.

A função de translação para os grupos espaciais $P222$, $P222_1$, $P2_12_12$ foram calculados usando as coordenadas da pepsina 3A como modelo de busca e os resultados estão listados na Tabela 7. Os coeficientes de correlação, após o cálculo da função de translação para os 3 grupos espaciais, estão entre 26,4 e 38,0% e o R_{factor} varia de 46,4 a 51,1%, o que fortemente indica que o grupo espacial correto é $P2_12_12_1$.

Tabela 7. Coordenadas fracionárias após função de translação para outros possíveis grupos espaciais.

Grupo Espacial	x	y	z	CC(%)	$R_{\text{factor}}(\%)*$
$P2_12_12$	0,0748	0,1244	0,1889	38,0	46,4
$P222_1$	0,0702	0,0254	0,3940	38,5	46,5
$P222$	0,0080	0,0045	0,0104	26,4	51,1

* $R_{\text{factor}} = 100 \times \Sigma |F_{\text{obs}} - F_{\text{calc}}| / \Sigma (F_{\text{obs}})$, a soma deve ser maior em todas as reflexões com $F/\sigma(F) > 2$.

7. Refinamento da uropepsina

Os valores de R_{factor} e R_{free} para cada etapa do refinamento da uropepsina estão listados na tabela 8.

Tabela 8. Valores de R_{factor} e R_{free} para cada etapa do refinamento da uropepsina.

Refina- mento	Corpo rígido	Posicional	Dinâmica molecular	Posicional	Fator B	Fator B + H ₂ O
R_{free}	0,323-0,320	0,324-0,320	0,320-0,330	0,325-0,321	0,308-0,284	0,254-0,251
R_{factor}	0,322-0,286	0,322-0,286	0,288-0,231	0,231-0,223	0,215-0,190	0,167-0,166

A uropepsina tem somente um resíduo de aminoácido diferente quando comparado à seqüência da pepsina gástrica humana usada na substituição molecular. Examinando o mapa de densidade eletrônica foi identificado que a estrutura se trata de uma isoforma da pepsina gástrica. Esta isoforma se origina da mutação na posição do resíduo 291 (Leu→Val). Um “omit map” para esta região é mostrado na Figura 21. Este mapa representa uma região do mapa de densidade eletrônica onde parte das coordenadas atômicas são retiradas do cálculo do fator de estrutura, eliminando, desta forma, informações estruturais de confiabilidade duvidosa, como por exemplo, resíduos de aminoácidos que necessitam de confirmação. A substituição da Leu291 pela Val291 foi realizada e o modelo modificado submetido ao refinamento cristalográfico usando protocolo de “slow-cooling” como implementado no programa X-PLOR (Brunger, 1992).



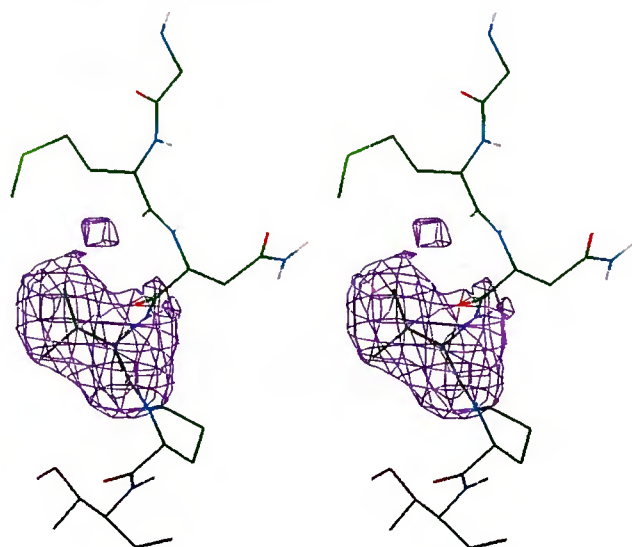


Figura 21. "Omit map" ($F_{\text{obs}} - F_{\text{calc}}$, 2σ) para a região em que existe a substituição na posição 291 (isoforma Leu-Val).

A evolução dos valores de R_{factor} e R_{free} nas seis etapas do refinamento está mostrada na Figura 22. Os valores de R_{factor} e R_{free} durante os 20 ciclos de refinamento de corpo rígido permaneceram próximos e estáveis. A diferença entre R_{factor} e R_{free} ocorreu durante o refinamento posicional. Os valores de R chegaram a uma diferença de 9% em média, entre eles. No final do refinamento, após adicionar as moléculas de água e analisar os fatores de temperatura (valores acima de 60,00 foram eliminados e nova análise do mapa $F_{\text{obs}} - F_{\text{calc}}$ foi realizada), chegamos a um R_{factor} de 16,6% e R_{free} de 25,1%, com 143 moléculas de água no modelo final. A figura 23 mostra mapas de densidade eletrônica ($2F_{\text{obs}} - F_{\text{calc}}$) na região do sítio catalítico, ao final de cada uma das seis etapas do refinamento cristalográfico, como definido na Figura 22. A uropepsina humana consiste de 2437 átomos da proteína. A qualidade geral do mapa de densidade eletrônica é boa como pode ser

vista na Figura 24. A região mostrada é a do sítio ativo, com os dois resíduos de ácido aspártico (Asp32 e Asp215).

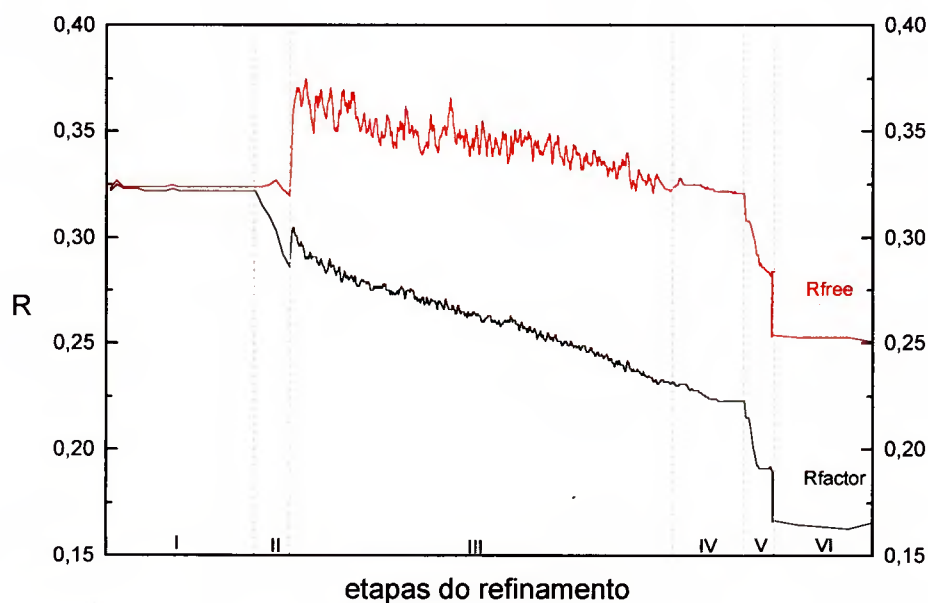


Figura 22. Gráfico mostrando o valor do R_{factor} e do R_{free} ao longo de todos os passos do refinamento da uropepsina. (I) refinamento de corpo rígido; (II) refinamento posicional; (III) "simulated annealing"; (IV) refinamento posicional; (V) refinamento do fator b; (VI) refinamento do fator b com 143 moléculas de água.

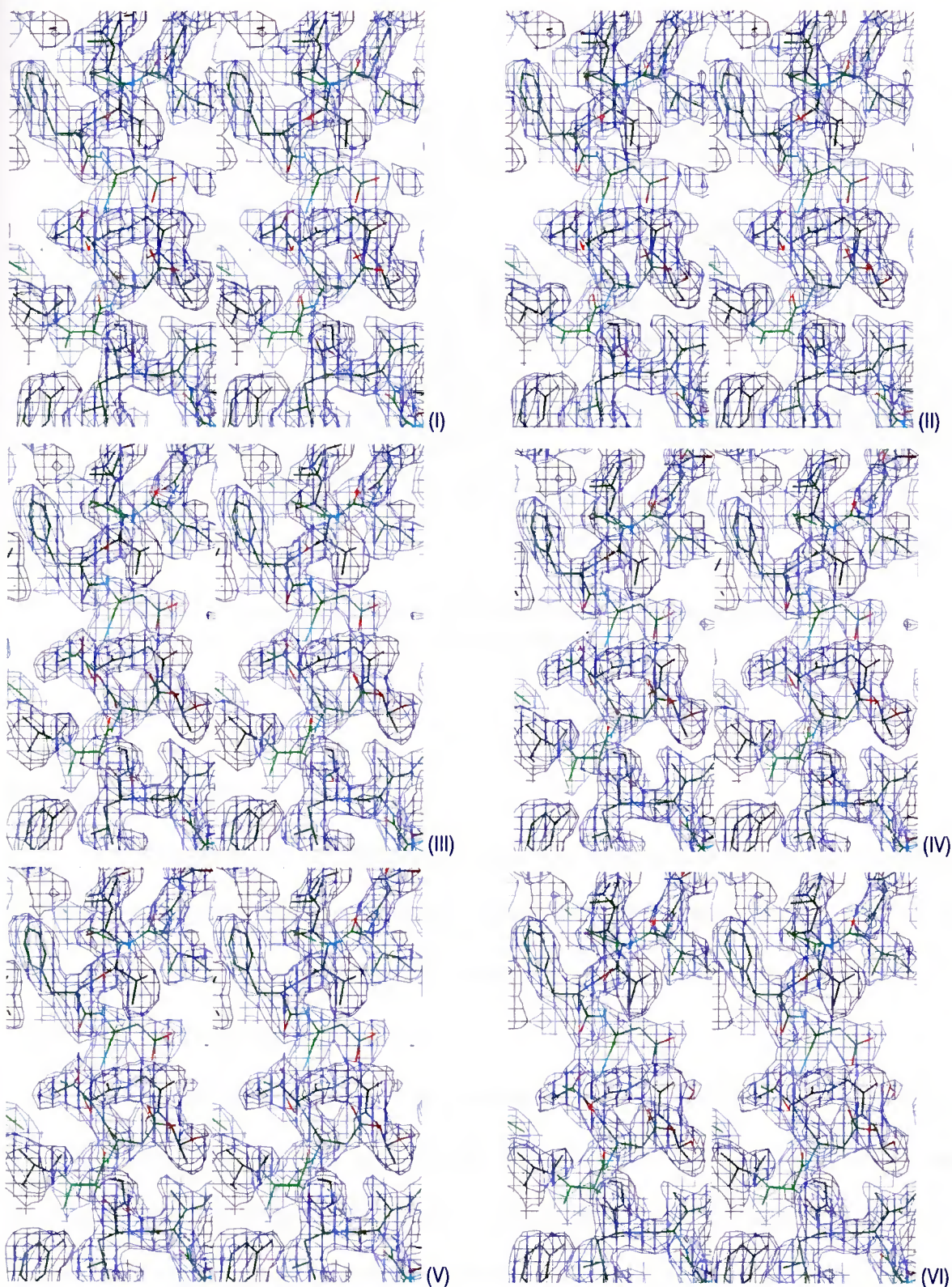


Figura 23. Mapas de densidade eletrônica ($2F_{\text{obs}} - F_{\text{calc}}$) $> 1\sigma(F_{\text{obs}})$ na região do sítio ativo ao final de cada etapa de refinamento: (I) refinamento de corpo rígido; (II) refinamento posicional; (III) "simulated annealing"; (IV) refinamento posicional; (V) refinamento do fator B ; (VI) refinamento do fator B com 143 moléculas de água.

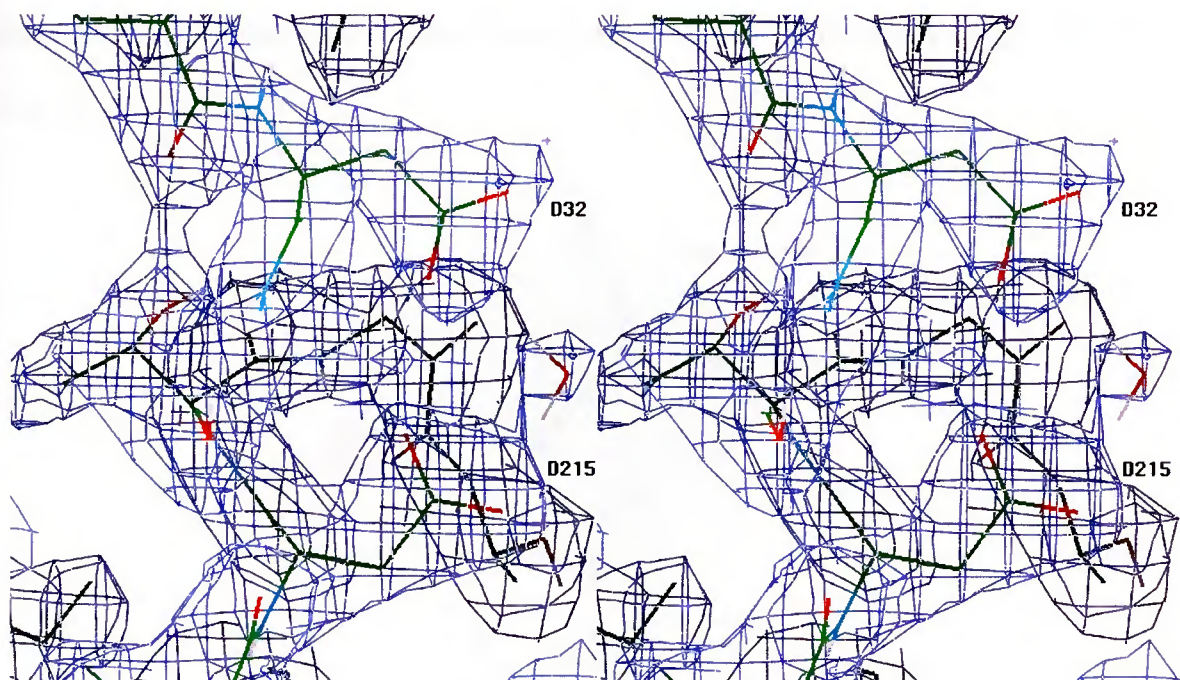


Figura 24. Visão estéreo do mapa de densidade eletrônica do sítio ativo da uropepsina humana, mostrando os dois resíduos de ácido aspártico 32 e 215.

8. Análise da estrutura

O enovelamento da cadeia principal da uropepsina nativa foi analisada através do diagrama de Ramachandran (Ramachandran & Srinivasan, 1961), no qual os ângulos ϕ são graficados contra os ângulos ψ de cada resíduo de aminoácido da proteína (Figura 25).

Pelo gráfico de Ramachandran feito no programa PROCHECK (Laskowski *et al.*, 1993), identificamos dois resíduos de aminoácidos (Asp11 e Asp158) fora das regiões permitidas (Figura 26). Estes aminoácidos, 0,7% do total dos resíduos, pertencem à região generosamente permitida do gráfico, mas pela análise do mapa de densidade eletrônica ($2F_{\text{obs}} - F_{\text{calc}}$), eles estão bem posicionados. Do total, 230 resíduos (84,2%) pertencem às regiões mais favoráveis e 41 resíduos (15,0%)

pertencem às regiões permitidas. Existem na proteína 34 resíduos de glicina e 17 de prolina.

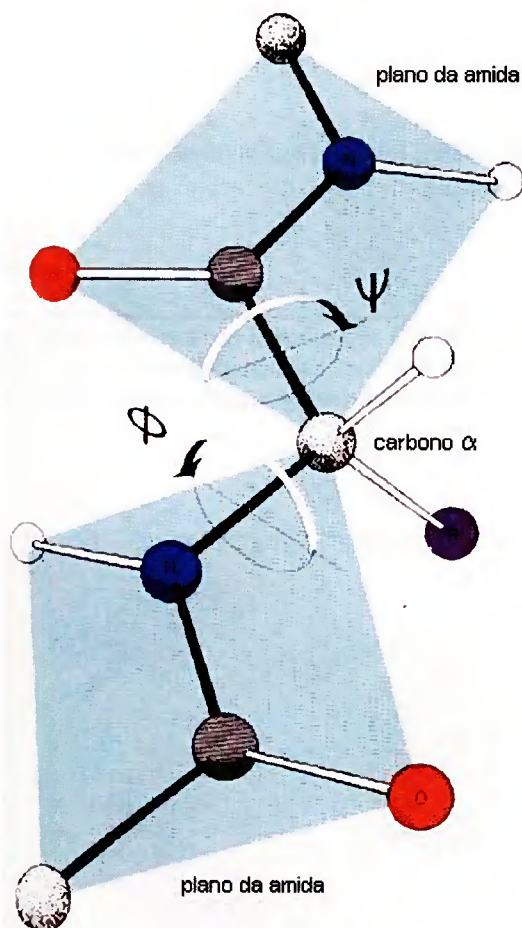


Figura 25. Porção da cadeia polipeptídica indicando os ângulos ϕ e ψ . Esferas vermelhas indicam átomos de oxigênio, esferas em azul, nitrogênio, cinza carbonos, em branco hidrogênio e roxo indica a cadeia lateral.

Diferença de desvios de r.m.s. a partir da geometria ideal para distância entre ligações, ângulos e ângulos diédricos estão apresentados na tabela 9.

Tabela 9. Estatísticas do refinamento para uropepsina humana.

Resolução (Å)	12,0-2,45
R_{factor}^* (%)	16,1
R_{free}^{**} (%)	25,1
Valores do fator B^+ (Å ²)	
Cadeia principal	13,9
Cadeias laterais	15,6
Moléculas de água	29,5
Desvio observado a partir da geometria ideal	
rms, distância entre ligações (Å)	0,009
rms, ângulo (graus)	1,61
rms, ângulos diédros (graus)	27,04
No. de moléculas de água	143

* $R_{\text{factor}} = 100 \times \Sigma |F_{\text{obs}} - F_{\text{calc}}| / \Sigma (F_{\text{obs}})$, soma sobre todas as reflexões com $F/\sigma(F) > 2$.
** $R_{\text{free}} = R_{\text{factor}}$ para 10% dos dados, os quais não estão incluídos no refinamento cristalográfico.
+fator B = valores médios do fator B para todos os átomos (não considerando átomos de hidrogênio).

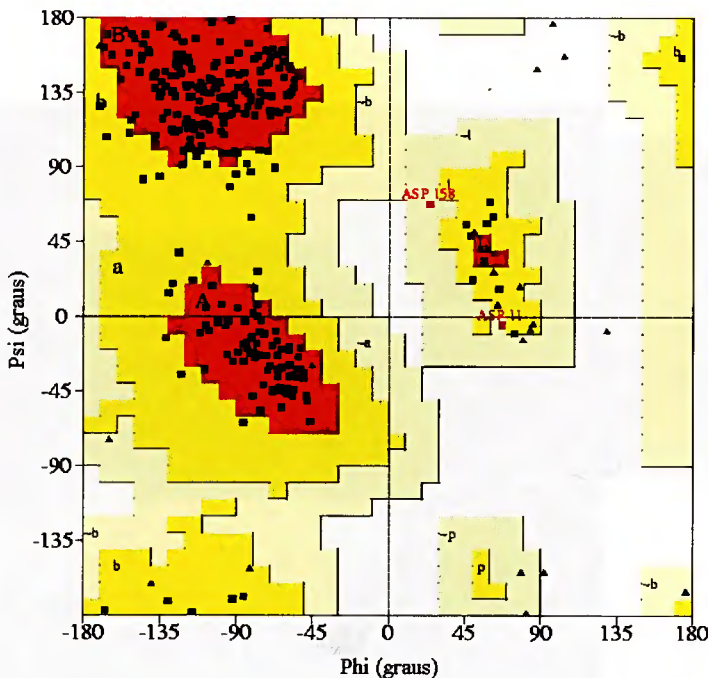


Figura 26. Gráfico de Ramachandran para a uropepsina. As regiões A, B, L são mais favoráveis, as regiões a, b, l, p são permitidas e ~a, ~b, ~l, ~p são regiões generosamente permitidas. Resíduos de glicina estão representados por triângulos.

O modelo refinado da uropepsina é bilobal, consistindo predominantemente de dois lobos de folhas β que, como observado em outras proteinases aspárticas, estão relacionados através de um pseudo eixo binário. A estrutura da uropepsina

humana é análoga à estrutura da pepsina de porco descrita previamente (Cooper *et al.*, 1990). Como foi feito para a análise da pepsina de porco (Sielecki *et al.*, 1990), a estrutura da uropepsina foi dividida em três domínios. O domínio central consiste de seis fitas β antiparalelas que servem como suporte para a região do sítio ativo da molécula. É composto pelos resíduos Val1-Leu6, Asp149-Val184 e Glu308-Ala326. O lobo N-terminal é composto dos resíduos Glu7-Gln148 e o lobo C-terminal pelos resíduos Thr185-Arg307. Os lobos consistem de folhas β empacotadas ortogonalmente com os lobos N- e C-terminal, tendo três e duas fitas, respectivamente. A estrutura e sua divisão em três domínios pode ser observado na Figura 27.

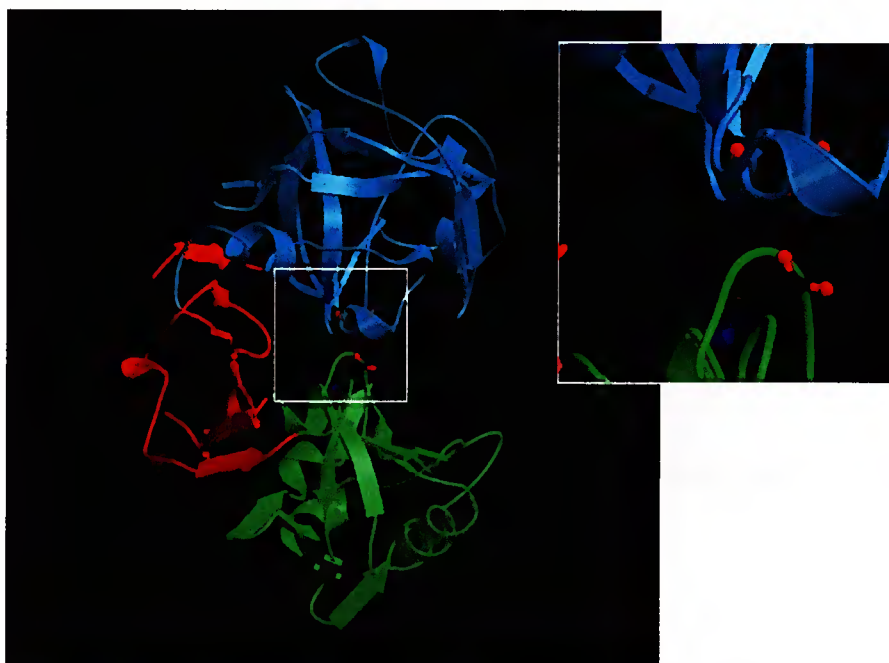


Figura 27. Elementos de estrutura secundária da uropepsina humana mostrando átomos da cadeia principal. Podem ser observados os três domínios: domínio N-terminal (azul); domínio central (vermelho); domínio C-terminal (verde).

A Figura 28 mostra o gráfico do desvio de r.m.s. ("root mean square" – raiz média quadrática) da sobreposição da uropepsina com a pepsina 3A. Este número indica a média da distância entre os carbonos α das duas proteínas sobrepostas. É importante frisar que a seqüência primária da uropepsina é a mesma da pepsina 3A, com exceção do resíduo 291 (Leu $\rightarrow$ Val). Realizamos 3 sobreposições: usando todos os átomos da proteína (exceto os átomos do resíduo 291), a cadeia principal (C, C α , N e O) e usando os carbonos α . O desvio de r.m.s. para a primeira sobreposição foi de 0,737Å, para a cadeia principal foi de 0,528Å e para os carbonos α foi de 0,529Å. As sobreposições foram feitas no programa PROFIT (McLachlan, 1982).

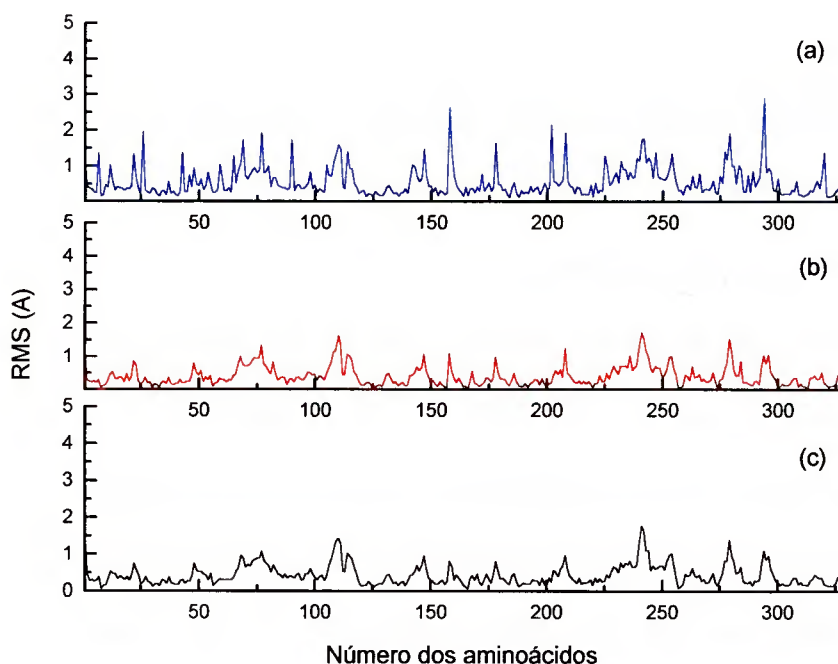


Figura 28. Gráfico mostrando o desvio de r.m.s. após a sobreposição da uropepsina e da pepsina 3A humana. (a) todos os átomos da proteína; (b) carbonos α (c) cadeia principal.

A figura 28a mostra que dois resíduos apresentam desvios de r.m.s. maiores que 2,5 Å. Um deles é o Asp158 (desvio de r.m.s.: 2,64 Å). Este resíduo pertence à região generosamente permitida do gráfico de Ramachandran para a estrutura da uropepsina, entretanto, está dentro do mapa de densidade eletrônica. O outro resíduo é o Glu 294 (desvio de r.m.s.: 2,91 Å). A razão para este valor é que este resíduo está exposto ao solvente, possuindo grande flexibilidade em ambas estruturas.

Ignorando dados à baixa resolução, um gráfico de Luzatti (Luzatti, 1952) nos oferece a melhor correlação entre os dados observados e calculados para um erro médio das coordenadas de 0,19 Å.

9. Parâmetros de Vibração Térmica

A figura 29 ilustra a ampla variação nos valores do fator de temperatura (fator B) médio para ambos, átomos da cadeia principal e lateral do modelo final. O fator B médio para átomos da cadeia principal foi de 15,0 Å², enquanto que para os átomos da cadeia lateral foi de 14,1 Å². Os fatores B para as moléculas de água estão na faixa de 12,2 a 46,8 Å², com um valor médio de 29,5 Å² (Tabela 3).

A estrutura da pepsina 3A foi resolvida à mais alta resolução do que a uropepsina. Muitas regiões da estrutura da pepsina tem altos valores do fator B , indicando regiões desordenadas. As mesmas regiões na estrutura da uropepsina apresentam valores menores de fator B , indicando regiões mais ordenadas da molécula. Isto provavelmente é devido ao menor conteúdo de solvente do cristal da uropepsina (43,3% contra 60,5% da pepsina).

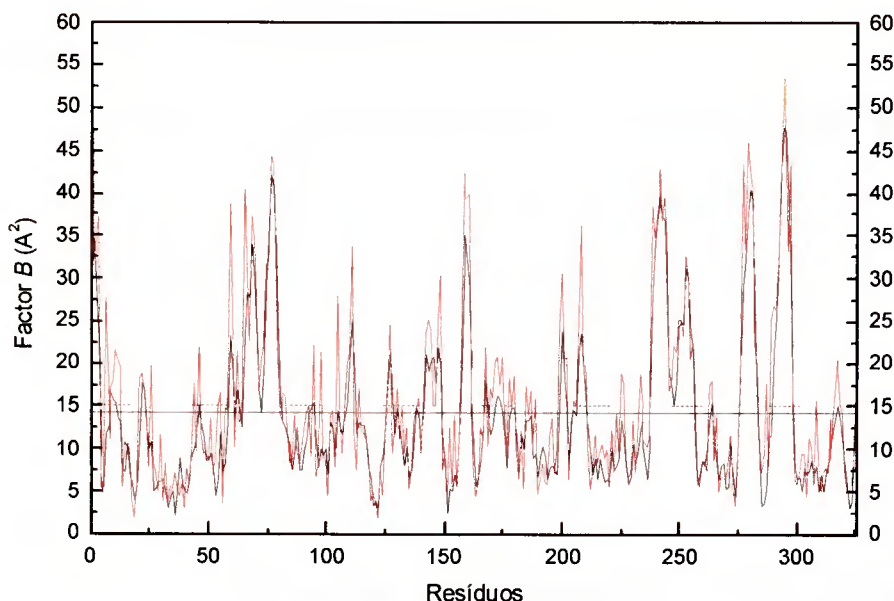


Figura 29. Valores de fator B usando a cadeia principal (linha vermelha) e a cadeia lateral (linha preta) para os 326 resíduos de aminoácidos da uropepsina. Linhas pontilhada e contínua, em preto mostram os valores de fator B médio ($15,0$ e $14,1 \text{ Å}^2$, respectivamente).

10. Contatos intermoleculares e empacotamento do cristal

Foram verificados os contatos intermoleculares no cristal de uropepsina utilizando o programa X-PLOR (Brunger, 1992; Engh & Huber, 1991). Existem muitas ligações de hidrogênio mediadas por moléculas de água. As pontes salinas ou ligações de hidrogênio formadas diretamente entre as moléculas de uropepsina estão listadas na Tabela 10. O mesmo foi feito para a pepsina 3A e não existem contatos comuns entre a uropepsina e a pepsina (Tabela 11). Especialmente interessante é o fato de que ambas enzimas foram cristalizadas no mesmo grupo espacial $P2_12_12_1$, entretanto com diferentes parâmetros de cela. Os parâmetros de cela para a pepsina são $a = 71,97 \text{ Å}$, $b = 151,59 \text{ Å}$ e $c = 41,15 \text{ Å}$, com volume da

cela unitária igual a $449 \times 10^3 \text{ \AA}^3$ e V_m de $3,24 \text{ \AA}^3/\text{Da}$. O conteúdo de solvente é 60,5% e a densidade do cristal é de $1,13 \text{ g cm}^{-3}$.

A Figura 30 mostra o empacotamento dos cristais da uropepsina humana e da pepsina 3A. Existem 4 moléculas na cela unitária de ambas estruturas, e o empacotamento é maior, com um menor conteúdo de solvente no cristal da uropepsina.

Tabela 10. Contatos por ligações de hidrogênio entre a uropepsina humana e seus quatorze vizinhos relacionados por simetria.

Grupo em x,y,z		Grupo relacionado por simetria		Elemento de simetria ^a	Distância entre as ligações (Å)
Glu69	OE1	Cys249	O	2	3,68
Glu69	OE1	Ser250	O	2	3,14
Glu69	OE1	Ser253	OG	2	2,73
Tyr86	OH	Ser248	OG	2	3,41
Ser131	O	Ser250	OG	2	3,92
Asn142	O	Ser241	OG	2	3,74
Glu3	O	Thr51	O	4	3,40
Thr17	OG1	Ser46	O	4	2,82
Ala24	O	Ser110	OG	4	3,21
Gln90	OE1	Tyr113	OH	4	2,63
Tyr175	O	Glu202	OE2	4	3,83
Ser178	OG	Asp234	O	4	3,41
Asn180	OD1	Ser233	O	4	3,28
Ser226	OG	Asp257	OD2	4	2,66

^aOperadores de simetria: 1, x,y,z; 2, -x+1/2,-y,z+1/2; 3, -x,y+1/2,-z+1/2; 4, x+1/2,-y+1/2,-z.

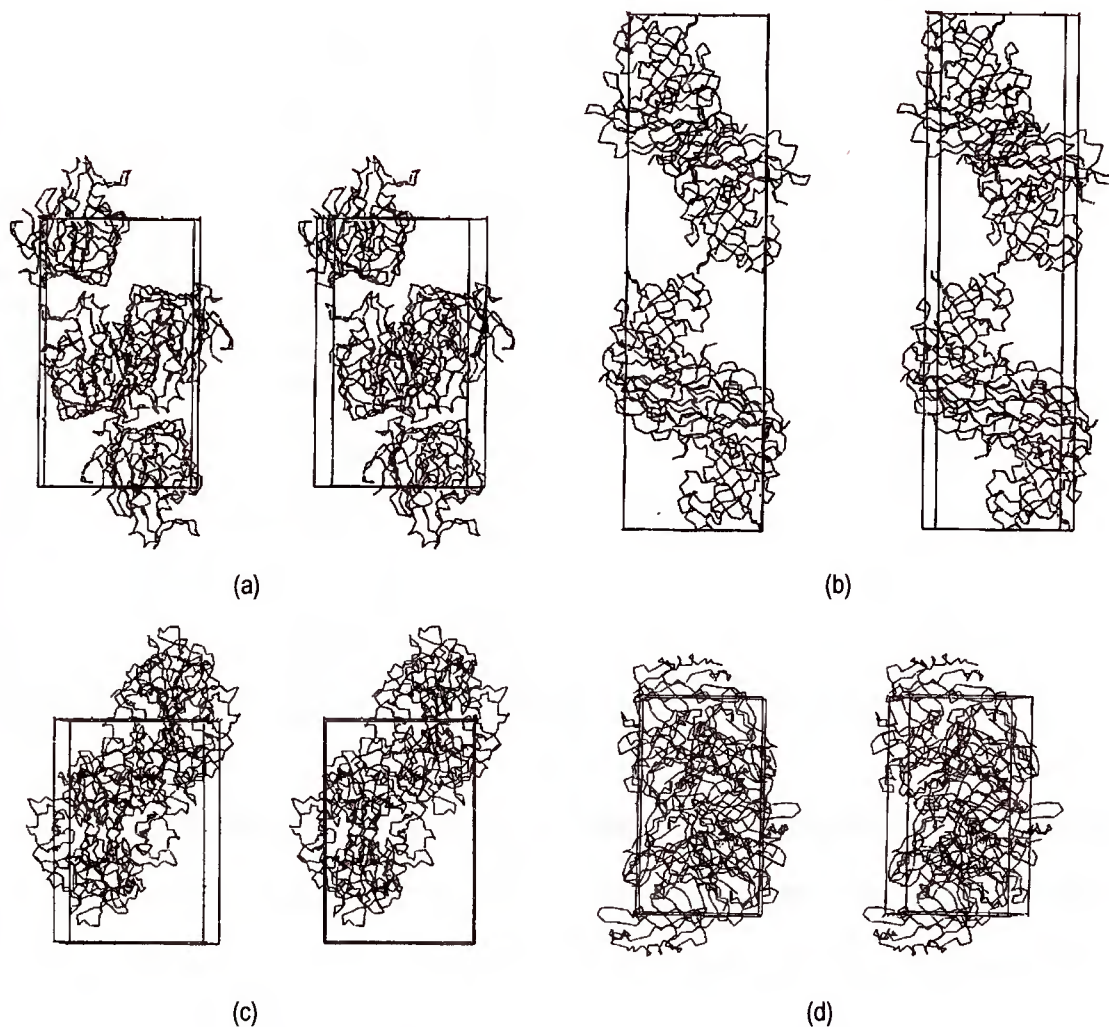


Figura 30. Visão estéreo do empacotamento do cristal de uropepsina (a) ao longo do eixo b e (c) ao longo do eixo c e do cristal de pepsina humana (b) ao longo do eixo a e (d) ao longo do eixo b.

Tabela 11. Contatos por ligações de hidrogênio entre pepsina humana e seus vinte e dois vizinhos relacionados por simetria.

Grupo x,y,z		Grupo relacionado por simetria		Elemento de simetria ^a	Distância entre as ligações(Å)
Glu3	OE1	Ser254	OG	4	3,80
Glu3	OE2	Ser250	O	4	3,80
Gly144	O	Gln266	OE1	4	3,60
Val146	O	Gln266	OE1	4	3,69
Ser147	OG	Thr261	OG1	4	2,90
Ser147	O	Thr198	OG1	4	3,70
Asp171	OD1	Glu208	OE1	4	3,73
Glu279	OE1	Glu294	OE2	1	3,26
Glu279	OE2	Glu294	OE2	1	3,95
Ser46	O	Glu69	OE1	2	3,68
Ser46	O	Thr70	OG1	2	2,23
Ser47	OG	Glu69	OE2	2	2,90
Glu3	OE1	Ser250	O	4	3,26
Asp171	OD1	Ser248	OG	4	3,41
Asp171	OD2	Ser248	OG	4	2,35
Ser172	OG	Glu208	OE1	4	3,73
Ser172	OG	Glu208	OE2	4	2,56
Glu202	OE1	Asp314	OD1	4	3,21
Glu202	OE1	Asp314	OD2	4	3,35
Glu202	OE1	Asn317	OD1	4	3,27
Glu202	OE2	Asp314	OD2	4	3,30
Glu202	OE2	Asn317	OD1	4	3,42

^aOperadores de simetria:: 1, x,y,z; 2, -x+1/2,-y,z+1/2; 3, -x,y+1/2,-z+1/2; 4, x+1/2,-y+1/2,-z.

A estrutura da uropepsina tem um total de 10 contatos intermoleculares comparados a 28 observados para a pepsina humana. Os contatos foram calculados usando o programa DISTANG (CCP4, 1994). O intervalo de distâncias para os contatos intermoleculares foi de 2,5 Å a 3,7 Å, dependendo do tipo de átomo, usando o raio de van der Waals padrão. Os resíduos envolvidos nos contatos são diferentes em ambas estruturas. Os principais resíduos envolvidos nos contatos intermoleculares são: Glu69, Gln90, Tyr113, Ser226, Ser253 e Asp257 para a uropepsina e Glu3, Ser46, Leu48, Ser68, Thr70, Asp171, Ser248 e Ser250 para a pepsina. Esta diferença é devida à uma orientação diferente no empacotamento molecular observado entre as estruturas nos dois cristais.

Resultados e Discussão

11. Comparação com outras proteinases aspárticas

A sequência primária da uropepsina foi comparada à outras proteinases aspárticas humanas, assim como à pepsina porcina (Figura 31). A identidade sequencial entre a uropepsina e outras proteinases são 86,5% para pepsina porcina, 52% para catepsina E humana, 54% para renina humana, 48% para catepsina D humana e 34% para gastricsina humana (pepsina I). A tabela 12 mostra o desvio de r.m.s. dos átomos de carbono α após a sobreposição realizado no programa PROFIT (McLachlan, 1982). Os maiores desvios de r.m.s. foram observados entre uropepsina e catepsina D humana, e entre uropepsina e renina. A similaridade estrutural correlaciona-se à similaridade nas sequências.

Tabela 12. Valores dos desvios de r.m.s. para a sobreposição da uropepsina com outras proteinases aspárticas.

	Pepsina Humana (1psn)	Pepsina Humana (1pso)	Renina Humana (1me)	Catepsina D Humana (1lyb)	Pepsina Porcina (4pep)
Uropepsina					
Desvios r.m.s. (Å)	0,53	0,63	1,54	1,57	0,55
Número de átomos de carbono α sobrepostos	326	326	292	304	326

[illegible]

Figura 31. Seqüência primária da uropepsina comparada as outras proteinases aspárticas: UROP_HUM: uropepsina humana; PEPS_PIG: pepsina porcina; GAST_HUM: gastricsina humana; CATE_HUM: catepsina E humana; CATD_HUM: catepsina D humana; RENI_HUM: renina humana.

12. Interação com pepstatina e sítio de ligação do substrato

A especificidade e a afinidade entre a enzima e seu inibidor dependem das ligações de hidrogênio direcionais e de interações iônicas, assim como da complementaridade da superfície de contato de ambos componentes (de Azevedo *et al.*, 1996; de Azevedo *et al.*, 1997; Kim *et al.*, 1996). A figura 32 ilustra o sítio de ligação da uropepsina, com os resíduos que participam da ligação ao inibidor. Foi observado um total de 14 ligações de hidrogênio entre a uropepsina e a pepstatina, muitas delas envolvendo os aspartatos catalíticos (Asp32 e Asp215). As ligações de hidrogênio padrões entre o inibidor e a enzima são bem conservadas em outros complexos com pepstatina estruturalmente determinados (Suguna *et al.*, 1992; Bailey *et al.*, 1993; Baldwin *et al.*, 1993). As distâncias das ligações de hidrogênio entre Asp32 e Asp215 na uropepsina e Sta404 na pepstatina são compatíveis com o complexo pepsina:pepstatina. Entretanto, as distâncias das ligações de hidrogênio entre Thr77-Val403 e Gly217-Sta404 (uropepsina e inibidor, respectivamente) são maiores do que aquelas observadas para a pepsina (Figuras 12 e 35; Tabela 13). Como observado para outros complexos de inibidores ligados a proteinases aspárticas, a pepstatina adota uma conformação estendida com a hidroxila da primeira estatina ocupando a posição no sítio ativo entre os grupos carboxil do Asp32 e do Asp215.

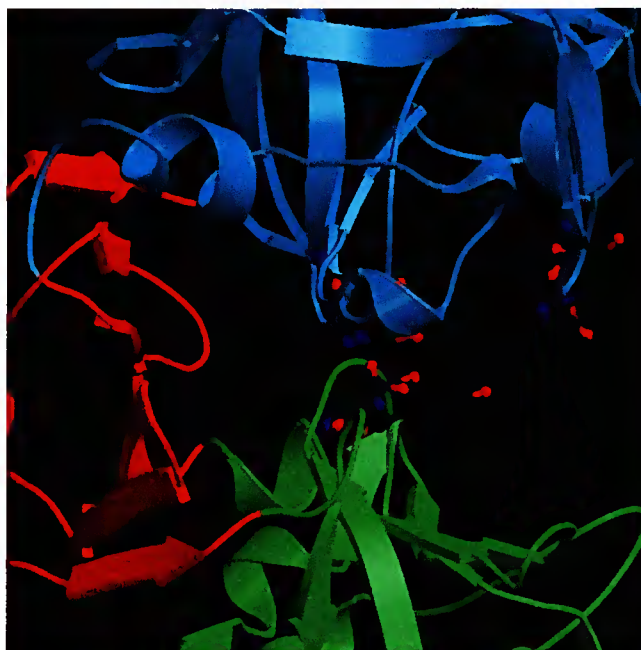


Figura 32. Sítio de ligação da uropepsina com os resíduos de aminoácidos que participam da catálise enzimática.

Tabela 13. Ligações de hidrogênio intermoleculares da pepsina e da uropepsina com a pepstatina. Alguns valores da distância estão em *itálico*. Indicando que não existem ligações de hidrogênio nestas posições na estrutura da uropepsina. Está presente para comparação.

Ligações de hidrogênio entre o sítio ativo e o inibidor				Distância (Å)		Diferença das distâncias (Å)
Pepstatina		Enzima		Uropepsina	Pepsina	
Sta404	OH	Asp215	O ^{δ2}	2,69	2,69	0
		Asp215	O ^{δ1}	2,71	2,96	0,25
		Asp32	O ^{δ2}	3,10	3,36	0,26
		Asp32	O ^{δ1}	2,71	2,68	0,03
Sta404	N	Gly217	O	3,49	2,95	0,54
Sta404	OH	Gly217	O	4,05	3,52	0,53
Ala405	N	Gly34	O	3,21	2,91	0,30
Ala405	O	Tyr189	O ^η	3,24	2,66	0,58
Val402	N	Ser219	O ^γ	2,76	2,98	0,22
Val402	O	Ser219	N	2,75	2,99	0,24
Val403	N	Thr77	O ^γ	5,33	3,08	2,25
Val403	O	Thr77	N	4,12	2,95	1,17
		Gly76	N	3,54	2,99	0,55
Sta404	O	Gly76	N	3,15	2,92	0,23
Sta406	N	Thr74	O	3,14	2,96	0,18

13. Potencial eletrostático

O potencial eletrostático da superfície da uropepsina nativa e do modelo na presença de pepstatina foram calculados com GRASP (Nicholls *et al.*, 1991). O mesmo foi feito para a pepsina 3A nativa e inibida. As duas superfícies moleculares foram comparadas considerando as coordenadas das proteínas nativas e inibidas. Existe uma mudança conformacional na estrutura quando o inibidor liga-se ao sítio ativo. A mudança é relativamente pequena, com uma diferença de desvio de r.m.s. de 0,33Å nas coordenadas de todos os átomos de carbono α depois da sobreposição para a pepsina 3A e de 0,44Å para uropepsina. Esta mudança pode ser claramente vista como um movimento relativo dos domínios quando o inibidor está ligado (Fujinaga *et al.*, 1995) (Figura 33).

As estruturas da uropepsina e da pepsina 3A são negativamente carregadas. As estruturas contêm apenas 1 resíduo de histidina, 3 resíduos de arginina e nenhum resíduo de lisina. O sítio ativo é fortemente negativo, como mostrado na figura 34.

Os subsítios de ligação S_4 a S_3' são definidos pelas interações dos resíduos P_4 a P_3' do inibidor com a enzima. É improvável que existam sítios de ligação adicionais além destes sítios. O N da cadeia principal do resíduo P_3' forma uma ligação de hidrogênio com Thr74. O sítio S_4 é um plano muito acessível ao solvente. Os subsítios S_1 e S_3 são contíguos, com o átomo de O da carbonila da Gly217 fazendo uma separação dos dois sítios. O subsítio S_1 tende a ser hidrofóbico, enquanto que S_3 é principalmente polar. Os subsítios S_2 e S_1' são principalmente hidrofóbicos. O subsítio S_2' é claramente definido pelo resíduo de alanina P_2' . A cadeia principal do inibidor faz uma ligação de hidrogênio com o

átomo de O da Gly34 e aceita uma ligação de hidrogênio do OH da Tyr189 em S₂' (Figura 8, Tabela 14) (Fujinaga *et al.*, 2000).

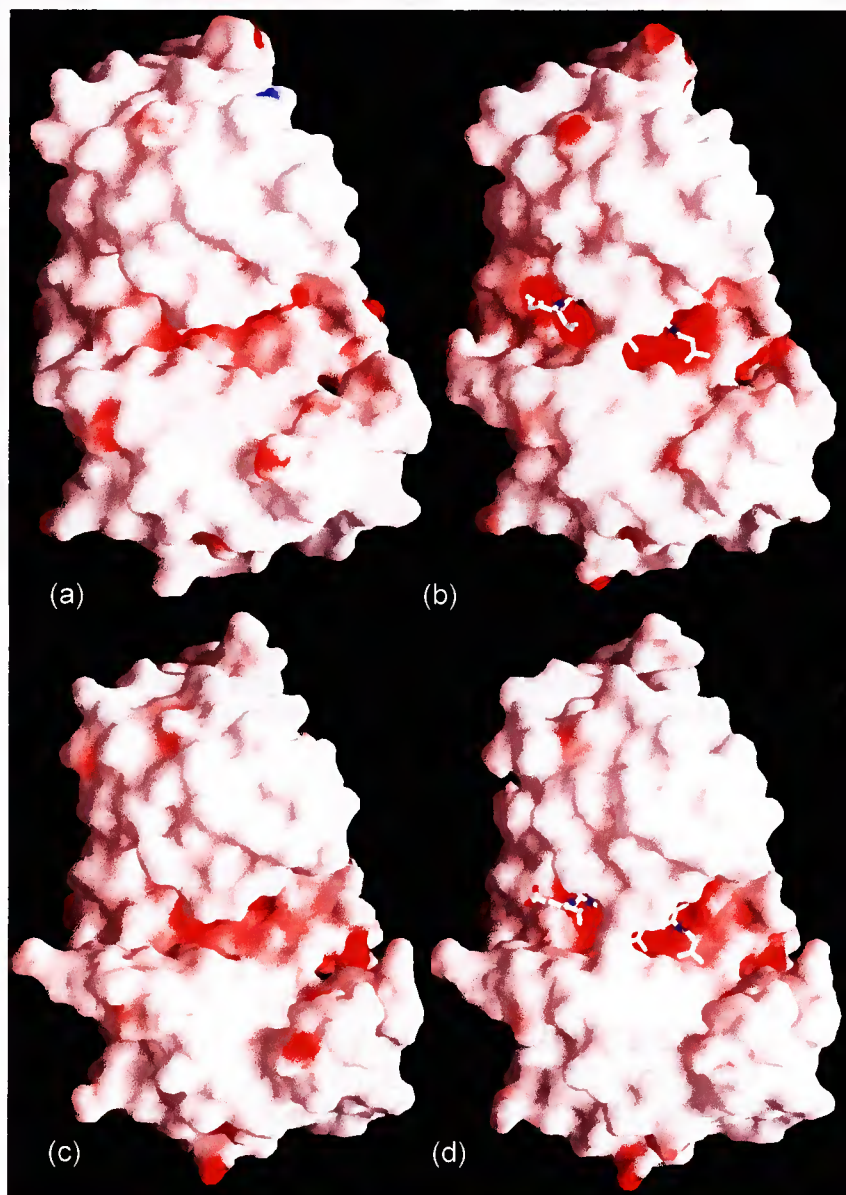


Figura 33. Potencial eletrostático da superfície da pepsina. (a) sem o inibidor e (b) com o inibidor, e da uropepsina (c) sem o inibidor e (d) com o inibidor, calculado no programa GRASP (-50kT: vermelho e +50kT: azul). Regiões não carregadas estão em branco.

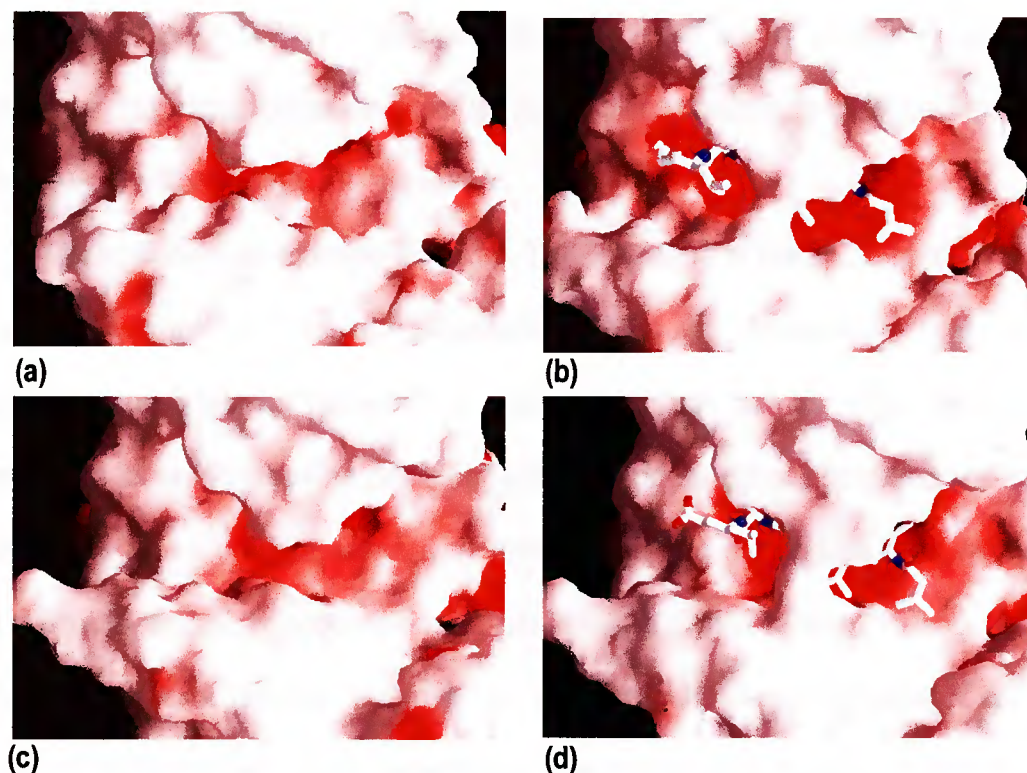


Figura 34. Potencial eletrostático da região do sítio catalítico da superfície da pepsina. (a) sem o inibidor e (b) com o inibidor, e da uropepsina (c) sem o inibidor e (d) com o inibidor. Além da mudança conformacional, podemos observar que a região é negativamente carregada.

Tabela 14. Resíduos envolvidos nos subsítios de especificidade da uropepsina humana.

Sítio	Uropepsina
S ₅	-
S ₄	F111, S219
S ₃	E13, V30, F111, F117, G217, T218, S219
S ₂	G76, T77, G217, T218, T222, Q287, M289
S ₁	V30, D32, Y75, T77, F111, F117, D215, G217, T218
S ₁ '	G34, G76, Y189, I213, D215, M289, V291, I300
S ₂ '	G34, T74, I128, Y189
S ₃ '	T74, Y189, V291

A única diferença observada entre a pepsina 3A e a uropepsina é a substituição Leu→Val na posição 291, localizada em S₃'. Contudo, esta substituição presente no sítio de ligação não afeta os valores de k_{cat} , desde que os valores

determinados para uropepsina e pepsina, usando o mesmo substrato são $6,01 \pm 0,11 \text{ s}^{-1}$ e $5,92 \pm 0,21 \text{ s}^{-1}$, respectivamente. Além disso, a substituição Leu→Val mantém a hidrofobicidade do sítio S_3' e a posição adotada pela cadeia lateral da valina não afeta a ligação do substrato.

As ligações de hidrogênio entre Thr77-Val403 e Gly217-Sta404 observadas para o complexo pepsina:pepstatina não é observado no complexo uropepsina:pepstatina (Tabela 13). Entretanto, um exame mais detalhado dos complexos binários pepsina:pepstatina e do modelo uropepsina:pepstatina mostra claramente que o inibidor adota a mesma orientação em ambos complexos (Figura 35). Além disso, as ligações de hidrogênio entre Asp32 e Asp215 e o inibidor são conservados em ambos complexos. Esta similaridade estrutural confirma a similaridade funcional contra pepstatina, observado entre as duas enzimas (Gomes R.A.S. comunicação pessoal).

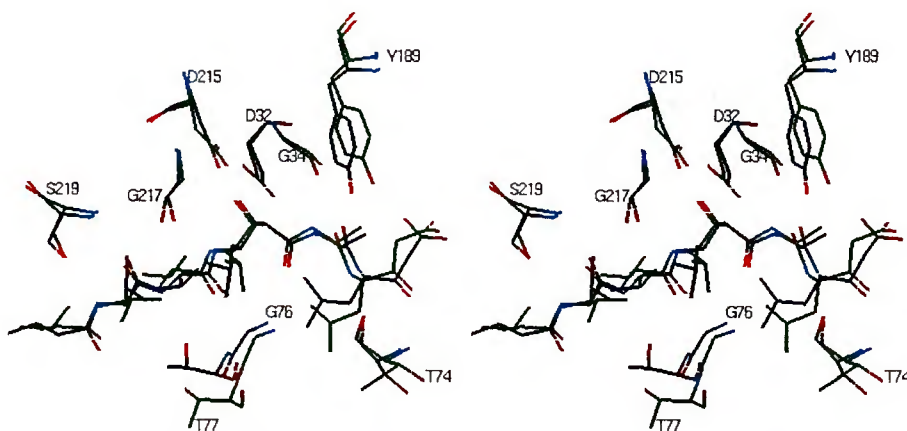


Figura 35. Representação estéreo do sítio de ligação da uropepsina (em verde) sobreposto ao sítio de ligação da pepsina 3A (em cinza) na presença da pepstatina.

Conclusões

CONCLUSÃO

A tese aqui apresentada visou a análise estrutural da uropepsina, membro da família das proteinases aspárticas, através das técnicas de cristalografia de proteínas, com o objetivo de se estabelecer uma correlação entre a estrutura e a sua função biológica.

Tentativas de cristalizar a pepsina I e a uropepsina na presença de pepstatina, não resultaram em bons cristais e a possível razão para este fato se deve à facilidade que as proteinases aspárticas possuem para auto-hidrólise.

Os nossos estudos cristalográficos com a enzima cristalizada no seu estado nativo foram importantes no sentido de que mais uma isoforma da pepsina humana foi elucidada, a mutação Leu→Val na posição 291. Isoformas de pepsina são de interesse potencial devido à sua evolução molecular. O pepsinogênio A pertence a uma família em que todos seus variantes derivaram de um gene ancestral comum e recente. Consequentemente, os variantes assemelham-se muito, podendo ocorrer poucas mutações na estrutura primária depois da duplicação dos genes.

A pepsina é a principal enzima gástrica, e suas isoformas, apesar de serem produzidas no estômago, nem todas permanecem ali quando na sua forma ativa. A urina contém relativamente grande quantidade de pepsinogênio A.

Através da modelagem e comparações com o complexo pepsina:pepstatina, verificamos que a cavidade do sítio ativo é conservado e as ligações com o inibidor também obedecem o mesmo padrão. Isto é confirmado

Conclusão

quando comparamos os valores obtidos da eficiência catalítica (k_{cat}) para ambas as enzimas, pepsina e uropepsina. Estudos futuros serão realizados na tentativa de resolver a estrutura do complexo uropepsina-pepstatina e investigar sua variação conformacional em relação ao pH.

Somente com a visão integrada de todos os aspectos possíveis é que poderemos descrever de maneira mais precisa o que ocorre na natureza. E a cristalografia de proteínas vem justamente de encontro a preencher a lacuna da descrição estrutural das macromoléculas biológicas, possibilitando uma visão holística dos processos biológicos envolvendo as proteínas.

Conclusão

Referências Bibliográficas

REFERÊNCIAS BIBLIOGRÁFICAS

- Andreeva, N. S., Zdfanov, A. S., Gustchina, A. E., Federov, A. A., (1984). *J. Biol. Chem.*, **259**, 11353-11364.
- Anson, M. L. (1939). *J. Gen. Physiol.* **22**, 79-89.
- Azároff, L. V. (1968). *Elements of X-ray crystallography* McGraw-Hill, Inc. USA.
- Bailey, D., Copper, J. B., Veerapandian, B., Blundell, T. L., Atrash, B., Jones, D. M., Szelke, M. (1993). *Biochem. J.* **289**, 363-371.
- Baldwin, E. T., Bhat, T. N., Gulnik, S., Hosur, M. V., Sowder, R. C. II, Cachau, R. E., Collins, J., Silva, A. M., Erickson, J. W. (1993). *Proc. Natl. Acad. Sci.* **90**, 6796-6800.
- Bayliss, J. R. K. & Wybrandt, G. B. (1969). *Biochem. J.* **113**, 377.
- Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*. Academic Press Inc. (London) Ltd., pp. 151-164.
- Blundell, T. L., Foundling, S. I., Hemmings, A. M., Pearl, L. H., Sibanda, B. L., Watson, F. E., Wood, S. P., Pedersen, V. B. (1984). in *Aspartic Proteinases and Their Inhibitors*, (Kostka, V. eds.), pp. 151-161, Walter de Gruyter, Berlin, New York.
- Boistelle, R. & Astier, J. P. (1988). *J. Crystal Growth*, **90**, 14.
- Bott, R., Subramanian, E., Davies, D. R. (1982). *Biochemistry*, **21**, 6956-6962.
- Brünger, A. T. (1992). *X-PLOR Version 3.1: A system for crystallography and NMR*. Yale University Press, New Haven.



- Brünger, A. T., Adams, P. D., Marius Clore, G., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. -S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., Warren, G. L. (1998). *Acta Cryst.* **D54**, 905-921.
- Campbell, M. K. (2000). *Bioquímica* ArtMed Editora, 3ª edição, 752p. Porto Alegre, RS, Brasil.
- Canduri, F., Teodoro, L. G. V. L., Lorenzi, C. C. B., Gomes, R. A. S., Fontes, M. R. M., Arni, R. K. & de Azevedo, W. F., Jr. (1998). *Biochemistry and Molecular Biology International*, **46**, 2, 355-363.
- Cantor, R. C. & Schimmel, P. R. (1980). *Biophysical Chemistry*, Vol. II, W.H. Freeman & Company. USA.
- Carter, C. W., Jr. & Carter, C. W. (1979). *J. Biol. Chem.*, **254**, 12219.
- CCP4 (1994). *Acta Cryst.* **D50**, 760-763.
- Cooper, J. B., Khan, G., Taylor, G., Tickle, I. J., Blundell, T.L. (1990). *J. Mol. Biol.* **214**, 199-222.
- Cullity, B. D. (1956). *Elements of X-ray crystallography*, Addison-Wesley Publishing Company, Inc. USA.
- de Azevedo, W. F. Jr., Mueller-Dieckmann, H. -J., Schulze-Gahmen, U, Worland, P. J., Sausville, E., Kim, S. -H. (1996). *Proc. Natl. Acad. Sci. USA.* **93**, 2735-2740.
- de Azevedo, W. F., Jr. (1997). Tese de Doutorado: *Introdução à biocristalografia com o estudo estrutural da quinase dependente de ciclina 2 (CDK2) complexada com inibidores*. IFSC-USP São Carlos-SP.
- Denburg, J. L., Nelson, R., Silver, M. S. (1968). *J. Am. Chem. Soc.* **90**, 479.

Referências Bibliográficas



- Drenth, J. (1994). *Principles of Protein X-ray Crystallography*. Springer-Verlag. New York, Inc.USA.
- Ducruix, A. & Giegé, R. (1992). *Crystallization of Nucleic Acids and Proteins: A Practical Approach* (Rickwood, D. & Hames, B.B. eds.), 331p. IRL Press, New York.
- Engh, R. A. & Huber, R. (1991). Accurate Bond and Angle Parameters for X-ray Protein-Structure Refinement, *Acta Cryst.* **A47**, 392-400.
- Etherington, D. J. & Taylor, W. H. (1967). *Nature* **216**, 279-280.
- Etherington, D. J., Roberts, N. B., Taylor, W. H. (1980). *Clin. Sci.* **58**, 30P.
- Frants, R., Pronk, J. C., Pals, G., Defize, J., Westervend, B. D., Meuwissen, S. G. M., Kreuning, J., Eriksson, A.W. (1984). *Hum. Genet.*, **65**, 385-390.
- Foltmann, B. (1981). *Essays Biochem.*, **17**, 52-84.
- Foltmann, B. (1988). *FEBS LETT.* **241**, 69-72.
- Fujinaga, M., Chernaia, M. M., Tarasova, N. I., Mosimann, S. C. & James, M. N. G. (1995). *Protein Sci.*, **4**, 960-972.
- Fujinaga, M., Cherney, M. M., Tarasova, N. I., Bartlett, P. A., Hanson, J. E., James, M. N. G. (2000). *Acta Cryst.* **D56**, 272-279.
- Garrett, R. H. & Grisham, C. M. (1995). *Biochemistry*, Saunders College Publishing, 1100p. USA.
- Gedde-Dahl, T. Jr., Korsnes, L., Thorsby, E., Olaisen, B, Bratlie, A., Siverts, A. (1978). *Cytogenet. Cell Genet.*, **22**, 301-303.
- Gewirth, D. (1995). *The HKL Manual*. Yale University. New haven Connecticut 06511 USA.

- Gomes, R. A. S., Chagas, J. R., Juliano, L., Hial, V. (1996). *Immunopharmacology* **32**, 76-79.
- Gomes, R. A. S. (1999-2000). comunicação pessoal.
- Gordon, A. H. (1975). Electrophoresis of proteins in polyacrilamide and starch gels, in Work, T.S. and Work, E. (Eds), *Laboratory Techniques in Biochemistry and Molecular Biology*, vol.1, Part 1, revised edition, North-Holland.
- Green, A. A. (1932). *J. Biol. Chem.*, **95**, 47.
- Green, M. L., Valler, M. J., Kay, J. (1984). *J. Dairy Res.*, **51**, 331-340.
- Hartsuck, J. A. & Tang, J. (1972). *J. Biol. Chem.* **247**, 2575.
- Henderson, P. J. (1972). *Biochem. J.* **127**, 2, 321-333.
- Higashi, T. J. (1990). *J. Appl. Cryst.*, **23**, 253-257.
- Hirsch-Marie, H. (1968). *Biol. Gastroenterology*, **2**, 109-122.
- Hirsch-Marie, H. & Conte, M. (1969). *Rev. Franc. Etudes Clin. Biol.*, **14**, 977-983.
- Ichinose, M, Miki, K., Furihata, C., Kageyama, T., Hayashi, R., Niwa, H., Oka, H., Matsushima, T. & Takahashi, K. (1982). *Clin. Chim. Acta*, **126**, 183-191.
- James, M. N. G. & Sielecki, A. R. (1983). *J. Mol. Biol.*, **163**, 299-361.
- James, M. N. G., Sielecki, A. R., Hofmann, I. (1985). in *Aspartic Proteinases and Their inhibitors*, (Kostka, V., eds.) pp. 163-177, Walter de Gruyter, Berlin, New York.
- James, M. N. G. & Sielecki, A. R. (1986). *Nature* **319**, 33-38.
- Jancarik, J. & Kim, S. -H. (1991). *J. Appl. Cryst.* **24**, 409-411.
- Jones, T.A., Zou, J. -Y., Cowan, S.W. (1991). *Acta Cryst.* **A47**, 110-120.
- Kageyama, T. & Takahashi, K. (1982). *J. Biochem.* **92**, 1179-1188.

Referências Bibliográficas



- Kageyama, T. & Takahashi, K. (1987). *Eur. J. Biochem.* **165**, 483-490.
- Kay, J. (1982). *Biochem. Soc. Trans.*, **10**, 277-280.
- Kim, S. –H. Schulze-Gahmen, U., Brandsen, J., de Azevedo, Jr. W. F. (1996). *Progress in Cell Cycle Research*, 2, 137-145. Meijer, L., Guidet, S. & Vogel, L. eds. Cap. 14. Plenum Press, N. Y., USA.
- Korsnes, L. & Gedde-Dahl, T. Jr. (1980). *Ann. Hum. Genet.*, **43**, 199-212.
- Kushner, I., Rapp, W., Burtin, P. (1964). *J.Clin. Invest.*, **43**, 1983-2002.
- Laemmli, U. K. (1970). *Nature (London)* **227**, 680-685.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., Thornton, J. M. (1993). *J. App. Cryst.*, **26**, 283.
- Lee, W. H. (2001). Tese de Doutorado: *Estudos cristalográficos de macromoléculas biológicas: 3-dehidroquinase de Salmonella typhi e duas fosfolipases A₂ extraídos do veneno de Bothrops pirajai*. UNICAMP, Instituto de Biologia, Depto de Genética e Evolução Campinas-SP.
- Lehninger, A. L. (1985). *Princípios de Bioquímica* Sarvier Ed. 725p., São Paulo, Brasil.
- Lipson, H. & Taylor, C. A. (1958) *Fourier Transforms and X-ray Diffraction*. G. Bell & Sons Ltd. Glasgow. Great Britain.
- Luzatti, P. V. (1952). *Acta Cryst.* **5**, 802-810.
- Marciniszyn, J. Jr., Huang, J. S., Hartsuck, J. A., Tang, J. (1976). *J. Biol. Chem.*, **251**, 7095-7102.
- ©Martin, A. C. R., SciTech Software, 1992-8.



- Marzzoco, A. & Torres, B. B. (1999). *Bioquímica Básica*, 2^a ed. 360p. Guanabara-Koogan, Rio de Janeiro, Brasil.
- Mathews, B. W. (1968). *J. Mol. Biol.*, **33**, 491-497.
- McLachlan, A. D. (1982). *Acta Cryst.*, **A38**, 871-873.
- McPherson, A., Koszelak, S., Axelrod, H., Day, J., Williams, R., Robinson, L., McGrath, M., Cascio, D. (1986). *J. Biol. Chem.*, **261**, 1969.
- McPhie, P. (1974). *Biochem. Biophys. Res. Commun.*, **56**, 789-792.
- McRee, D. E. (1993). *Practical Protein Crystallography*. Academic Press, 386p.
- Mikol, V. & Giegé, R. (1989). *J. Crystal Growth*, **97**, 324.
- Minamiura, N., Ito, K., Kobayashi, M., Kobayashi, O., Yamamoto, T. (1984). *J. Biochem.*, **96**, 1061-1069.
- Navaza, J. (1994). *Acta. Cryst.*, **A50**, 157-163
- Nicholls, A., Sharp, K., Honig, B. (1991). *Proteins Struct. Funct. Genet.* **11**, 281.
- Pearl, L. (1985). in *Aspartic Proteinases and Their Inhibitors* (Kostka, V. eds.), pp. 189-195, Walter de Gruyter, Berlin, New York.
- Pearson, J. P., Roberts, N. B., Taylor, W. H. (1984). *Clin. Sci.* **66**, 61P.
- Powell, M. J. D. (1977) *Mathematical Programming*, **12**, 241.
- Powers, J. C., Harley, A. D., Myers, D. V. (1977). *Acid Proteinases, Structure, Function & Biology* (Tang, J. ed), pp. 141-157, Plenum, N.Y.
- Raab, W. P. (1972). *Clinical Chemistry* **18**, 1, 5-25.
- Ramachandran, G. N. & Srinivasan, R. (1961). *Nature*, **190**, 159-161.
- Rich, D.H., Bernatowicz, M.S., Agarwal, N.S., Kawai, M., Salituro, F.G., Schmidt, P.G. (1985). *Biochemistry*, **24**, 13, 3165-3173.



- Riès-Krautt, M. M. & Ducruix, A. (1989). *J. Biol. Chem.*, **264**, 745.
- Samloff, I. M. (1969). *Gastroenterology*, **57**, 659-669.
- Samloff, I. M. & Townes, P. L. (1970). *Gastroenterology*, **58**, 462-469.
- Samloff, I. M. (1971). *J. Immunol.*, **106**, 962-968.
- Samloff, I. M. & Liebman, W. M. (1972). *Clin. Exp. Immunol.*, **11**, 405-414.
- Samloff, I. M. (1982). *Gastroenterology*, **82**, 26-33.
- Seijffers, M. J., Segal, H. L. & Miller, L. L. (1963). *Am. J. Physiol.*, **205**, 1099-1105.
- Seijffers, M. J., Segal, H. L. & Miller, L. L. (1964). *Am. J. Physiol.*, **206**, 1106-1110.
- Seijffers, M. J., Miller, L. L. & Segal, H. L. (1965). *Proc. Soc. Exp. Biol. Med.*, **118**, 405-409.
- Schechter, I. & Berger, A. (1967). *Biochem. Biophys. Res. Commun.*, **27**, 157-162.
- Sham, H. L., Zhao, C., Stewart, K. D., Betebenner, D. A., Lin, S., Park, C.H., Kong, X. P., Rosenbrook, W. Junior, Herrin, T., Madigan, D., Vasavanonda, S., Lyons, N., Molla, A., Saldivar, A., Marsh, K. C., McDonald, E., Wideburg, N. E., Denissen, J. F., Robins, T., Kempf, D. J., Plattner, J. J., Norbeck, D. W. (1996). A Novel, Picomolar Inhibitor of Human Immunodeficiency Virus Type 1 Protease. *J. Med. Chem.*, **39**, 392.
- Sherwood, D. (1976). *Crystal, X-rays and Proteins*. Longman Group Ltd. London, Great Britain.
- Shewale, J. G., Takahashi, T. & Tang, J. (1985). *Aspartic Proteinases and their Inhibitors* (Kostka, V., ed.) Walter de Gruyter, Berlin, New York.
- Sielecki, A. R., Fedorov, A. A., Boodhoo, A., Andreeva, N. S. & James, M. N. G. (1990). *J. Mol. Biol.* **214**, 143-170.

- Suguna, K., Padlan, E. A., Bott, R., Boger, J., Parris, K. D. & Davies, D. R. (1992). *Proteins Struct. Funct. Genet.* **13**, 195-205.
- Taggart, R. T., Yu, P. L., Karn, R. C., Conneally, P. M. & Merritt, A. D (1978). *Cytogenet. Cell Genet.*, **22**, 335-340.
- Taggart, R. T., Karn, R. C., Merritt, A. D., Yu, P. L. & Conneally, P. M. (1979). *Hum. Genet.*, **52**, 227-238.
- Tang, J.; James, M. N. G.; Hsu, I. N.; Jenkins, J. A. & Blundell, T. L. (1978). *Nature*, **271**, 618-621.
- Usha, R., Johnson, J. E., Moras, D., Thierry, J. -C., Fourme, R., Kahn, R. (1984). *J. Appl. Cryst.*, **17**, 147.
- Voet, D. & Voet, J. G. (1995). *Biochemistry*, John Wiley & Sons 2^a ed., 1360p., New York, USA.
- von Hippel, P. H. & Schleich, T. (1969). *Accounts of Chemical Research*, **2**, 257.
- Walker, V. & Taylor, W. H. (1980). *Gut* **21**, 766-771.
- Woolfson, M. M. (1970). *An Introduction to X-ray Crystallography*. Cambridge University Press, England.
- Zinchenko, A. A., Rumsh, L. D. & Antonov, V. K. (1976). *Bioorg. Chem. (USSR)* **2**, 803.



Apêndice I

Uma visão geral sobre cristalografia de macromoléculas biológicas	95
Difração de raios X	114
Substituição molecular	125
Refinamento das coordenadas	130
Construção do modelo e validação	134

UMA VISÃO GERAL SOBRE CRISTALOGRAFIA DE MACROMOLÉCULAS BIOLÓGICAS

1. Cristalização de macromoléculas biológicas

Uma macromolécula biológica é um polímero de aminoácidos ou nucleotídeos, os quais estão enovelados em uma estrutura tridimensional, principalmente devido a: (1) interações dipolo-dipolo, ligações de hidrogênio e interações de van der Waals; (2) algumas ligações covalentes; (3) eventualmente, pontes salinas entre resíduos carregados.

Como o próprio nome cristalografia sugere, o material em estudo precisa estar na forma de um cristal para ser analisado. As macromoléculas biológicas podem formar um arranjo cristalino, ou seja, um retículo altamente regular e periódico, com unidades fundamentais idênticas – as celas unitárias.

1.1. Principais diferenças entre cristais de macromoléculas biológicas e cristais minerais

Nos cristais minerais, cada unidade é formada por um ou poucos átomos, enquanto nos cristais de macromoléculas cada unidade inclui uma ou várias moléculas inteiras (cada uma composta de muitos milhares de átomos), podendo ter parâmetros de cela unitária de até algumas dezenas de milhares de ângstrons (Usha *et al.*, 1984). Cristais de macromoléculas biológicas são normalmente bem menores, dificilmente excedem 1mm^3 , e são muito mais frágeis. A fragilidade destes cristais é uma consequência das interações fracas entre as macromoléculas dentro da rede cristalina e do alto conteúdo de solvente nestes cristais (de 20% a mais de 80%). As ligações fracas que mantêm a estrutura



cristalina destas macromoléculas são do tipo hidrofóbica e ligação de hidrogênio. Por esta razão, estes cristais devem ser mantidos em um ambiente saturado de solvente, caso contrário, a desidratação conduzirá à quebra destas ligações fracas, e conseqüente destruição dos cristais. O alto conteúdo de solvente, contudo, tem conseqüências úteis, pois os canais de solvente nos cristais de macromoléculas biológicas permitem a difusão de pequenas moléculas, uma propriedade usada na preparação de derivados isomorfos e de complexos binários de proteínas e ligantes (Ducruix & Giegé, 1992).

1.2. Técnicas de cristalização de macromoléculas biológicas

Quatro métodos são utilizados para a cristalização de macromoléculas biológicas (Ducruix & Giegé, 1992):

- a. Diálise;
- b. Difusão em interface livre;
- c. Difusão de vapor;
- d. "Batch".

A seguir são discutidos brevemente cada método.

a. Diálise

O princípio do método de diálise é bem conhecido pelos bioquímicos: o equilíbrio de pequenos solutos via membrana semi-permeável. A vantagem sobre os outros métodos é sua reversibilidade: se as proteínas precipitam ou formam cristais de forma indesejável, podem ser redissolvidos por diálise contra uma solução sem precipitante. Isto também é útil quando o processo de cristalização

requer uma mudança de pH. Por outro lado, precipitantes de peso molecular grande não podem ser utilizados.

b. Difusão em interface livre

Esta técnica, normalmente realizada em pequena escala, não é normalmente utilizada em experimentos de “screening”. Consiste de dois solventes de diferentes densidades, nas quais a macromolécula não é igualmente solúvel. Os componentes da solução de cada camada difundem lentamente entre as camadas, permitindo as condições de supersaturação para a macromolécula na região intermediária, e os cristais se formam nessa região. A concentração de proteína é usualmente maior (50mg/ml) e deve-se tomar cuidado para evitar a nucleação heterogênea por poeira ou outras partículas.

c. Difusão de vapor

Três variações do método são utilizadas na prática: “hanging”, “sitting” e “sandwich” “drop” (gota pendurada, gota sentada e gota sanduíche) (Figura 1). A técnica de “hanging drop” é frequentemente a mais utilizada.

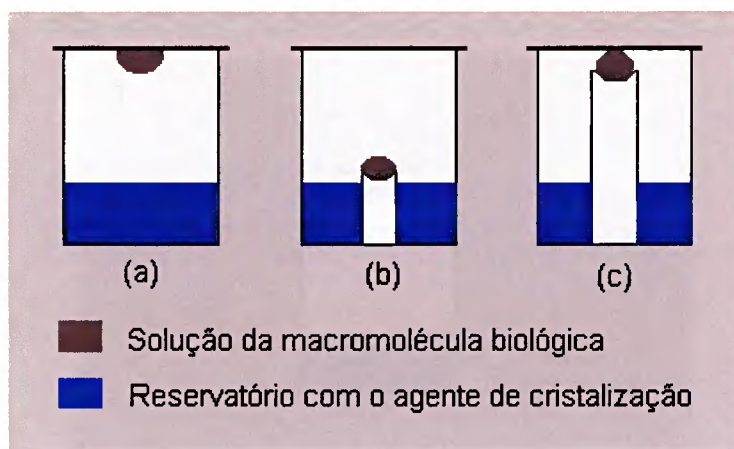


Figura 1. Representação esquemática de (a) gota pendurada ("hanging drop"), (b) gota sentada ("sitting drop") e (c) gota sanduíche ("sandwich drop").

O princípio deste método é o processo de equilíbrio entre duas soluções através da fase de vapor num meio fechado. A solução menos concentrada poderá perder seu solvente volátil até que os potenciais químicos das duas soluções sejam iguais.

Para poder controlar a concentração final da solução de proteína, realizamos o experimento de difusão de vapor com um volume pequeno de solução de proteína contra um volume grande ("infinito") de solução precipitante – o reservatório – conforme pode ser observado na figura 2, para a técnica de "hanging drop". Uma gota contendo a macromolécula biológica a ser cristalizada, com o tampão, agentes precipitantes e aditivos, é equilibrada contra um reservatório contendo uma solução do agente precipitante à uma concentração mais alta que a da gota.

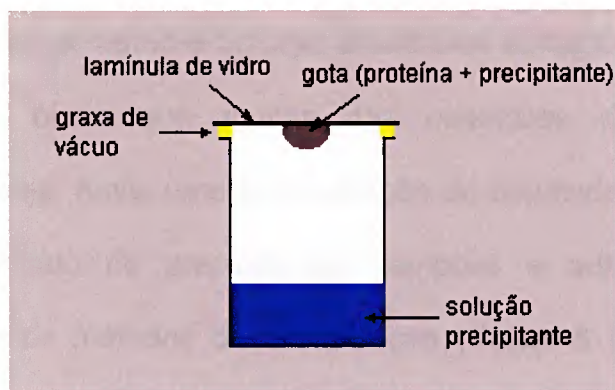


Figura 2. Método da difusão de vapor "hanging drop".

Se o volume da solução do reservatório é suficientemente grande comparado com o de uma gota, podemos assumir que quando o equilíbrio é atingido, a concentração do precipitante na solução de proteína será igual à concentração inicial na solução do reservatório.

d. "Batch"

Geralmente utilizada para produção de cristais em larga escala, uma vez que as condições de cristalização já sejam conhecidas.

Consiste na preparação de uma solução de proteína de tal forma que cristalizará espontaneamente no tempo. Podemos aplicar um gradiente de temperatura, se por exemplo, a proteína for mais solúvel a temperaturas mais altas. Aquecendo-se a solução, e permitindo-se o seu resfriamento lento, pode-se levar à formação de cristais.

1.3. O Método da matriz esparsa (Método Fatorial)

Com o aumento do número de macromoléculas biológicas cristalizadas com sucesso, tornou-se óbvio que muitas das condições de cristalização se assemelhavam, ou seja, havia uma concentração de resultados positivos usando-se um número limitado de precipitantes, tampões e aditivos. Isto levou à proposição de diversos métodos de cristalização (Carter & Carter, 1979), onde algumas condições foram testadas, usando-se pequenas quantidades (mg) da macromolécula biológica. A partir da observação dos resultados preliminares destes experimentos, foi possível determinar que tampão, aditivo e agente precipitante seriam os mais favoráveis, e a partir daí proceder-se a sucessivos melhoramentos até se conseguir cristais adequados. Ou ainda, em casos favoráveis, obter-se cristais adequados já na primeira tentativa com as condições padrões. Dentro deste raciocínio, foi proposto o método da matriz esparsa, onde diversas condições diferentes são usadas para se cristalizar uma macromolécula biológica (Jancarick & Kim, 1991).

Visto que o número de variáveis que afetam a cristalização é grande e combinatória (concentração, temperatura, pH, força iônica, aditivos específicos e precipitantes), o número total de possíveis condições de cristalização a serem testadas é tão grande que uma procura completa com todas as condições de cristalização em potencial seria inviável. No método da matriz esparsa foram escolhidas três categorias de parâmetros que afetam a cristalização como variáveis principais: pH e materiais tampões, aditivos e agentes precipitantes. Para a variável pH e tampões, foram escolhidos cinco diferentes pHs: 4,6; 5,6; 6,5; 7,5 e



8,5, e para cada valor de pH foi escolhido o tampão químico que tinha provado ser o mais adequado para a cristalização de proteínas no passado. A escolha de aditivos também foi baseada na experiência passada de muitos laboratórios. A tabela 1 lista os quatro tipos de agentes precipitantes: não-voláteis, voláteis, vários agentes de “salting-out” e misturas.

Tabela 1. Parâmetros da matriz de cristalização (Jancarik & Kim, 1991).

Agentes precipitantes*			
Não-Voláteis	Sais	Voláteis	Mistura
MPD	Tartarato de Na,K	2-Propanol	Sulfato de NH ₄ + PEG
PEG 400	Fosfato de NH ₄		2-Propanol + PEG
PEG 4000	Sulfato de NH ₄		
PEG 8000	Acetato de Na		
	Sulfato de Li		
	Formiato de Na		
	Fosfato de Na,K		
	Citrato de Na		
	Formiato de Mg		
Faixa de pH:	4,6; 5,6; 6,5; 7,5 e 8,5		

*Aditivos: Cloreto de Ca, Citrato de Na, Cloreto de Mg, Acetato de NH₄, Sulfato de NH₄, Acetato de Mg, Acetato de Zn e Acetato de Ca.

Por tentativa e erro a matriz multimensional foi simplificada, eliminando-se as condições que poderiam ser parcialmente representadas por resultados de outras condições. As 50 condições finais estão listadas no Apêndice II.



1.4. Principais parâmetros a serem controlados num experimento de cristalização

1.4.1. Pureza da macromolécula biológica.

Devido ao fato de que as macromoléculas são extraídas de complexas misturas biológicas, a purificação é um passo importante para a cristalogênese. Entretanto, não é uma condição primordial para a cristalização, visto que cristais de macromoléculas biológicas podem algumas vezes ser obtidos de misturas de macromoléculas, onde temos uma ou mais macromoléculas biológicas como contaminante. Mas, na maioria das vezes, são cristais bem pequenos, ou então, crescem como massas policristalinas, que não são adequadas para estudos cristalográficos por difração de raios X. Para o uso de técnicas de difração de raios X, é necessário o uso de cristais com a menor dimensão por volta de 0,2mm. Contudo, com o uso de fontes de raios X bem intensas, tais com síncrotrons, tem sido possível a resolução de estruturas de macromoléculas biológicas cujos cristais apresentavam dimensões menores que 0,1mm. Como regra geral, pode-se dizer que a baixa pureza da macromolécula biológica é a causa mais comum da falta de sucesso na sua cristalização.

1.4.2. Solubilidade e supersaturação

Para crescer cristais de qualquer composto, as moléculas ou íons precisam ser trazidos a um estado de supersaturação, estado que é termodinamicamente instável, no qual pode-se chegar a uma fase amorfa ou cristalina, quando este retornar ao equilíbrio. A supersaturação pode ser alcançada pela evaporação lenta

do solvente ou pela variação de alguns dos parâmetros listados na tabela 2. Portanto, o conhecimento das solubilidades macromoleculares é um pré-requisito para o controle das condições de cristalização. Assim, como as bases teóricas da solubilidade são ainda controversas, especialmente no que tange efeitos dos sais, os dados sobre a solubilidade quase sempre se originam de determinações experimentais. Nos últimos anos, métodos quantitativos permitiram tais determinações, usando pequenas quantidades de proteína (Mikol & Giegé, 1989; Ries-Krautt & Ducruix, 1989). O principal resultado destes métodos foi a demonstração experimental da complexidade do comportamento da solubilidade, enfatizando a importância da determinação de diagramas de fase para o planejamento racional de experimentos de crescimento de cristais de macromoléculas biológicas.

Tabela 2. Parâmetros que afetam a cristalização (e/ou solubilidade) de macromoléculas biológicas

Parâmetros físico-químicos intrínsecos:
• Supersaturação (Concentração das macromoléculas e precipitantes) • Temperatura, pH (Flutuações destes parâmetros) • Tempo (Cinética do equilíbrio e crescimento) • Força iônica e pureza das substâncias químicas utilizadas • Difusão e convecção (géis e microgravidade) • Volume e geometria das amostras (superfície das câmaras de cristalização) • Partículas sólidas, efeitos de parede e interface • Efeitos de densidade e viscosidade (diferenças entre cristal e líquido-mãe) • Pressão, campos elétricos e magnéticos • Vibrações e som (ondas acústicas) • Sequência de eventos (experimentalista versus robô)
Parâmetros bioquímicos e biofísicos:
• Sensibilidade das conformações aos parâmetros físicos (temperatura, pH, força iônica) • Ligação de substratos, cofatores, íons metálicos e outros íons • Aditivos específicos (agentes redutores, detergentes não iônicos, poliaminas) • Envelhecimento da amostra (degradação e desnaturação)



Tabela 2. Parâmetros que afetam a cristalização (e/ou solubilidade) de macromoléculas biológicas (continuação)**Parâmetros biológicos:**

- Pequena quantidade da maioria das macromoléculas biológicas
- Fontes biológicas e estado fisiológico dos organismos ou células

Pureza das macromoléculas:

- Contaminantes macromoleculares
- Heterogeneidade da sequência de aminoácidos
- Heterogeneidade conformacional

1.4.3. Nucleação, crescimento e cessação de crescimento

Devido ao fato de que proteínas e ácidos nucleicos necessitam de pH e força iônica definidos para a sua estabilidade e função, seus cristais precisam ser crescidos em soluções aquosas complexas. A cristalização começa com a fase de nucleação (formação dos primeiros agregados ordenados), e é seguida pela fase de crescimento. As condições para se atingir a nucleação são algumas vezes difíceis de se repetir, e desta forma os procedimentos de semeadura ("seeding") com um material cristalino pré-formado se faz necessário. Em muitos casos, este é o único método para se obter resultados reprodutíveis. A nucleação requer um estado de supersaturação maior do que aquele da fase de crescimento. A velocidade da cristalização aumenta com a supersaturação. Desta forma, a nucleação e o crescimento deveriam ser desacoplados, o que quase nunca é realizado conscientemente. O término de crescimento de uma macromolécula biológica pode ter diversas causas. Desconsiderando a remoção da macromolécula do meio de cristalização, defeitos de crescimento, envenenamento das faces ou envelhecimento da macromolécula são causas do término de crescimento de um cristal.

1.5. Solubilidade de macromoléculas biológicas

Uma macromolécula biológica segue as mesmas leis da termodinâmica, como qualquer outro tipo de molécula (inorgânica ou orgânica) no que se refere à supersaturação, nucleação, e crescimento de cristais (Boistelle & Astier, 1988). Todavia, macromoléculas biológicas apresentam peculiaridades, porque as interações necessárias para solubilizá-las num solvente são similares, e podem ser competitivas com as interações intermoleculares, responsáveis pela estrutura tridimensional da macromolécula (Ducruix & Giegé, 1992).

Em geral, a solubilidade de um composto químico em um solvente, a uma dada temperatura, é definida como a quantidade de composto dissolvido em uma solução em equilíbrio com um excesso de composto não dissolvido. A solubilidade é definida pelas características do solvente. Como o solvente consiste usualmente de água, a qual foram adicionadas diversas substâncias químicas para levar a solução até um dado pH (tampão), força iônica (sais), e eventualmente outros aditivos, estes compostos podem afetar a solubilidade da macromolécula biológica das seguintes formas:

- (a) diretamente, pela interação com diferentes grupos funcionais da macromolécula e, eventualmente, modificando a conformação da macromolécula. Trabalhar em diferentes condições de tampão e sais é como trabalhar com uma macromolécula diferente do ponto de vista físico-químico;
- (b) indiretamente, através da modificação das propriedades físico-químicas do solvente, ou seja, variando-se o pH, modifica-se a carga líquida da proteína e



portanto, a natureza do polieletrólito. Além disso, se pontes salinas estão envolvidas na estrutura tridimensional, estas podem se quebrar e a macromolécula modificar sua conformação.

A estabilidade das macromoléculas biológicas em solução baseia-se na competição das interações solvente-soluto com interações intramoleculares, as quais são necessárias para manter a estrutura da macromolécula. O balanço das interações que controlam a solubilidade e/ou a conformação de uma macromolécula podem ser modificadas (von Hippel & Schleich, 1969) por:

- (a) temperatura: um aumento na temperatura aumenta a desordem das moléculas de solvente e permite também que conformações da macromolécula de mais alta energia livre total se formem;
- (b) pH: variações no pH afetam ambos, soluto e solvente. Todavia, alterações nas concentrações de H^+ e OH^- são minoritárias, se comparadas com a protonação e a desprotonação de grupos da macromolécula;
- (c) sais: podem agir de diferentes formas. (c.1) São responsáveis pela força iônica e, desta forma, afetam as interações eletrostáticas da macromolécula por meio de blindagem de cargas; (c.2) podem formar interações eletrostáticas diretas com resíduos carregados na superfície das macromoléculas. (c.3) podem agir através de interações monopolo-dipolo com grupos dipolares da macromolécula biológica;
- (d) competidores de ligações de hidrogênio: moléculas que possuem um potencial para formarem ligações de hidrogênio competem a alta concentração ($>4M$)



com ligações de hidrogênio da água e ligações de hidrogênio intramoleculares na proteína;

(e) aditivos hidrofóbicos: podem agir diretamente com partes hidrofóbicas da macromolécula e podem também alterar a estrutura do solvente. Detergentes não-iônicos agem desta forma (McPherson *et al.*, 1986);

(f) solventes orgânicos: a sua adição implica na modificação da constante dielétrica e, conseqüentemente, em variações em diversas interações.

1.6. “Salting-in” e “salting-out”

A dependência da solubilidade da proteína com a concentração salina foi estudada em termos dos fenômenos de “salting-in” e “salting-out” (Green, 1932). O aumento da solubilidade de uma macromolécula biológica a baixa concentração salina ($<0,5M$) é chamada “salting-in”. Este fenômeno é explicado pelas interações eletrostáticas não específicas entre a macromolécula carregada e as espécies iônicas do sal. Segundo a teoria de Debye-Hückel para soluções iônicas, um aumento na força iônica reduz a atividade dos íons em solução, aumentando a solubilidade do composto iônico. Uma forma alternativa de se tratar este fenômeno é considerar o “salting-in” como o resultado da competição entre grupos carregados na superfície da macromolécula e os íons em solução. Na ausência de íons de solvente, a macromolécula precipita devido à atração de Coulomb entre cargas opostas em diferentes partes da macromolécula. Se íons são adicionados à solução, eles blindam os grupos carregados da macromolécula e aumentam a sua solubilidade (Drenth, 1994; Ducruix & Giegé, 1992) (Figura 3).



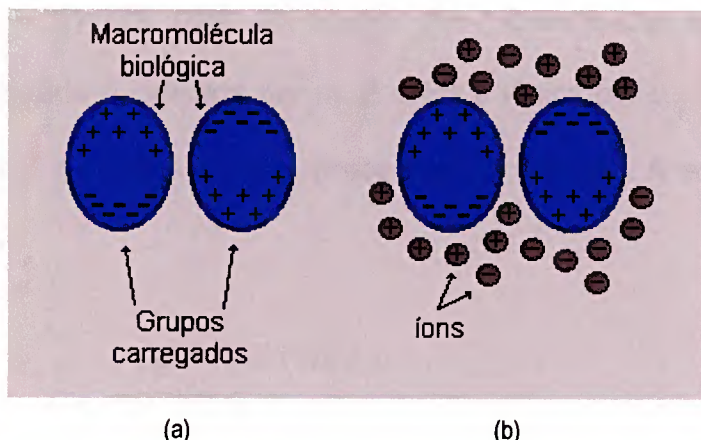


Figura 3. Fenômeno de “salting-in”. a) Macromolécula biológica sem a presença de íons dissolvidos na solução. A atração eletrostática entre os grupos carregados em duas ou mais macromoléculas causa a aglomeração e precipitação. b) A adição de íons na solução blindam a interação eletrostática entre as macromoléculas, aumentando a solubilidade.

Outra forma de precipitar uma macromolécula é feito pela adição de sal (“salting-out”). Este serve para imobilizar as moléculas de água, aumentando a concentração efetiva da macromolécula, ou seja, diminuindo a sua solubilidade. No fenômeno de “salting-out” a solubilidade da macromolécula biológica é governada principalmente por efeitos hidrofóbicos. De acordo com a equação linear empírica de Cohn-Green a solubilidade da macromolécula numa situação de “salting-out” decresce da seguinte forma:

$$\log S = \beta - K_S m, \text{ ou } \log S = \beta - K_S' \mu \quad (1.1)$$

onde S é a solubilidade da macromolécula biológica, β o intercepto para $m=0$. β é uma constante para altas concentrações salinas e é função da carga líquida da macromolécula e altamente dependente do pH da solução (no ponto isoelétrico (pI), β tende a zero). A magnitude de β , bem como a distribuição de cargas, variam

com a temperatura. K_S e K_S' são constantes de “salting-out”. São independentes da temperatura e do pH, mas dependem da natureza dos sais presentes na solução. A molaridade é indicada por m . E a força iônica por μ . Esta força leva em conta a valência Z_i de todos os íons presentes na solução. A expressão da força iônica é dada por:

$$\mu = \frac{1}{2} \sum (m_i Z_i^2). \quad (1.2)$$

Um composto adicionado a um sistema macromolécula biológica + água (soluto+solvente) pode se ligar à macromolécula (ligação preferencial) ou ser excluído (exclusão preferencial). A interação líquida do “salting-out” é a exclusão preferencial, mesmo que moléculas possam se ligar à macromolécula biológica.

A variação da solubilidade das macromoléculas biológicas (S) na faixa completa de concentração salina reflete o efeito de ambas interações, eletrostática e hidrofóbica. A primeira sendo predominante a baixas concentrações salinas e a segunda a altas concentrações salinas, como mostrado na figura 4. Isto é expresso pela seguinte equação (Green, 1932):

$$\log S = \log S_0 + k_i C^{1/2} - k_o C, \quad (1.3)$$

onde k_i e k_o são as constantes de “salting-in” e “salting-out”, respectivamente; S_0 é a solubilidade da macromolécula biológica em água pura e C a concentração molar salina.



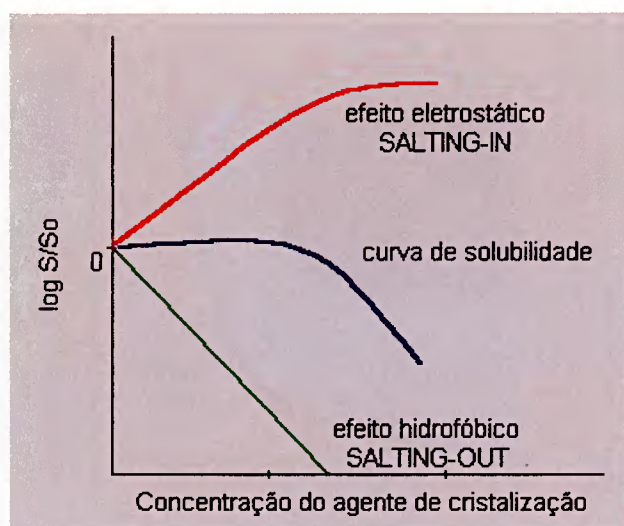


Figura 4. Contribuição dos efeitos eletrostático e hidrofóbico sobre a solubilidade S , normalizada em relação à solubilidade em água pura S_0 .

1.7. Diagramas de fase de macromoléculas biológicas

Como a solubilidade de uma macromolécula biológica depende de diversos parâmetros (Tabela 2), um diagrama bidimensional é uma representação desta solubilidade (mg/ml ou mM de macromolécula em solução) como função de dois parâmetros, com todos os outros sendo mantidos constantes (Ducruix & Giegé, 1992). Um diagrama deste tipo está representado na Figura 5, e compreende as seguintes zonas:

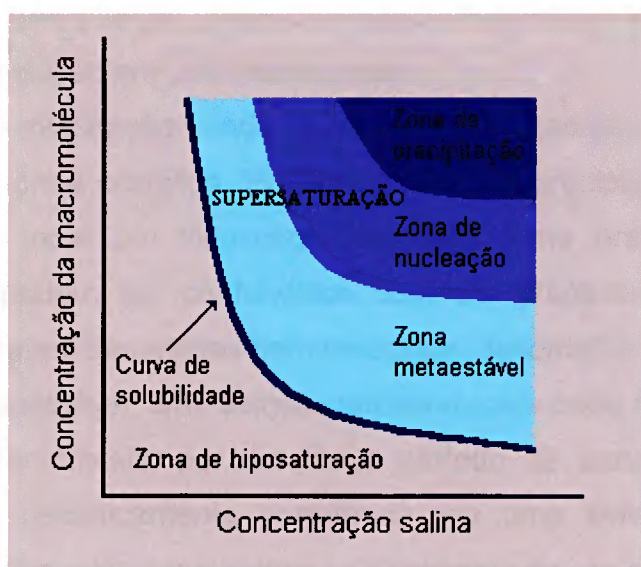


Figura 5. Descrição esquemática de um diagrama bidimensional da solubilidade de uma macromolécula biológica.

1. Curva de solubilidade (S): separa as zonas supersaturadas e hiposaturadas. Em um experimento onde as concentrações do agente de cristalização e da macromolécula biológica correspondem às condições da curva S, a solução saturada da macromolécula está em equilíbrio com a macromolécula cristalizada. Isto corresponde à situação ao final do processo de crescimento do cristal a partir de uma solução supersaturada, ou da dissolução de cristais em uma solução hiposaturada.
2. Abaixo da curva de solubilidade (S): a solução está hiposaturada e a macromolécula biológica nunca irá cristalizar.
3. Acima da curva de solubilidade (S): a concentração da macromolécula biológica é maior que a concentração no equilíbrio para uma dada concentração salina. Isto corresponde à zona de supersaturação. Dependendo da cinética para se atingir o equilíbrio e do nível de supersaturação, esta região pode ser dividida em três zonas:

(3.1) Zona de precipitação: onde o excesso de macromolécula imediatamente se separa da solução em um estado amorfo.

(3.2) Zona de nucleação: onde o excesso da macromolécula biológica se separa sob a forma cristalina. Próxima à zona de precipitação, a cristalização pode ocorrer como um 'chuveiro' ("shower") (uma grande quantidade de cristais) que podem ser confundidos com um precipitado amorfo, pois os cristais apresentam dimensões bem reduzidas, assemelhando-se a um pó.

(3.3) Zona metaestável: uma solução supersaturada pode não nuclear, ou seja, pode não gerar cristais, por um longo período de tempo, a menos que a solução seja mecanicamente perturbada, ou uma semente (um pequeno cristal) seja introduzida (semeadura). Classicamente, este processo se divide em macrossemeadura ("macroseeding"), onde um cristal de dimensões "macroscópicas" é introduzido à uma solução na zona metaestável, e microssemeadura ("microseeding"), onde normalmente vários cristais bem pequenos são inseridos numa solução na zona metaestável. Contudo, esta diferenciação entre macro e micro é um tanto arbitrária, sendo que o principal é que se tenha uma ou mais sementes inseridas numa solução com a macromolécula na zona metaestável. Esta zona corresponde, idealmente, ao crescimento de cristais sem a nucleação de novos cristais.

Uma solução saturada contém uma quantidade de soluto tal que não pode haver crescimento ou dissolução, se cristais são adicionados à solução, ou seja, qualquer ponto sobre a curva (S) no diagrama de fases esquemático (Figura 5). Isto corresponde ao equilíbrio termodinâmico entre as duas fases (líquido e sólido). O potencial químico (μ) de cada espécie química, i , é a mesma em ambas as fases (cristal e solução):

$$\mu_{ic} = \mu_{is} = \mu_{io} + RT \ln \gamma c_i \quad (1.4)$$

onde μ_{io} é o potencial químico padrão, γ é o coeficiente de atividade, e c_i a concentração da espécie química i . Isto é verdadeiro para qualquer espécie química. A supersaturação é alcançada quando o potencial químico do soluto em solução é maior que o do cristal. Isto pode ser atingido através da variação de alguns dos parâmetros que afetam a solubilidade (Tabela 2). A supersaturação é a força motriz para a cristalização; e é definida como a diferença do potencial químico entre a solução e o cristal. Entretanto, as aproximações mais comuns são as relações c/c_s e $(c-c_s)/c_s$ onde c é a concentração da proteína e c_s a concentração do soluto em uma solução saturada, ou seja, um ponto sobre a curva de solubilidade da figura 5. Estas quantidades têm a vantagem de serem adimensionais. Deve-se notar que o valor da supersaturação (definido como a relação entre a concentração de macromolécula biológica c e a solubilidade c_s) realmente encontrada nos experimentos de cristalização de macromoléculas biológicas ficam na faixa de 2 a 10. Esta faixa é bem mais alta que aquela encontrada para o crescimento de cristais de pequenas moléculas.

1.8. Melhoria das Condições de Cristalização

Considerando a existência de pequenos cristais nos experimentos iniciais, com ou sem precipitado amorfo acompanhando-os, dá-se início a tentativas de melhoria dos cristais e eliminação do precipitado amorfo.

1º passo: Diminuir a concentração da macromolécula e tentar as mesmas condições em que foram obtidos os primeiros cristais. Tentar várias concentrações:



se os primeiros cristais cresceram à concentração de 10mg/ml, tentar 9, 8, 7, 6 e 5 mg/ml. Esperar por uma semana e, se não houver melhora nos cristais,

2º passo: Manter a concentração da macromolécula constante e abaixar a concentração do agente precipitante. Esperar por uma semana e se não houver melhora nos cristais,

3º passo: Adicionar 0,2M NaCl ao reservatório. Se a gota de cristalização ficar completamente clara após uma semana (sem precipitado ou cristais), aumentar a concentração do precipitante em incrementos de 5% a cada 3 dias. Se a gota continuar sem cristais ou precipitado, tentar o mesmo procedimento usando-se 20mM NaCl.

A eliminação de um possível precipitado é importante, visto que o mesmo age como sítio de nucleação e complica a manipulação futura do cristal.

Se não houver melhora nos cristais, pode-se tentar microsemeadura ("microseeding"). Caso tal tentativa não funcione, pode-se tentar variar a temperatura.

2. Difração de Raios X

Após a cristalização da macromolécula, o próximo passo é expor o cristal ao feixe de raios X, pois para distinguir átomos individuais, o comprimento de onda da radiação utilizada deve ser próximo às distâncias entre os átomos. Quando um feixe de raios X atravessa um cristal, ocorre uma interação entre os elétrons dos átomos e os raios. Os elétrons acelerados pela força do campo elétrico da



radiação, passam a espalhar radiação para todas as direções, num processo denominado espalhamento. Existem dois tipos de espalhamento: Thomson e Compton.

2.1. Espalhamento Thomson

O campo elétrico oscilante associado ao feixe de raios X que incide sobre um elétron obriga este elétron a oscilar em torno da sua posição de equilíbrio. Sabemos que toda partícula carregada acelerada emite radiação. Assim, o elétron submetido a um campo elétrico oscilante emite uma onda eletromagnética que possui o mesmo comprimento de onda da radiação incidente (espalhamento elástico). A intensidade do feixe de raios X espalhados por um elétron de carga $-e$ e massa m a uma distância r do elétron, é dada por,

$$I = I_0 \frac{e^4}{r^2 m^2 c^4} \left(\frac{1 + \cos^2 2\theta}{2} \right) \quad (2.1)$$

onde I_0 é a intensidade do feixe incidente e 2θ o ângulo de espalhamento da radiação, sendo a onda eletromagnética plana e polarizada (Woolfson, 1970; Cullity, 1956).

2.2. Espalhamento Compton

Há uma forma completamente diferente pela qual um elétron pode espalhar raios X, conhecida como efeito Compton. Esse efeito ocorre quando raios X incidem sobre elétrons livres ou fracamente ligados e pode ser melhor entendido a

partir da teoria corpuscular da luz. Assim, a partir da conservação do momento e da energia da colisão do fóton com o elétron, obtemos a expressão abaixo,

$$\Delta\lambda = 0,0243(1-\cos 2\theta) \text{ (Å)} \quad (2.2)$$

onde 2θ é o ângulo de espalhamento e $\Delta\lambda$ é a diferença entre o comprimento de onda da radiação espalhada e incidente. Assim, temos que a radiação espalhada possui um comprimento de onda maior que o da radiação incidente (espalhamento inelástico), devido a transferência de energia do fóton para o elétron.

Experimentalmente, a radiação espalhada pelos materiais consiste de duas partes. A primeira parte é aquela associada ao espalhamento Thomson e possui o mesmo comprimento de onda da radiação incidente; a segunda parte tem um comprimento de onda maior que a radiação incidente, com o aumento do comprimento de onda sendo dependente do ângulo de espalhamento (Woolfson, 1970; Cullity, 1956).

2.3. Espalhamento por elétrons

Para analisar o espalhamento de raios X por elétrons vamos considerar a geometria de um experimento típico de espalhamento, como mostrado na figura 6. Um feixe de raios X colimado incidindo sobre um elétron localizado na origem do sistema de coordenadas.



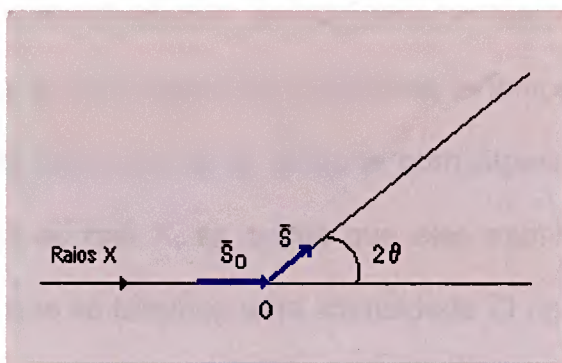


Figura 6. Espalhamento de raios X por um elétron.

Um vetor unitário s_0 descreve a direção da radiação incidente. A direção de espalhamento é indicada por outro vetor unitário s e o ângulo de espalhamento é 2θ . A figura 6 mostra o vetor de espalhamento S , dado pela expressão

$$\vec{S} = \frac{\vec{s} - \vec{s}_0}{\lambda} \quad (2.3)$$

a partir da figura 6, vemos que o módulo S é função do ângulo de espalhamento, como segue,

$$|\vec{S}| = \frac{2|\sin \theta|}{\lambda} \quad (2.4)$$

O valor do módulo de S pode variar de 0 a $2/\lambda$. Dessa forma, o vetor S está descrito num espaço onde cada eixo de seu sistema de coordenadas tem dimensão de recíproco da distância. Este espaço de coordenadas é chamado espaço recíproco (Cantor & Schimmel, 1980).

2.4. Fator de espalhamento atômico

Considerando que um elétron isolado espalha raios X com intensidade I , seria de se esperar que num átomo de Z elétrons teríamos uma intensidade ZI . Porém, devido às distâncias entre os elétrons num átomo serem da ordem do comprimento de onda do raio X, as ondas que eles espalham interferem umas com outras de forma que só teremos uma intensidade ZI na direção de incidência do raio X. Para o espalhamento em outras direções, temos interferência parcialmente destrutiva, de maneira que a amplitude total cai com o aumento do ângulo de espalhamento. O fator de espalhamento atômico é definido como a relação entre a amplitude espalhada por um átomo (E_a) e a amplitude espalhada por um elétron (E_e) isolado, sob condições idênticas

$$f = \frac{E_a}{E_e} \quad (2.5)$$

O valor máximo de f é Z (número atômico do átomo) e ocorre quando os elétrons espalham em fase, na direção de incidência ($2\theta = 0$).

O fator de espalhamento atômico também depende do comprimento de onda da radiação incidente. Para um valor fixo de θ , f será menor para comprimentos de onda mais curtos, visto que a diferença de caminho será maior com relação ao comprimento de onda, levando a uma maior interferência.

Considerando um átomo esférico com o seu centro coincidente com a origem do sistema de coordenadas, a onda total espalhada por um pequeno volume dv numa posição $\mathbf{r}$ relativa à onda espalhada na origem terá uma amplitude proporcional a $\rho(\mathbf{r})dv$ e uma fase $2\pi\mathbf{r}\cdot\mathbf{S}$, ou seja, a amplitude da onda

espalhada será igual a $\rho(\mathbf{r})\exp(2\pi i \mathbf{r} \cdot \mathbf{S})d\mathbf{v}$ (Cantor & Schimmel, 1980; Azároff, 1968).

Consequentemente, a onda total espalhada por um átomo é calculada pela soma das ondas espalhadas pelos elementos de volume $d\mathbf{v}$

$$f(\mathbf{S}) = \int_{\text{vol. átomo}} \rho(\mathbf{r})\exp(2\pi i \mathbf{r} \cdot \mathbf{S})d\mathbf{v} \quad (2.6)$$

A expressão acima representa o fator de espalhamento atômico. Curvas do espalhamento atômico para diversos átomos estão tabeladas no Volume III da “International Tables for X-ray Crystallography”.

2.5. Espalhamento de raios X por uma molécula

Vamos analisar agora o espalhamento de raios X de um conjunto de átomos colocados em posições definidas pelos vetores posição $\mathbf{r}_i$ (Blundell & Johnson, 1976; Lipson & Taylor, 1958).

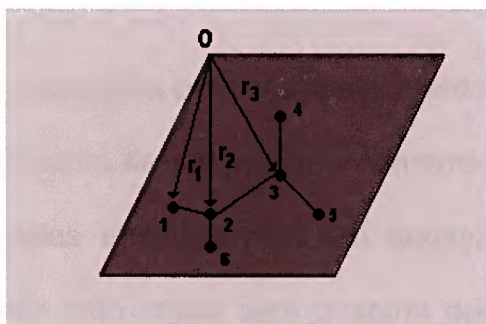


Figura 7. Posições atômicas em uma célula unitária.

Consideremos o átomo 1 da figura 7 que está a uma distância r_1 da origem (O). Este deslocamento de centro do átomo significa que a distância r na equação (2.6) é substituída por $r+r_1$. Assim, temos que o espalhamento do átomo 1 será dado pela seguinte expressão

$$f_1 = \int_{\text{vol. átomo}} \rho(\mathbf{r}) \exp(2\pi i (\mathbf{r}_1 + \mathbf{r}) \cdot \mathbf{S}) d\mathbf{v} = f_1' \exp(2\pi i \mathbf{r}_1 \cdot \mathbf{S}) \quad (2.7)$$

onde,

$$f_1' = \int_{\text{vol. átomo}} \rho(\mathbf{r}) \exp(2\pi i \mathbf{r} \cdot \mathbf{S}) d\mathbf{v} \quad (2.8)$$

Expressões similares podem ser obtidas para os outros átomos.

A onda total espalhada por todos os átomos é dada pela soma vetorial das contribuições de cada átomo.

$$\mathbf{G}(\mathbf{S}) = \sum_{j=1}^N f_j \exp(2\pi i \mathbf{r}_j \cdot \mathbf{S}) \quad (2.9)$$

2.6. Espalhamento de raios X por um cristal

Afim de obter a expressão para o espalhamento por um cristal, primeiro consideramos o espalhamento de um cristal unidimensional, que é composto de um arranjo linear de celas unitárias com um espaçamento a entre elas. A amplitude total espalhada pelo cristal será a soma das ondas espalhadas por cada cela unitária. A amplitude da onda espalhada pela primeira cela unitária relativa a origem é simplesmente $\mathbf{G}(\mathbf{S})$. A amplitude espalhada pela segunda cela

unitária relativa à mesma origem é $\mathbf{G}(\mathbf{S})\exp(2\pi i \mathbf{a} \cdot \mathbf{S})$, visto que todas as distâncias estão deslocadas pelo vetor $\mathbf{a}$. A amplitude da onda espalhada pela n -ésima cela unitária é $\mathbf{G}(\mathbf{S})\exp(2\pi i(n-1)\mathbf{a} \cdot \mathbf{S})$. Consequentemente, a amplitude total espalhada é

$$\mathbf{F}(\mathbf{S}) = \sum_{n=1}^T \mathbf{G}(\mathbf{S}) \exp 2\pi i(n-1)\mathbf{a} \cdot \mathbf{S} \quad (2.10)$$

onde T é o número total de celas unitárias.

A onda de cada cela unitária está fora de fase com sua vizinha por uma quantidade $2\pi \mathbf{a} \cdot \mathbf{S}$. Consequentemente, conforme o número de celas unitárias aumenta, a amplitude total espalhada $\mathbf{F}(\mathbf{S})$ fica da mesma ordem de $\mathbf{G}(\mathbf{S})$ que, para raios X é muito pequena para ser observada.

O espalhamento só será observado quando a diferença de fase entre as ondas espalhadas por celas unitárias sucessivas for um múltiplo inteiro de 2π , ou seja, $\mathbf{a} \cdot \mathbf{S} = h$, onde h é um número inteiro.

Sob estas circunstâncias, as ondas se somam para formar uma onda espalhada mais intensa, que é proporcional em magnitude a $T \cdot |\mathbf{G}(\mathbf{S})|$. Em resumo, para uma rede unidimensional, só observamos espalhamento quando $\mathbf{a} \cdot \mathbf{S} = h$. Quando o problema é estendido para três dimensões com uma cela unitária definida pelos vetores $\mathbf{a}$, $\mathbf{b}$ e $\mathbf{c}$, a condição para ocorrer a difração é que as condições $\mathbf{a} \cdot \mathbf{S} = h$, $\mathbf{b} \cdot \mathbf{S} = k$ e $\mathbf{c} \cdot \mathbf{S} = l$ sejam simultaneamente satisfeitas. Estas condições correspondem às conhecidas equações de Laue.

Assim podemos reescrever a amplitude total da seguinte forma

$$F(S) = \sum_{j=1}^N f_j \exp 2\pi i(\mathbf{r}_j \cdot \mathbf{S}) \quad (2.11)$$

onde $\mathbf{r}_j = a\mathbf{x}_j + b\mathbf{y}_j + c\mathbf{z}_j$ são as coordenadas fracionárias do j -ésimo átomo. Sendo que a constante de proporcionalidade T foi omitida. As coordenadas fracionárias (x,y,z) são definidas como

$$x = \frac{X}{a}, y = \frac{Y}{b} \text{ e } z = \frac{Z}{c},$$

onde X, Y, Z são as coordenadas absolutas do átomo na cela unitária de eixos a, b e c .

Considerando as equações de Laue, temos que

$$\begin{aligned} \mathbf{r}_j \cdot \mathbf{S} &= x_j \mathbf{a} \cdot \mathbf{S} + y_j \mathbf{b} \cdot \mathbf{S} + z_j \mathbf{c} \cdot \mathbf{S} \\ &= h x_j + k y_j + l z_j \end{aligned}$$

portanto,

$$F(hkl) = \sum_{j=1}^N f_j \exp 2\pi i(h x_j + k y_j + l z_j) \quad (2.12)$$

onde $a \cdot S, b \cdot S$ e $c \cdot S$ foram substituídos por h, k, l no lado esquerdo da equação.

A equação (2.12) é conhecida como equação do fator de estrutura. Se as posições de todos os átomos na cela unitária são conhecidas, então o correspondente padrão de difração pode ser calculado (Blundell & Johnson, 1976; Lipson & Taylor, 1958).

2.7. Densidade eletrônica

O padrão de difração é a transformada de Fourier da densidade eletrônica da estrutura e inversamente, a densidade eletrônica da estrutura é a transformada de Fourier do padrão de difração. Para mostrar isto, podemos reescrever a equação do fator de estrutura (equação 2.12) em termos de uma integral sobre o volume da cela unitária (V) (Sherwood, 1976).

$$F(S) = \sum_{j=1}^N f_j \exp 2\pi i(\mathbf{r}_j \cdot \mathbf{S})$$

$$= \int_V \rho(\mathbf{r}) \exp 2\pi i(\mathbf{r} \cdot \mathbf{S}) d\mathbf{v},$$

onde S é usado para representar a posição no espaço recíproco e $\rho(\mathbf{r})$ é a densidade eletrônica. Multiplicando ambos os lados por $(\exp -2\pi i(\mathbf{r}' \cdot \mathbf{S}))$ e integrando sobre o volume recíproco ($V^*=1/V$), temos

$$\rho(\mathbf{r}) = \int_{V^*} F(\mathbf{S}) \exp -2\pi i(\mathbf{r} \cdot \mathbf{S}) d\mathbf{v}^*$$

onde $d\mathbf{v}^*$ é o elemento de volume no espaço recíproco.

A integração pode ser substituída por uma somatória, visto que F(S) não é contínuo e é diferente de zero somente nos pontos do retículo recíproco. Consequentemente,

$$\rho(hkl) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F(hkl) \exp -2\pi i(hx + ky + lz) \quad (2.13)$$

Desta forma, se os fatores de estrutura F(hkl) são conhecidos para todas as reflexões hkl, então a densidade eletrônica $\rho(xyz)$ pode ser calculada para

cada ponto xyz na cela unitária. A densidade eletrônica representa a estrutura do cristal.

2.8. O problema da fase

Para calcular a densidade eletrônica é necessário o conhecimento do módulo $F(hkl)$ e da fase $\alpha(hkl)$ do fator de estrutura. Isto é enfatizado quando reescrevemos a equação 2.13 como segue:

$$\rho(xyz) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F(hkl) \exp i\alpha(hkl) \exp - 2\pi i (hx + ky + lz)$$

Durante um experimento de difração de raios X, só se registram as intensidades, sendo que toda a informação sobre a fase é perdida. Portanto, é impossível determinar a estrutura diretamente pelas medidas do padrão de difração, visto que parte da informação é perdida. O problema da determinação da fase é o problema básico em qualquer determinação de estrutura (Blundell & Johnson, 1976).

Para resolvermos o problema da fase lançamos mão de alguns métodos indiretos. Atualmente existem quatro métodos principais para se determinar as fases dos fatores de estrutura: (i) a substituição molecular, (ii) a substituição isomorfa múltipla, (iii) a difração anômala a múltiplos comprimentos de onda, (iv) métodos diretos. Daremos ênfase apenas ao método de substituição molecular, pois foi a única das três técnicas utilizada no trabalho apresentado nesta tese, e é o mais usado.

3. Substituição Molecular

Uma das formas de se obter fases para os fatores de estrutura é através de modelos de busca. A idéia é a de construir um modelo e então simular a sua difração e comparar os seus fatores de estrutura com os fatores coletados experimentalmente. Isso costuma ser razoavelmente simples para pequenas moléculas, uma vez que se conheça o composto (e os átomos que o constituem) e o arranjo cristalino que estes poderiam assumir. No entanto, para macromoléculas biológicas, esta abordagem torna-se mais complexa, devido ao grande número de átomos (Lee, 2001).

O método de substituição molecular vem para aplicar a técnica de modelos às proteínas. Este método exige o conhecimento prévio da estrutura tridimensional de uma proteína (um modelo) com pelo menos 30% a 50% de sua sequência de aminoácidos idêntica à que está em análise. O maior problema com procuras usando-se modelos com baixa identidade é que uma vez que a solução seja obtida, as fases serão uma estimativa grosseira das fases verdadeiras, o que pode causar problemas durante o refinamento destas coordenadas. Tais modelos estão disponíveis em bancos de dados cristalográficos (Protein Data Bank – PDB), onde são depositadas as estruturas já determinadas. O modelo prévio permite atribuir o valor das fases iniciais para os dados de difração coletados, ao assumir que sequências parecidas tenham enovelamento geral também parecidos. Desta maneira, a posição dos átomos do modelo devem estar

relativamente próximos à posição dos átomos da proteína que se quer determinar a estrutura (Lee, 2001).

Porém, um dos problemas é a orientação do conjunto inteiro dos átomos de uma estrutura. Um mesmo modelo, posicionado de diferentes formas no espaço, gera fatores de estruturas diferentes, caso mantivéssemos todos os outros parâmetros. Para orientarmos o modelo corretamente precisamos, a princípio, de informações sobre o posicionamento espacial dos átomos da proteína em estudo, e isso depende da fase (Lee, 2001).

A solução para este problema é dado através do uso da função de Patterson $P(u)$ ou $P(uvw)$:

$$P(u) = \int \rho(r) \cdot \rho(r + u) dv \quad (3.1)$$

onde a integração é feita sobre o volume da cela unitária. Por esta equação, a função $P(u)$ atingirá valores diferentes de zero quando o vetor u (ou uvw) representar uma distância entre átomos. Esta mesma função pode ser escrita como uma soma de Fourier cujos coeficientes são os módulos do fator de estrutura ao quadrado e fases iguais a zero:

$$P(u, v, w) = \frac{1}{V_c} \sum_h \sum_k \sum_l |F(hkl)|^2 \cos 2\pi (hu + kv + lw) \quad (3.2)$$

de forma que será utilizado apenas a informação da intensidade da reflexão (dado pelo termo $|F(hkl)|^2$), ignorando a informação que seja dada pela fase. Em suma, a

função de Patterson fornece as distâncias vetoriais entre os átomos (mas não suas coordenadas) somente com os dados parciais dos quais dispomos.

Os vetores obtidos são divididos em intramoleculares e intermoleculares. Se os pares de átomos pertencem à mesma molécula, os vetores de Patterson correspondentes (intramoleculares) são relativamente curtos, e são semelhantes em estruturas similares. Por outro lado, os vetores intermoleculares resultam de pares de átomos colocados mais distantes, sendo relacionados por simetrias cristalográficas e não-cristalográficas.

Um aspecto importante dos mapas de Patterson, onde todos os vetores são desenhados com uma origem comum, é a sua relação intrínseca com a orientação do arranjo atômico em relação à cela unitária. Isso significa que um mesmo arranjo de átomos (uma estrutura) pode gerar mapas distintos, dependendo da orientação. Contudo, a relação entre diferentes mapas pode ser encontrada via comparação entre eles, e conseqüentemente, a relação entre as possíveis orientações desse arranjo.

Da mesma maneira, em estruturas semelhantes, podemos encontrar mapas dos dados experimentais com os mapas gerados do modelo com a finalidade de encontrar a orientação entre eles e, subseqüentemente, utilizar as informações de fase do modelo para resolvermos a estrutura em questão. No caso dos cristais de proteínas, precisamos realizar duas buscas distintas. A primeira busca é a chamada rotação, onde determinamos a orientação relativa. A segunda busca, já com a orientação estabelecida, é feita dentro da cela unitária, de forma a localizar a molécula dentro do seu espaço.

A função de rotação considera apenas os vetores intramoleculares e é expressa pela seguinte função:

$$R(\alpha, \beta, \gamma) = \int_U P(u) \cdot P_r(u_r) du \quad (3.3)$$

onde $P(u)$ é a função de Patterson da molécula desconhecida e $P_r(u_r)$ é a função Patterson da molécula modelo, submetida a uma rotação. A integração é feita sobre o volume U que contém os vetores intramoleculares (esfera mínima que contém a molécula modelo).

As funções de Patterson $P(u)$ e $P_r(u_r)$ podem ser expressas como somas de Fourier:

$$P(u) = \frac{1}{V} \sum |F(h)|^2 e^{-2\pi i h u} \quad (3.4)$$

$$P_r(u_r) = \frac{1}{V} \sum |F(h)|^2 e^{-2\pi i h [C]u} \quad (3.5)$$

onde $[C]$ é a matriz de rotação que relaciona os vetores u e u_r .

A função de rotação $R(\alpha, \beta, \gamma)$ terá o seu valor máximo quando forem encontrados os três ângulos que proporcionem a sobreposição entre os picos correspondentes nos mapas de Patterson do modelo e da molécula desconhecida.

Uma vez encontrada a orientação correta, o próximo passo é a busca de um vetor de translação que posicione corretamente a molécula modelo dentro da nova cela unitária. A função de translação vai fornecer a correlação entre os

vetores intermoleculares da molécula modelo e a função de Patterson observada, e é dada por:

$$T(t) = \int_V P_{1,2}(u, t) \cdot P(u) du \quad (3.6)$$

onde $P_{1,2}(u, t)$ é a função de Patterson, já orientada pelos ângulos de rotação, que leva em conta os vetores intermoleculares entre duas moléculas modelo relacionadas por simetria e $P(u)$ é a função de Patterson observada. A função $T(t)$ apresentará um máximo quando o vetor t corresponder a um vetor intermolecular, presente no mapa de Patterson observado.

Durante cada etapa das buscas, os processos são ainda monitorados através dos fatores de correlação CC (função de rotação) e do fator R cristalográfico R_{factor} (função de translação) para as diferentes posições da molécula modelo ao longo da nova cela unitária. Estes fatores são dados por:

$$CC = \frac{\sum_{hkl} \left(|F_{obs}|^2 - \overline{|F_{obs}|^2} \right) \cdot \left(|F_{calc}|^2 - \overline{|F_{calc}|^2} \right)}{\left[\sum_{hkl} \left(|F_{obs}|^2 - \overline{|F_{obs}|^2} \right)^2 \cdot \sum_{hkl} \left(|F_{calc}|^2 - \overline{|F_{calc}|^2} \right)^2 \right]^{1/2}} \quad (3.7)$$

$$R_{factor} = \frac{\sum_{hkl} \left| |F_{obs}| - k |F_{calc}| \right|}{\sum_{hkl} |F_{obs}|} \quad (3.8)$$

onde $|F_{obs}|$ e $|F_{calc}|$ são, respectivamente, os módulos dos fatores de estrutura experimental e do modelo (na posição relacionada com t). A substituição

molecular terá encontrado a solução mais aproximada para o problema das fases quando um conjunto de ângulos de rotação e vetor de translação fornecer um fator de correlação alto (~60%) e um fator R baixo (~30% a 35%) que contrastem significativamente com os mesmos parâmetros dos demais conjuntos.

Os procedimentos utilizados pelo método de substituição molecular se encontram implementados, com pequenas variações, em diversos programas tais como o AMoRe (Navaza, 1994) e CNS (Brunger *et al.*, 1998).

4. Refinamento das coordenadas

As soluções encontradas pelos diversos métodos assumem as fases em decorrência de cálculos aproximados. No caso da substituição molecular, essa aproximação dada por uma molécula semelhante (mas não idêntica) é suficiente para uma determinação grosseira da estrutura. Para chegarmos aos níveis de detalhamento desejado, faz-se necessário corrigir estas fases, de forma que elas correspondam da maneira mais precisa possível às fases reais dos fatores de estrutura. Este é o objetivo principal do refinamento, etapa que se segue às determinações das fases (Brunger, 1992).

Os principais passos para o refinamento das coordenadas de uma macromolécula usando o programa X-PLOR está mostrado na figura 8.



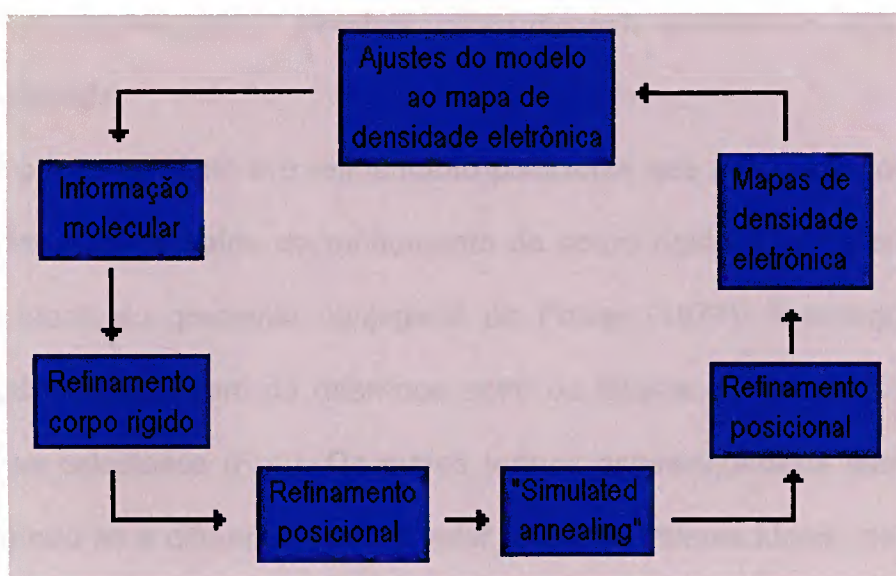


Figura 8. Diagrama em blocos do processo de refinamento usando o programa X-PLOR (Brunger, 1992).

O primeiro passo, que não é um refinamento, é a geração das informações moleculares da estrutura que terá suas coordenadas refinadas. Estas informações são divididas em dois arquivos, um com a topologia da macromolécula e outro com os parâmetros químicos. As informações sobre a topologia incluem as massas dos átomos, as ligações (conectividade), os ângulos diédricos, os tipos de átomos, a carga elétrica, as ligações de hidrogênio etc. O arquivo de parâmetros químicos armazena as constantes a serem usadas no cálculo da função energia (E_{total}), que será minimizada durante o refinamento. Estes dois arquivos são gerados a partir das coordenadas em formato PDB do modelo a ser refinado. Em seguida é realizado o refinamento de corpo rígido, que é o processo de refinamento das posições de grupos rígidos de átomos, contra as intensidades observadas. O modelo a ser refinado pode ser dividido em um ou mais grupos

rígidos. Para cada grupo é possível refinar até seis parâmetros, três rotações e três translações.

O passo seguinte é o refinamento posicional que usa como modelo inicial as coordenadas de saída do refinamento de corpo rígido. O programa X-PLOR usa o método do gradiente conjugado de Powell (1977). A energia do termo cristalográfico E_{cristal} vem da diferença entre os fatores de estrutura observados (F_{obs}) e os calculados (F_{calc}). Os outros termos estereoquímicos são calculados considerando-se a diferença entre o valor real e os valores ideais, de forma que, quando átomos se desviam do valor ideal ocorre um aumento no termo da energia. Os valores ideais foram tabulados para proteínas e ácidos nucleicos por meio do exame da estrutura de pequenas moléculas de aminoácidos, peptídeos, nucleotídeos e outros compostos similares. O programa X-PLOR fornece arquivos com a geometria de todos os aminoácidos, alguns nucleotídeos e moléculas de água, além de adicionar um termo eletrostático na função energia E_{total} , para os grupos carregados presentes na macromolécula. Durante cada ciclo de refinamento, o programa X-PLOR calcula dois índices de discordância (equações 3.7 e 3.8).

4.1. "Simulated annealing"

Para se sair do mínimo local, usa-se a técnica de dinâmica molecular ou "simulated annealing", que permite a superação de barreiras energéticas da

função E_{total} (Brunger, 1992). Na dinâmica molecular, o comportamento dinâmico de um sistema de partículas, neste caso átomos, é simulado.

O cálculo da dinâmica molecular começa com a atribuição de velocidades aos átomos da macromolécula, estas velocidades são derivadas de uma distribuição de Maxwell a uma dada temperatura. Num tempo $t=0\text{s}$, os átomos estão em uma configuração inicial com energia potencial E_{total} . Sobre cada átomo i na posição $\mathbf{r}_i$ existe uma força $\mathbf{f}_i$ agindo que é a derivada da energia potencial, $\mathbf{f}_i = -\partial E_{\text{total}} / \partial \mathbf{r}_i$. Da mecânica Newtoniana temos, $\mathbf{f}_i = m_i d^2 \mathbf{r}_i / dt^2$ e a aceleração $d^2 \mathbf{r}_i / dt^2$ para o átomo i de massa m_i pode ser calculada e então aplicada. Após um curto intervalo de tempo Δt , na faixa de femtosegundos ($1\text{fs} = 10^{-15}\text{s}$), o processo é repetido com os átomos em novas posições. Se o número de passo for suficiente, entre 103 e 104 passos, um novo mínimo de energia potencial E_{total} será atingido. Teremos também conseguido informações sobre o comportamento dinâmico dos átomos na estrutura, o que revela partes flexíveis presentes na macromolécula. Se a temperatura do sistema é elevada a um valor mais alto, mais átomos terão velocidades maiores, e uma energia cinética mais alta, e poderão superar barreiras energéticas.

A idéia básica da dinâmica molecular, como um método superior de refinamento, é elevar a temperatura do sistema alta o suficiente, entre 1000 e 4000 K, e então resfriá-lo lentamente para que a energia se aproxime do mínimo. A seguir, se processa mais alguns passos de refinamento posicional.

As coordenadas refinadas são usadas agora para gerar mapas de densidade eletrônica. Geralmente faz-se uso de dois tipos de mapas de

densidade eletrônica, os mapas $F_{\text{obs}} - F_{\text{calc}}$, onde F_{obs} é o fator de estrutura observado e F_{calc} o fator de estrutura calculado, as fases α_{calc} são aquelas calculadas a partir do modelo refinado. Este tipo de mapa é útil na localização de partes da estrutura que não fazem parte do modelo, como ligantes e moléculas de água, que não estão ainda na estrutura do modelo refinado. A densidade eletrônica é calculada usando-se a equação (2.13) e o algoritmo da transformada de Fourier Rápida (Drenth, 1994), o fator de estrutura $F(hkl)$ da equação (2.13) é substituído pela diferença, $(|F_{\text{obs}}| - |F_{\text{calc}}|)\exp(2\pi i\alpha_{\text{calc}})$. O outro tipo de mapa utilizado é o $2F_{\text{obs}} - F_{\text{calc}}$, com fases α_{calc} . Este mapa é útil na localização de pequenos desvios das cadeias laterais e para a observação do espectro geral da estrutura. Tal visualização é feita com a superposição das coordenadas do modelo refinado com o mapa de densidade eletrônica em terminais gráficos utilizando pacotes de programas, tais como XtalView (McRee, 1993) e O (Jones *et al.*, 1991). Após os ajustes feitos no modelo, novo ciclo de refinamento deve ser realizado, até que se tenham adicionado todos os ligantes e todas moléculas de água ao modelo, e o R_{free} e o R_{factor} tenham convergido.

5. Construção do modelo e validação

Durante o refinamento é muito importante avaliar se o mesmo está sendo realizado de acordo com o esperado, ou seja, se o modelo construído está em concordância com relação aos dados experimentais. Novamente lançamos mão do fator R (eq. 3.8) que vai verificar se está ocorrendo uma convergência de todos

os fatores de estrutura envolvidos F_{obs} e F_{calc} , respeitando-se as restrições químicas. No entanto, devido aos métodos usados, muitas vezes os programas de refinamento introduzem erros ('over-refinement') na sua busca de minimização. Para contornar este tipo de problema, foi criado um método que consiste em remover uma parte dos fatores de estrutura dos procedimentos de refinamento (Brunger, 1992). Dessa maneira, os seus dados não seriam utilizados pelo programa para serem conduzidos ao 'over-refinement'. Estes fatores seriam utilizados como parâmetros de controle: caso o refinamento estivesse conduzindo ao mínimo correto, os fatores livres também aumentariam a sua concordância quando verificados através do mesmo fator R (eq. 3.8). Caso contrário, o fator R calculado para os fatores livres (R_{free}) mostrariam um aumento de discordância (Lee, 2001).

Como vimos, o refinamento pode introduzir erros. Fases mal calculadas podem resultar em picos falsos, erroneamente ocupados por átomos do modelo. Assim, o fator R apesar de ser muito útil na orientação do refinamento de um modo geral, não nos dá informações precisas sobre detalhes do modelo.

Como nos certificamos, então, de que a estrutura real está correta? Existem várias formas de verificar se o modelo gerado representa a realidade da proteína, baseados principalmente em parâmetros estereoquímicos aos quais os átomos de cada resíduo de proteína estão relacionados (Lee, 2001).

Uma das formas de se verificar, é checando o enovelamento da cadeia principal, que pode ser analisado através do diagrama de Ramachandran (Ramachandran & Srinivasan, 1961), no qual os ângulos ϕ e ψ para cada resíduo

são representados em um gráfico. Estes dois ângulos podem variar de -180° a $+180^\circ$ na teoria, porém na prática existem combinações destes ângulos que levam a choques entre componentes dos resíduos, constituindo valores proibitivos. O gráfico de Ramachandran mostra exatamente as combinações de ϕ e ψ permitidas e o posicionamento dos resíduos nesse gráfico, permite visualizar os resíduos que estão infringindo essa regra.

Além disso, outros parâmetros como distância e ângulo de ligação entre os diferentes átomos, planaridade das ligações peptídicas e de algumas cadeias laterais, distância das interações não-covalentes, quiralidade de algumas cadeias laterais, acesso ao solvente de resíduos, natureza dos resíduos em regiões específicas da molécula, como um resíduo carregado estar ou não dentro de um bolsão hidrofóbico, e outros parâmetros estereoquímicos que podem ajudar a avaliar a qualidade do modelo refinado. Todos esses valores são comparados a um conjunto padrão de valores, extraídos de estruturas selecionadas que foram determinadas a alta resolução, bem como dados provenientes de cristalografia de pequenas moléculas (Lee, 2001).

Ao analisar estes valores, é sempre importante lembrar que a correspondência exata aos padrões não é imperativa. Poucos e determinados desvios do padrão geralmente requer uma inspeção mais cuidadosa, pois estes casos podem refletir exceções às regras. Estas exceções, mantidas às custas de estados não-mínimos de energia, podem ser cruciais e indispensáveis à atividade da proteína ou enzima, sendo geralmente justificáveis. No entanto, uma estrutura que apresente consistentemente erros de natureza estereoquímica geralmente

reflete erros de refinamento ou em alguma etapa do estudo, e devem ser vistos com maior cautela.



"Screen I" (Jancarik e Kim, 1991).

Sal	Tampão	Precipitante
1. 0,02M Cloreto de cálcio	1. 0,1M Acetato pH4,6	1. 30% MPD
2. ----	2. ----	2. 0,4M Tartarato de sódio e potássio
3. ----	3. ----	3. 0,4M Fosfato de amônio
4. ----	4. 0,1M TRIS pH8,5	4. 2,0M Sulfato de amônio
5. 0,2M Citrato de sódio	5. 0,1M HEPES pH7,5	5. 30% MPD
6. 0,2M Cloreto de magnésio	6. 0,1M TRIS pH8,5	6. 30% PEG 4000
7. ----	7. 0,1M Cacodilato pH6,5	7. 1,4M Acetato de sódio
8. 0,2M Citrato de sódio	8. 0,1M Cacodilato pH6,5	8. 30% Isopropanol
9. 0,2M Acetato de amônio	9. 0,1M Citrato pH5,6	9. 30% PEG 4000
10. 0,2M Acetato de amônio	10. 0,1M Acetato pH4,6	10. 30% PEG 4000
11. ----	11. 0,1M Citrato pH5,6	11. 1,0M Fosfato de amônio monobásico
12. 0,2M Cloreto de magnésio	12. 0,1M HEPES pH7,5	12. 30% Isopropanol
13. 0,2M Citrato de sódio	13. 0,1M TRIS pH8,5	13. 30% PEG 400
14. 0,2M Cloreto de cálcio	14. 0,1M HEPES pH7,5	14. 28% PEG 400
15. 0,2M Sulfato de amônio	15. 0,1M Cacodilato pH6,5	15. 30% PEG 8000
16. ----	16. 0,1M HEPES pH7,5	16. 1,5M Sulfato de lítio monobásico
17. 0,2M Sulfato de lítio	17. 0,1M TRIS pH8,5	17. 30% PEG 4000
18. 0,2M Acetato de magnésio	18. 0,1M Cacodilato pH6,5	18. 20% PEG 8000
19. 0,2M Acetato de amônio	19. 0,1M TRIS pH8,5	19. 30% Isopropanol
20. 0,2M Sulfato de amônio	20. 0,1M Acetato pH4,6	20. 25% PEG 4000
21. 0,2M Acetato de magnésio	21. 0,1M Cacodilato pH6,5	21. 30% MPD
22. 0,2M Acetato de sódio	22. 0,1M TRIS pH8,5	22. 30% PEG 4000
23. 0,2M Cloreto de magnésio	23. 0,1M HEPES pH7,5	23. 30% PEG 400
24. 0,2M Cloreto de cálcio	24. 0,1M Acetato pH4,6	24. 20% Isopropanol
25. ----	25. 0,1M Imidazol pH6,5	25. 1,0M Acetato de sódio
26. 0,2M Acetato de amônio	26. 0,1M Citrato pH5,6	26. 30% MPD
27. 0,2M Citrato de sódio	27. 0,1M HEPES pH7,5	27. 20% Isopropanol

28. 0,2M Acetato de sódio	28. 0,1M Cacodilato pH6,5	28. 30% PEG 8000
29. ----	29. 0,1M HEPES pH7,5	29. 0,8M Tartarato de sódio e potássio
30. 0,2M Sulfato de amônio	30. ----	30. 30% PEG 8000
31. 0,2M Sulfato de amônio	31. ----	31. 30% PEG 4000
32. ----	32. ----	32. 2,0M Sulfato de amônio
33. ----	33. ----	33. 4,0M Formiato de sódio
34. ----	34. 0,1M Acetato pH4,6	34. 2,0M Formiato de sódio
35. ----	35. 0,1M HEPES pH7,5	35. 0,8M Fosfato de potássio monobásico; 0,8M Fosfato de sódio monobásico
36. ----	36. 0,1M TRIS pH8,5	36. 8% PEG 8000
37. ----	37. 0,1M Acetato pH4,6	37. 8% PEG 4000
38. ----	38. 0,1M HEPES pH7,5	38. 1,4M Citrato de sódio
39. ----	39. 0,1M HEPES pH7,5	39. 2% PEG 400 + 2,0 M Sulfato de amônio
40. ----	40. 0,1M Citrato pH5,6	40. 20% Isopropanol + 20% PEG 4000
41. ----	41. 0,1M HEPES pH7,5	41. 10% Isopropanol + 20% PEG 4000
42. 0,05M Fosfato de potássio monobásico	42. ----	42. 20% PEG 8000
43. ----	43. ----	43. 30% PEG 1500
44. ----	44. ----	44. 0,2M Formiato de magnésio
45. 0,2M Acetato de zinco	45. 0,1M Cacodilato pH6,5	45. 18% PEG 8000
46. 0,2M Acetato de cálcio	46. 0,1M Cacodilato pH6,5	46. 18% PEG 8000
47. ----	47. 0,1M Acetato pH4,6	47. 2,0M Sulfato de amônio
48. ----	48. 0,1M TRIS pH8,5	48. 2M Fosfato de amônio monobásico
49. 1,0M Sulfato de lítio monobásico	49. ----	49. 2% PEG 8K
50. 0,5M Sulfato de lítio	50. ----	50. 15% PEG 8K



