

UNIVERSIDADE ESTADUAL DE CAMPINAS

SANDRA KRAUCHENCO

**ESTUDOS CRISTALOGRAFICOS DE DOIS INIBIDORES DE PROTEINASES
TIPO STI-KUNITZ EXTRAÍDOS DAS SEMENTES DE *Copaífera langsdorffii* E DE
Delonix regia RESPECTIVAMENTE.**

Tese apresentada ao Instituto de Biologia
para obtenção do Título de Doutora em
Biologia Funcional e Molecular, na área
de Bioquímica.

Orientador: Prof. Dr. IGOR POLIKARPOV

2004

Data da Defesa: 12 / 03 / 2004

Banca Examinadora

Titulares

Prof. Dr. IGOR POLIKARPOV (Orientador)

Prof. Dr. JOSÉ CAMILLO NOVELLO

Prof. Dr. SÉRGIO MARANGONI

Prof. Dra. PAULA REGINA KUSER

Prof. Dr. RICHARD CHARLES GARRAT

Suplentes

Prof. Dr. MUNIR SALOMÃO SCAF

Esse trabalho é dedicado a todas as pessoas que se foram antes de nós;

Pessoas que tornaram todas as nossas vidas melhores.

AGRADECIMENTOS

A DEUS que torna todas as coisas possíveis por me amar tanto assim.

Ao Prof Dr Igor Polikarpov meu orientador a quem devo muito, mas muito mesmo.

Ao Dr Ronaldo Alves Pinto Nagem cujo trabalho foi fundamental para a resolução da estrutura de CTI e que me ensinou muito sobre cristalografia e física.

A Valéria Peyrl Forrer, Renata Carmona e Ferreira e Jackson Antônio Lamounier Camargo Resende pela valiosa ajuda com a cristalização.

Ao José Ribeiro Brandão Neto pela ajuda fundamental com a coleta e processamento dos dados de difração.

Ao José Antônio da Silva por fornecer CTI purificado, pela ajuda com a preparação do complexo CTI-Tripsina e por esclarecer sempre minhas muitas dúvidas biológicas.

A Silvana Cristina Pando por fornecer DrTI purificado e pela ajuda com a preparação do complexo DrTI-Tripsina.

Ao Prof Dr Sérgio Marangoni coordenador do Laboratório de Química de Proteínas do Departamento de Bioquímica do Instituto de Biologia da UNICAMP cuja colaboração com meu orientador permitiu o contato com os alunos José Antônio da Silva e Silvana Cristina Pando.

Agradecimentos especiais a Dra Paula Regina Kuser Falcão que me ensinou muito sobre tudo e não apenas cristalografia. Muito obrigada.

Ao Dr Carlos Basílio Pinheiro por me mostrar o caminho e ter jogado um pouco de luz na minha vida num dos momentos cruciais desse doutorado. Valeu mesmo bicho!

Aos meus pais cujo suporte me permitiu a escalada até aqui.

Aos meus irmãos e amigos de verdade com quem aprendo e enriqueço todos os dias por quem agradeço simplesmente por estar viva porque vale à pena.

Aproveite bem cada momento da sua vida,
porque a vida é um delicioso fruto maduro que amanhã estará podre.

A vida é agora!

RESUMO

As estruturas cristalográficas de dois novos inibidores pertencentes à família STI Kunitz foram resolvidas. O inibidor extraído das sementes de *Copaifera langsdorffii* (CTI) foi resolvido usando MIRAS com derivados preparados por *quick cryo-soaking* e refinado até os residuais cristalográficos R_{factor} de 17,3% e R_{free} de 20,3% a 1,8Å de resolução. E o inibidor extraído das sementes de *Delonix regia* (DrTI) foi resolvido por substituição molecular usando a estrutura de CTI como modelo de busca e refinado até valores de 21,5% para R_{factor} e de 25,3% para R_{free} a 1,75Å de resolução.

Embora sejam membros da mesma família, cada inibidor é particularmente interessante por suas peculiaridades. CTI é composto por duas cadeias polipeptídicas não covalentemente ligadas e possui apenas uma ponte dissulfeto ao invés de uma cadeia polipeptídica única e duas pontes dissulfeto que caracterizam um inibidor padrão pertencente à família STI-Kunitz. Apesar dessas diferenças estruturais, CTI é um inibidor efetivo de tripsina com uma constante de dissociação de 1,2nM da mesma ordem de grandeza que os inibidores padrões como STI, ETI ou WCI. A sua estrutura cristalográfica revela que isso se deve ao fato de que as discrepâncias estruturais não afetam a conformação canônica do seu loop reativo.

Já a estrutura cristalográfica de DrTI revela que, embora ele tenha uma cadeia polipeptídica com duas pontes dissulfeto como o esperado, a conformação do seu loop reativo é distorcida pela inserção de um resíduo, Glu A68, entre os resíduos na posição P1 e P2, o que enfraquece a sua atividade contra tripsina e DrTI apresenta uma constante de dissociação de 22nM. Além disso, a composição do loop reativo de DrTI também é diferente com a presença de dois ácidos glutâmicos A67 e A68 e DrTI apresenta uma atividade inibitória diferenciada sendo um inibidor efetivo de calicreína do plasma humano com uma constante de dissociação de 5nM embora não apresente qualquer atividade contra calicreína de tecido.

Em ambos os casos a estrutura cristalográfica ajuda a compreender a base molecular para o comportamento particular de cada inibidor e esses resultados são apresentados e discutidos aqui.

ABSTRACT

The crystallographic structures of two new inhibitors that belong to the Kunitz STI family were solved. The inhibitor extracted from seeds of *Copaifera langsdorffii* (CTI) was solved by quick cryo-soaking MIRAS procedure and refined to a crystallographic residual R_{factor} of 17.3% and R_{free} of 20.3% at 1.8Å resolution. And the inhibitor from seeds of *Delonix regia* (DrTI) was solved by molecular replacement method using the CTI structure as search model and refined to values of 21.5% of R_{factor} and 25.3% of R_{free} at 1.75Å resolution.

Although be members from the same family, each inhibitor is particularly interesting due to its peculiarities. CTI is composed by two non-covalently bound polypeptide chains and only one disulfide bridge. In spite of these structural differences, CTI is an effective trypsin inhibitor with a dissociation constant of 1.2nM, a value comparable to other standard Kunitz-type inhibitors like STI, ETI or WCI. Its crystallographic structure reveals that this is because the structural discrepancies do not affect the canonical conformation of the reactive loop of CTI.

In the same way, the crystallographic structure of DrTI reveals that, although it has one polypeptide chain with two disulfide bridges as expected, the conformation of its reactive loop is distorted from the canonical by the insertion of one residue, Glu A68, between the residues at P1 and P2 positions. This weakens the inhibitory activity of DrTI that presents a constant dissociation of 22nM against trypsin. Moreover, the composition of DrTI reactive loop is also different with the presence of two glutamic acids A67 and A68 and DrTI has a different inhibitory activity being an effective inhibitor of human plasma kallikrein with a dissociation constant value of 5nM although it does not present any activity against tissue kallikrein.

For both inhibitors the crystallographic structure helps to comprehend the molecular basis of particular behavior of each one. The results are presented and discussed here.

ÍNDICE

RESUMO.....	vii
ABSTRACT.....	viii
INTRODUÇÃO.....	1
PARTE I – AS PROTEINASES E SEUS INIBIDORES	
CAPÍTULO 1: AS PROTEINASES.....	5
CAPÍTULO 2: OS INIBIDORES DE PROTEINASES.....	21
CAPÍTULO 3: OS INIBIDORES DE SERINOPROTEINASES.....	35
CAPÍTULO 4: FUNÇÕES E APLICAÇÕES DOS INIBIDORES DE PROTEINASES....	75
PARTE II – CRISTALOGRAFIA DE PROTEÍNAS	
CAPÍTULO 5: FUNDAMENTOS DA CRISTALOGRAFIA DE PROTEÍNAS.....	83
CAPÍTULO 6: PRINCÍPIOS DA CRISTALIZAÇÃO DE MACROMOLÉCULAS.....	115
CAPÍTULO 7: AQUISIÇÃO E TRATAMENTO DOS DADOS DE DIFRAÇÃO.....	135
CAPÍTULO 8: MÉTODOS DE FASEAMENTO.....	151
CAPÍTULO 9: REFINAMENTO E VALIDAÇÃO DO MODELO.....	189
PARTE III – RESULTADOS E DISCUSSÃO	
CAPÍTULO 10: A HEXOQUINASE PII E A MUP.....	219
CAPÍTULO 11: CTI – O INIBIDOR EXTRAÍDO DAS SEMENTES DE <i>Copaífera langsdorffii</i>	221
CAPÍTULO 12: DfTI – O INIBIDOR EXTRAÍDO DAS SEMENTES DE <i>Delonix regia</i>	251
CONCLUSÕES.....	267
REFERÊNCIAS BIBLIOGRÁFICAS.....	269
SITES INTERESSANTE.....	283
ANEXOS: TRABALHOS PRODUZIDOS.....	285

INTRODUÇÃO

A cristalografia geométrica é um ramo das ciências exatas que se ocupa do estudo da matéria numa escala atômica e tem como objetivos a determinação, a classificação e a interpretação de estruturas geométricas de sólidos cristalinos. A cristalografia de proteínas ou biocristalografia se dedica particularmente ao estudo da estrutura tridimensional de macromoléculas de interesse biológico.

A estrutura tridimensional é uma peça fundamental na compreensão dos processos biológicos em seu nível mais básico: quais moléculas interagem, como elas interagem, como enzimas catalisam reações, como drogas atuam, etc. Com o conhecimento de estruturas tridimensionais é possível o desenvolvimento de drogas que atuem de forma específica minimizando efeitos colaterais. Pode-se citar como exemplo os inibidores que compõem o coquetel usado no tratamento da AIDS, que foram desenhados para uma ação específica contra a HIV-protease sem afetar colateralmente qualquer proteína humana. Obviamente isso se tornou possível a partir do conhecimento da estrutura atômica tridimensional, particularmente do sítio ativo, da protease do vírus.

Nessa tese estão descritas as estruturas tridimensionais determinadas por biocristalografia de dois inibidores da família STI Kunitz extraídos das sementes de *Copaifera langsdorffii* e de *Delonix regia*. Esse é o fruto de um trabalho de colaboração interdisciplinar envolvendo estudos bioquímicos que levantaram alguns aspectos

instigantes sobre esses inibidores e estudos estruturais que ajudaram a compreender em parte o comportamento bioquímico desses inibidores.

A tese é composta de três partes distintas. A primeira parte apresenta um apanhado geral da literatura disponível sobre as proteinases e os seus inibidores, particularmente sobre os inibidores de serinoproteinases, mostrando um pouco sobre os seus mecanismos de ação, seus papéis biológicos e as suas aplicações e potenciais na medicina, agricultura e no uso laboratorial. A segunda parte aborda a cristalografia de proteínas, descrevendo seu fundamento teórico e delineando as suas várias etapas e métodos usados na resolução de estruturas de macromoléculas biológicas de interesse. E na terceira parte são descritos os resultados obtidos e a discussão desses resultados, relacionando as estruturas com as particularidades funcionais de cada inibidor.

Para maiores informações é fortemente recomendado que a literatura aqui referenciada seja consultada, particularmente os *reviews* e os livros citados, e os *sites* recomendados sejam visitados. Muitas informações e idéias usadas na elaboração dessa tese foram tiradas desses *sites* e, como se pode constatar, muitas figuras usadas aqui foram adaptadas deles. A maioria dos *sites* foi elaborada e é constantemente atualizada por excelentes pesquisadores e suas equipes, que são os grandes responsáveis pela literatura disponível em livros e artigos sobre o assunto.

PARTE

I

**PROTEINASES E
OS SEUS INIBIDORES**

CAPÍTULO 1

AS PROTEINASES

1. INTRODUÇÃO

HANS NEURATH foi um dos primeiros cientistas a reconhecer que proteases não atuam somente como enzimas digestivas, mas desempenham numerosas outras funções no organismo. Proteases têm uma relação do tipo ‘berço-ao-túmulo’ com as proteínas. Elas assistem ao nascimento das proteínas removendo os resíduos de metionina iniciais. Elas participam na entrega em um destino apropriado pela remoção dos peptídeos sinais. Na morte, elas convertem tanto as proteínas exógenas (digestão de alimentos) quanto as endógenas em aminoácidos, os quais são então utilizados para a síntese de novas proteínas. Entretanto, a maior interação das proteínas com proteases acontece em sua vida adulta. O processamento de proteínas controla numerosas atividades do tipo ‘liga’ e ‘desliga’ e essas atividades por sua vez, são responsáveis por uma grande faixa de fenômenos biológicos como coagulação sanguínea, dissolução de coágulos, ação de hormônios, penetração das camadas internas do óvulo pelo espermatozóide, diferenciação, morte celular e apoptose. (NEURATH, 1989)

Dentre essa grande variedade de processos fisiológicos, a ação das proteases pode ser dividida em duas diferentes categorias (NEURATH, 1989):

1) **proteólise limitada**, na qual uma protease cliva apenas uma ou um número limitado de ligações peptídicas de uma determinada proteína levando à ativação ou maturação de uma proteína inativa e

2) **proteólise ilimitada**, na qual proteínas são degradadas em seus aminoácidos constituintes. A degradação das proteínas pode se dar de duas formas. Na primeira, as proteínas são primeiro conjugadas a um marcador e então rapidamente hidrolisadas pelo proteosoma na presença de ATP e a outra via consiste na compartimentalização das proteases, por exemplo, em lisossomas, com a transferência das proteínas para este compartimento onde são rapidamente degradadas.

2. NOMENCLATURA: PROTEASE, PROTEINASE OU PEPTIDASE?

A União Internacional de Bioquímica e Biologia Molecular (UIBBM) tem recomendado desde 1984 o uso do termo *peptidase* para o subgrupo de hidrolases de ligações peptídicas (Subclasse E.C 3.4.). O termo *protease* amplamente usado é sinônimo de *peptidase*. *Peptidases* compreendem dois grupos de enzimas: as endopeptidases e as exopeptidases, as quais clivam ligações peptídicas em pontos na proteína e removem aminoácidos sequencialmente do N- ou C-terminal respectivamente. O termo *proteinase* é também usado como sinônimo para *endopeptidase*. O esquema moderno de nomenclatura pode ser assim representado:

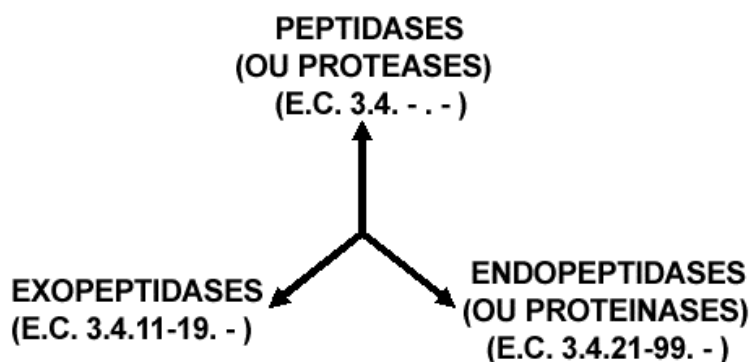


Figura 1-1: Esquema moderno da nomenclatura usada na classificação das proteases.

A nomenclatura para descrever a interação de um substrato com uma protease foi introduzida em 1967 por SCHECHTER e BERGER e é agora amplamente usada na literatura. Nesse sistema é considerado que os resíduos de aminoácidos do substrato polipeptídico se ligam em *subsítios* da enzima ativa. Por convenção, esses subsítios na protease são chamados S (para subsítios) e os resíduos de aminoácidos do substrato são chamados P (para peptídeo). Os resíduos de aminoácidos do lado N-terminal da ligação peptídica do substrato que é hidrolisada pela protease são numerados P_n, P(n-1), ..., P₃, P₂, P₁ e os resíduos do lado C-terminal são numerados P₁', P₂', P₃', ..., P(n-1)', P_n'. Os resíduos P₁–P₁' formam a ligação chamada “scissile” ou a ligação peptídica do substrato que é clivada pela protease na hidrólise. Os subsítios na protease que complementam os resíduos do substrato ligado são numerados S_n, S(n-1), ..., S₃, S₂, S₁, S₁', S₂', S₃', ..., S(n-1)', S_n' como mostrado na figura a seguir.

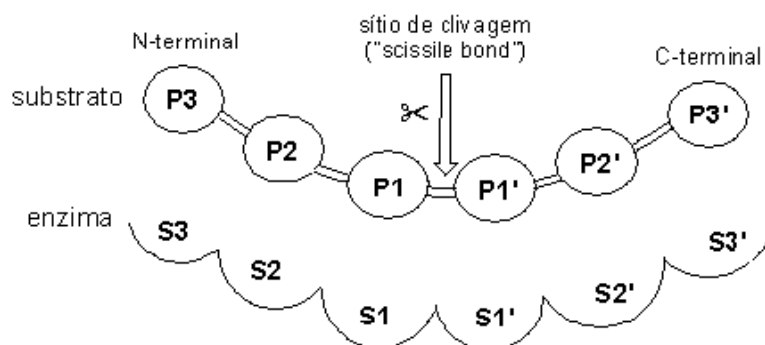


Figura 1-2: Nomenclatura dos subsítios de uma protease e os resíduos complementares do seu substrato.

Proteínases são classificadas atualmente de acordo com seus mecanismos catalíticos. Mas recentemente, tem se sugerido que essa classificação por tipos catalíticos seja estendida para uma classificação por famílias baseada nas relações evolutivas das proteases (RAWLINGS & BARRETT, 1993) e é essa classificação que está disponível no banco de dados SwissProt (<http://www.ebi.ac.uk/swissprot/>).

Quatro classes mecanísticas têm sido reconhecidas pela UIBBM: proteinases aspárticas, proteinases cisteínicas, metaloproteinases e serinoproteinases. Além destas quatro classes mecanísticas, existe ainda uma seção da nomenclatura das enzimas que é destinada para proteases cujo mecanismo catalítico ainda não tenha sido identificado.

3. PROTEINASES ASPÁRTICAS OU PROTEINASES ÁCIDAS

A maioria das proteinases ácidas (DAVIES, 1990) pertence à família da pepsina. A família da pepsina inclui enzimas digestivas como pepsina e quimosina bem como catepsinas D lisossomais e enzimas processadoras como renina e certas proteinases de fungos, como penicillopepsina, rizopuspepsina e endotiapepsina por exemplo. Uma segunda família compreende proteinases de vírus como a protease do vírus da AIDS (HIV) também chamada retropepsina.

Tabela 1-1: Atual classificação das proteinases aspárticas. (Tabela retirada do site <http://merops.sanger.ac.uk/indexes/families.html>)

FAMÍLIA	SUBFAMÍLIA	PROTEINASE REPRESENTANTE (EXEMPLO)
A1	A1	pepsin A (<i>Homo sapiens</i>)
A2	A2A	HIV-1 retropepsin (human immunodeficiency virus 1)
	A2B	Ty3 transposon (<i>Saccharomyces cerevisiae</i>)
	A2C	Gypsy transposon (<i>Drosophila melanogaster</i>)
	A2D	Osvaldo retrotransposon (<i>Drosophila buzzatii</i>)
	A2E	retrotransposon peptidase (<i>Schizosaccharomyces pombe</i>)
A3	A3A	cauliflower mosaic virus-type endopeptidase (cauliflower mosaic virus)
	A3B	bacilliform virus protease (rice tungro bacilliform virus)
A4	A4	aspergillopepsin II (<i>Aspergillus niger</i>)
A5	A5	thermopsin (<i>Sulfolobus acidocaldarius</i>)
A6	A6	nodavirus endopeptidase (flock house virus)
A8	A8	signal peptidase II (<i>Escherichia coli</i>)
A9	A9	spumapepsin (human spumaretrovirus)
A11	A11A	Copia transposon (<i>Drosophila melanogaster</i>)
	A11B	Ty1 transposon (<i>Saccharomyces cerevisiae</i>)
A12	A12	retrotransposon bs1 endopeptidase (<i>Zea mays</i>)
A21	A21	tetravirus endopeptidase (Nudaurelia capensis omega virus)
A22	A22A	presenilin 1 (<i>Homo sapiens</i>)
	A22B	signal peptide peptidase (<i>Homo sapiens</i>)
A24	A24A	type IV prepilin peptidase 1 (<i>Pseudomonas aeruginosa</i>)
	A24B	preflagellin peptidase (<i>Methanococcus maripaludis</i>)
A26	A26	omptin (<i>Escherichia coli</i>)
A29	A29	PibD g.p. (<i>Sulfolobus solfataricus</i>)

Estudos cristalográficos têm mostrado que essas enzimas são moléculas bilobuladas com o sítio ativo localizado entre os dois lóbulos homólogos. Cada lóbulo contribui com um resíduo aspartato da díade catalítica de aspartatos. Esses dois resíduos de ácido aspártico ficam em grande proximidade geométrica na molécula ativa e um ácido aspártico é ionizado enquanto o outro permanece não ionizado na faixa de pH ótima de 2–3.

Retropepsinas são monoméricas contendo apenas um ácido aspártico catalítico e se faz necessário haver uma dimerização para formar uma enzima ativa.

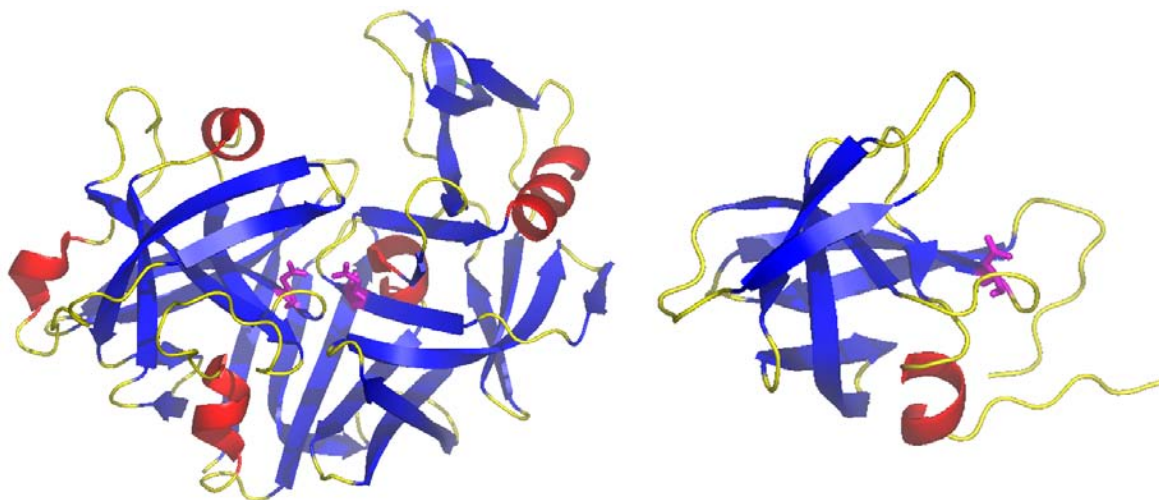


Figura 1-3: Arranjo característico das proteinases aspárticas. As folhas- β s estão em azul, as α -hélices em vermelho e os loops em amarelo. Esse esquema de cores será seguido para todas as estruturas representadas na tese, a menos que haja alguma indicação em contrário – o que estará devidamente explicado. Os dois resíduos de ácido aspártico do sítio ativo são mostrados em rosa. À esquerda está a estrutura da endotiapepsina que é bilobular e é uma molécula ativa e, à direita está a estrutura de retropepsina do HIV-1 que é monolobular e precisa que duas moléculas como essa se juntem para formar o dímero ativo.

Em contraste com as serinoproteinases e proteinases cisteínicas, a catálise por proteinases aspárticas não envolve um intermediário covalente ainda que um íon intermediário tetraédrico exista. O ataque nucleofílico é alcançado por duas transferências simultâneas de prótons: um de uma molécula de água para os dois grupos carboxila da díade e o segundo da díade para o oxigênio da carbonila do substrato com a clivagem da ligação C=O–NH. Essa catálise ácido–base, cujo mecanismo pode ser chamado “puxa–empurra”, leva à formação de um intermediário tetraédrico neutro não covalente. Uma representação esquemática do mecanismo catalítico das proteinases aspárticas é mostrada na figura a seguir. Os dois sucessivos “puxa–empurra” do mecanismo ocorrem nos passos indicados pelas setas (SALI & BLUNDELL, 1990).

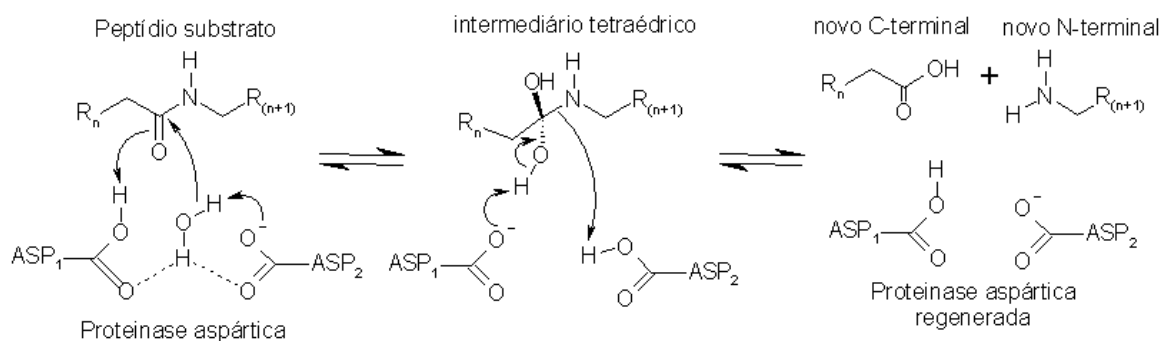


Figura 1-4: Representação esquemática do mecanismo de catálise ácido-base das proteinases aspartáticas.

4. PROTEINASES CISTEÍNICAS

Esta classe inclui proteases de origem vegetal como papaína, actinidina ou bromelaína, muitas catepsinas lisossomais de mamíferos, as calpaínas citosólicas (cálcio-ativadas) bem como diversas proteases de parasitas (por exemplos Trypanosoma, Schistosoma). Papaína é o arquétipo e o mais bem estudado membro da classe. Entretanto, a recente elucidação da estrutura cristalográfica da enzima interleucina-1- β revelou um novo tipo de arranjo para proteinases cisteínicas, como mostrado na figura 1-5.

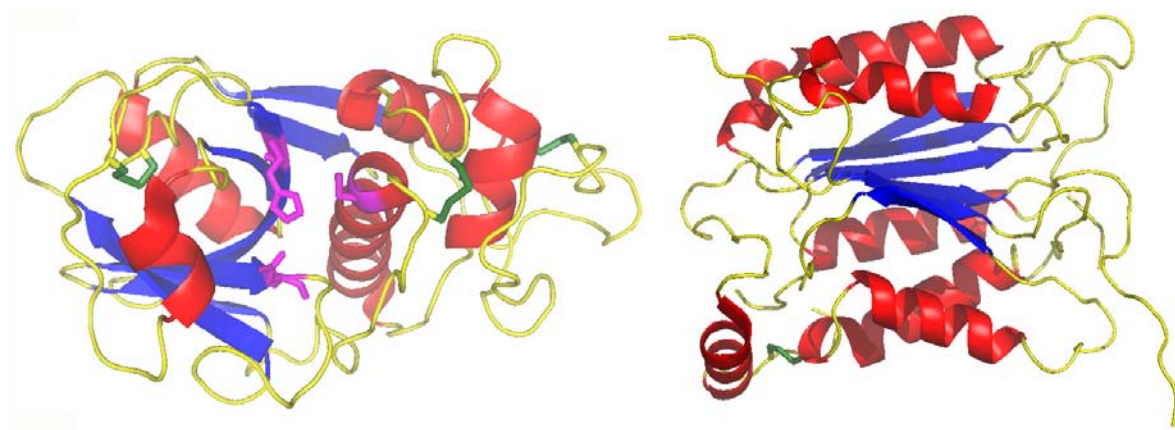


Figura 1-5: Arranjo característico das proteinases cisteínicas. Os resíduos da tríade catalítica Cys-His-Asn são mostrados em rosa. À direita está a papaína e à esquerda a interleucina-1- β : note como os arranjos são bastante distintos.

Para as proteinases cisteínicas, a catálise acontece através da formação de um intermediário tetraédrico e envolve uma cisteína e uma histidina. As essenciais Cys25 e His159 (numeradas de acordo com a papaína) desempenham o mesmo papel que as Ser195 e His57 nas serinoproteinases. O nucleófilo é um íon tiolato ao invés de um grupo hidroxila. O íon tiolato é estabilizado pela formação de um par iônico com o vizinho grupo imidazola da His159. O par iônico tiolato–imidazola é que promove o ataque nucleofílico em ambos os passos e então a molécula de água não é requerida. A representação esquemática do mecanismo das proteinases cisteínicas é mostrada na Figura 1-6.

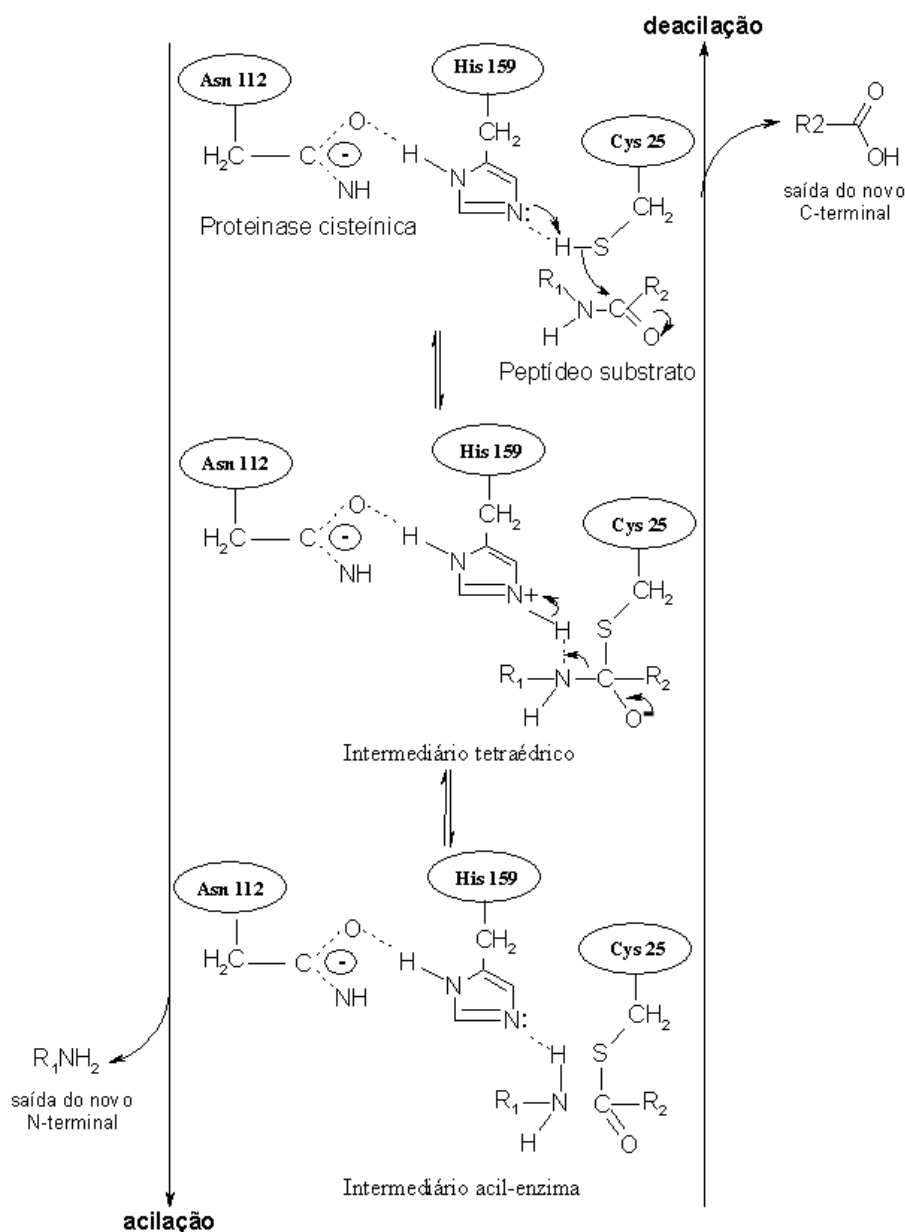


Figura 1-6: Representação esquemática do mecanismo catalítico das proteinases cisteínicas.

Tabela 1-2: Atual classificação das proteinases cisteínicas. (Tabela retirada do site <http://merops.sanger.ac.uk/indexes/families.html>)

FAMÍLIA	SUBFAMÍLIA	PROTEINASE REPRESENTANTE (EXEMPLO)
C1	C1A	papain (<i>Carica papaya</i>)
	C1B	bleomycin hydrolase (<i>Saccharomyces cerevisiae</i>)
C2	C2	Calpain-2 (<i>Homo sapiens</i>)
C3	C3A	poliovirus-type picornain 3C (human poliovirus type 1)
	C3B	poliovirus-type picornain 2A (human poliovirus type 1)
	C3C	foot-and-mouth disease virus picornain 3C (foot-and-mouth disease virus)
	C3D	Cowpea mosaic-type comovirus picornain 3C (cowpea mosaic virus)
	C3E	hepatitis A virus-type picornain 3C (hepatitis A virus)
	C3F	parechovirus picornain 3C (human parechovirus 1)
	C3G	rice tungro spherical virus-type endopeptidase (rice tungro spherical virus)
C4	C4	Nuclear-inclusion-a endopeptidase (plum pox virus)
C5	C5	Adenain (human adenovirus type 2)
C6	C6	potyvirus helper component proteinase (potato virus Y)
C7	C7	chestnut blight fungus virus p29 proteinase (<i>Cryphonectria hypovirus</i>)
C8	C8	chestnut blight fungus virus p48 proteinase (<i>Cryphonectria hypovirus</i>)
C9	C9	Sindbis virus-type nsP2 proteinase (Sindbis virus)
C10	C10	streptopain (<i>Streptococcus pyogenes</i>)
C11	C11	clostripain (<i>Clostridium histolyticum</i>)
C12	C12	ubiquitinyl hydrolase UCH-L1 (<i>Homo sapiens</i>)
C13	C13	legumain (<i>Canavalia ensiformis</i>)
C14	C14	Caspase-1 (<i>Rattus norvegicus</i>)
C15	C15	pyroglutamyl-peptidase I (<i>Bacillus amyloliquefaciens</i>)
C16	C16A	murine hepatitis coronavirus papain-like endopeptidase 1 (murine hepatitis virus)
	C16B	murine hepatitis coronavirus papain-like endopeptidase 2 (murine hepatitis virus)
C19	C19	ubiquitin-specific protease 14 (<i>Homo sapiens</i>)
C21	C21	tymovirus endopeptidase (turnip yellow mosaic virus)
C23	C23	carlavirus endopeptidase (apple stem pitting virus)
C24	C24	rabbit hemorrhagic disease virus 3C endopeptidase (rabbit hemorrhagic disease virus)
C25	C25	gingipain R (<i>Porphyromonas gingivalis</i>)
C26	C26	Gamma-glutamyl hydrolase (<i>Rattus norvegicus</i>)
C27	C27	rubella virus endopeptidase (Rubella virus)
C28	C28	foot-and-mouth disease virus L-proteinase (foot-and-mouth disease virus)
C30	C30	porcine transmissible gastroenteritis virus-type main protease (porcine transmissible gastroenteritis virus)
C31	C31	porcine respiratory and reproductive syndrome arterivirus-type cysteine proteinase alpha (lactate-dehydrogenase-elevating virus)
C32	C32	equine arteritis virus-type cysteine proteinase (porcine reproductive and respiratory syndrome virus)
C33	C33	equine arterivirus Nsp2-type cysteine proteinase (equine arteritis virus)
C36	C36	beet necrotic yellow vein furovirus-type papain-like endopeptidase (beet necrotic yellow vein virus)
C37	C37	calicivirin (Southampton virus)
C39	C39	bacteriocin-processing peptidase (<i>Pediococcus acidilactici</i>)
C40	C40	dipeptidyl-peptidase VI (<i>Bacillus sphaericus</i>)
C41	C41	cysteine proteinase (hepatitis E virus)
C42	C42	beet yellows virus-type papain-like endopeptidase (beet yellows virus)
C44	C44	glutamine phosphoribosylpyrophosphate amidotransferase precursor (<i>Homo sapiens</i>)

Tabela 1-2: Atual classificação das proteinases cisteínicas (continuação)

FAMÍLIA	SUBFAMÍLIA	PROTEINASE REPRESENTANTE (EXEMPLO)
C45	C45	acyl-coenzyme A:6-aminopenicillanic acid acyl-transferase precursor (<i>Penicillium chrysogenum</i>)
C46	C46	hedgehog protein (<i>Drosophila melanogaster</i>)
C47	C47	staphopain A (<i>Staphylococcus aureus</i>)
C48	C48	Ulp1 endopeptidase (<i>Saccharomyces cerevisiae</i>)
C50	C50	separase (<i>Saccharomyces cerevisiae</i>)
C51	C51	D-alanyl-glycyl endopeptidase (<i>Staphylococcus aureus</i>)
C53	C53	pestivirus Npro endopeptidase (classical swine fever virus)
C54	C54	Aut2 peptidase (<i>Saccharomyces cerevisiae</i>)
C55	C55	YopJ protease (<i>Yersinia pseudotuberculosis</i>)
C56	C56	PfpI endopeptidase (<i>Pyrococcus furiosus</i>)
C57	C57	vaccinia virus I7 processing peptidase (vaccinia virus)
C58	C58	YopT peptidase (<i>Yersinia pestis</i>)
C59	C59	penicillin V acylase (<i>Bacillus sphaericus</i>)
C60	C60A	sortase A (<i>Staphylococcus aureus</i>)
	C60B	sortase B (<i>Staphylococcus aureus</i>)
C61	C61	small protease (<i>Sulfolobus solfataricus</i>)
C62	C62	gill-associated virus 3C-like proteinase (gill-associated virus)
C63	C63	African swine fever virus processing peptidase (African swine fever virus)
C64	C64	Cezanne deubiquitinating peptidase (<i>Homo sapiens</i>)
C65	C65	Otubain 1 (<i>Homo sapiens</i>)
C66	C66	MAC protein (<i>Streptococcus pyogenes</i>)
C67	C67	CylD protein (<i>Homo sapiens</i>)

5. METALOPROTEINASES

As metaloproteinases podem ser uma das mais antigas classes de proteinases e são encontradas em bactérias, fungos e em organismos mais evoluídos. Elas diferem amplamente em suas seqüências e estruturas mas a grande maioria das enzimas contém um átomo de zinco o qual é cataliticamente ativo. Em alguns casos o zinco pode ser substituído por outro metal como cobalto ou níquel sem a perda da atividade.

Dentre as metaloproteinases, a termolisina bacteriana tem sido muito bem caracterizada e sua estrutura cristalográfica indica que o zinco está ligado a duas histidinas e um ácido glutâmico. Muitas enzimas contêm a seqüência HEXXH, a qual possui duas histidinas como ligantes para o zinco sendo que o terceiro ligante pode ser tanto um ácido glutâmico (termolisina, neprilisina, alanil aminopeptidase) ou uma outra histidina (astacina). Outras famílias exibem ligantes diferentes para o átomo de zinco.

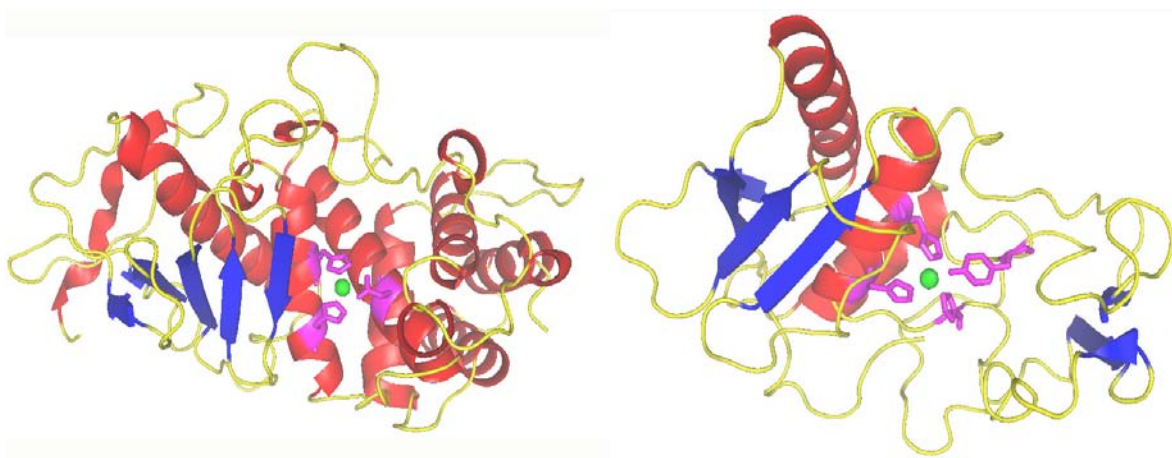


Figura 1-7: Arranjo característico das metaloproteinases. Os ligantes do metal zinco estão representados em rosa e o átomo de zinco em verde. À esquerda está a termolisina bacteriana com os ligantes HIS/HIS/GLU ao zinco e à direita está a astacina com três HIS e uma TYR ligadas ao átomo de zinco. Note como as estruturas são diferentes: termolisina possui estrutura secundária mais bem definida e um arranjo global mais estável.

O mecanismo catalítico das metaloproteinases leva à formação de um intermediário tetraédrico não covalente após o ataque do zinco ligado a uma molécula de água ao grupo carbonila da ligação “scissile” P1-P1’. Esse intermediário é decomposto pela transferência de um próton do ácido glutâmico para o grupo abandonador. Uma representação esquemática para o mecanismo catalítico das metaloproteinases é apresentada a seguir.

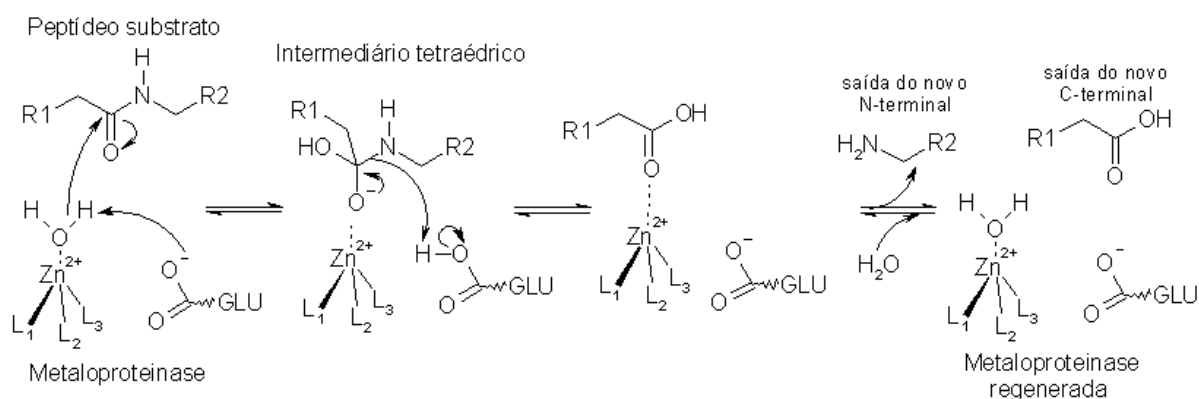


Figura 1-8: Representação esquemática do mecanismo catalítico das metaloproteinases.

Tabela 1-3: Atual classificação das metaloproteínases. (Tabela retirada do site <http://merops.sanger.ac.uk/indexes/families.html>)

FAMÍLIA	SUBFAMÍLIA	PROTEINASE REPRESENTANTE (EXEMPLO)
M1	M1	aminopeptidase N (<i>Homo sapiens</i>)
M2	M2	angiotensin-converting enzyme peptidase unit 1 (<i>Homo sapiens</i>)
M3	M3A	thimet oligopeptidase (<i>Rattus norvegicus</i>)
	M3B	oligopeptidase F (<i>Lactococcus lactis</i>)
M4	M4	thermolysin (<i>Bacillus thermoproteolyticus</i>)
M5	M5	mycolysin (<i>Streptomyces cacaoi</i>)
M6	M6	Immune inhibitor A (<i>Bacillus thuringiensis</i>)
M7	M7	snopalysin (<i>Streptomyces lividans</i>)
M8	M8	leishmanolysin (<i>Leishmania major</i>)
M9	M9A	microbial collagenase (<i>Vibrio alginolyticus</i>)
	M9B	collagenase colA (<i>Clostridium histolyticum</i>)
M10	M10A	collagenase 1 (<i>Homo sapiens</i>)
	M10B	serralysin (<i>Serratia marcescens</i>)
	M10C	fragilysin (<i>Bacteroides fragilis</i>)
M11	M11	gametolysin (<i>Chlamydomonas reinhardtii</i>)
M12	M12A	astacin (<i>Astacus astacus</i>)
	M12B	adamalysin (<i>Crotalus adamanteus</i>)
M13	M13	neprilysin (<i>Homo sapiens</i>)
M14	M14A	carboxypeptidase A1 (<i>Homo sapiens</i>)
	M14B	carboxypeptidase E (<i>Bos taurus</i>)
	M14C	Gamma-D-glutamyl (<i>Bacillus sphaericus</i>)
M15	M15A	zinc D-Ala-D-Ala carboxypeptidase (<i>Streptomyces albus</i>)
	M15B	vanY D-Ala-D-Ala carboxypeptidase (<i>Enterococcus faecium</i>)
	M15C	ply endolysin (bacteriophage A118)
	M15D	vanX D-Ala-D-Ala dipeptidase (<i>Enterococcus faecium</i>)
M16	M16A	pitrilysin (<i>Escherichia coli</i>)
	M16B	mitochondrial processing peptidase beta-subunit (<i>Saccharomyces cerevisiae</i>)
	M16C	eupitrilysin (<i>Homo sapiens</i>)
M17	M17	leucyl aminopeptidase (<i>Bos taurus</i>)
M18	M18	aminopeptidase I (<i>Saccharomyces cerevisiae</i>)
M19	M19	membrane dipeptidase (<i>Homo sapiens</i>)
M20	M20A	glutamate carboxypeptidase (<i>Pseudomonas sp.</i>)
	M20B	peptidase T (<i>Escherichia coli</i>)
	M20C	X-His dipeptidase (<i>Escherichia coli</i>)
	M20D	carboxypeptidase Ss1 (<i>Sulfolobus solfataricus</i>)
M22	M22	O-sialoglycoprotein endopeptidase (<i>Pasteurella haemolytica</i>)
M23	M23A	beta-lytic metalloendopeptidase (<i>Achromobacter lyticus</i>)
	M23B	lysostaphin (<i>Staphylococcus simulans</i>)
M24	M24A	methionyl aminopeptidase 1 (<i>Escherichia coli</i>)
	M24B	aminopeptidase P (<i>Escherichia coli</i>)
M26	M26	IgA1-specific metalloendopeptidase (<i>Streptococcus sanguinis</i>)
M27	M27	tentoxilysin (<i>Clostridium tetani</i>)
M28	M28A	aminopeptidase S (<i>Streptomyces griseus</i>)
	M28B	glutamate carboxypeptidase II (<i>Homo sapiens</i>)
	M28C	IAP aminopeptidase (<i>Escherichia coli</i>)
	M28D	aminopeptidase ES-62 (<i>Acanthocheilonema viteae</i>)
	M28E	aminopeptidase Ap1 (<i>Vibrio proteolyticus</i>)
M29	M29	aminopeptidase T (<i>Thermus aquaticus</i>)
M30	M30	hyicolysin (<i>Staphylococcus hyicus</i>)
M32	M32	carboxypeptidase Taq (<i>Thermus aquaticus</i>)
M34	M34	Anthrax lethal factor (<i>Bacillus anthracis</i>)
M35	M35	deuterolysin (<i>Aspergillus flavus</i>)

Tabela 1-3: Atual classificação das metaloproteinases (continuação).

FAMÍLIA	SUBFAMÍLIA	PROTEINASE REPRESENTANTE (EXEMPLO)
M36	M36	fungalsin (<i>Aspergillus fumigatus</i>)
M38	M38	beta-aspartyl dipeptidase (<i>Escherichia coli</i>)
M41	M41	FtsH endopeptidase (<i>Escherichia coli</i>)
M42	M42	glutamyl aminopeptidase (<i>Lactococcus lactis</i>)
M43	M43A	cytophagalsin (<i>Cytophaga sp.</i>)
	M43B	pappalysin-1 (<i>Homo sapiens</i>)
M44	M44	vaccinia virus-type metalloendopeptidase (vaccinia virus)
M48	M48A	Ste24 endopeptidase (<i>Saccharomyces cerevisiae</i>)
	M48B	HtpX endopeptidase (<i>Escherichia coli</i>)
M49	M49	dipeptidyl-peptidase III (<i>Rattus norvegicus</i>)
M50	M50A	S2P protease (<i>Homo sapiens</i>)
	M50B	sporulation factor SpoIVFB (<i>Bacillus subtilis</i>)
M52	M52	HybD endopeptidase (<i>Escherichia coli</i>)
M55	M55	D-aminopeptidase DppA (<i>Bacillus subtilis</i>)
M56	M56	BlaR1 peptidase (<i>Bacillus licheniformis</i>)
M57	M57	prtB g.p. (<i>Myxococcus xanthus</i>)
M60	M60	enhancin (Lymantria dispar nucleopolyhedrovirus)
M61	M61	glycyl aminopeptidase (<i>Sphingomonas capsulata</i>)
M63	M63	gpr protease (<i>Bacillus megaterium</i>)
M64	M64	IgA protease (<i>Clostridium ramosum</i>)
M66	M66	StcE protease (<i>Escherichia coli</i>)
M67	M67	Poh1 peptidase (<i>Saccharomyces cerevisiae</i>)
M72	M72	peptidyl-Asp metalloendopeptidase (<i>Pseudomonas aeruginosa</i>)

6. SERINOPROTEINASES

Essa classe compreende duas famílias distintas. A família da quimotripsina a qual inclui as enzimas de mamíferos como quimotripsina, tripsina, elastase e calicreína e a família da subtilisina a qual inclui as enzimas de bactérias como subtilisina. A estrutura 3D geral, veja na Figura 1-9, é diferente nas duas famílias mas elas têm a mesma geometria para o sítio ativo e por isso a catálise se dá pelo mesmo mecanismo. As serinoproteinases exibem diferentes especificidades por substrato, o que está relacionado com as substituições de aminoácidos nos vários subsítios das enzimas, os quais são nomeados de acordo com a nomenclatura de Schechter e Berger, interagindo com os resíduos do substrato. Algumas enzimas têm um sítio de interação estendida com o substrato enquanto outras interagem restritamente com o resíduo na posição P1 do substrato.

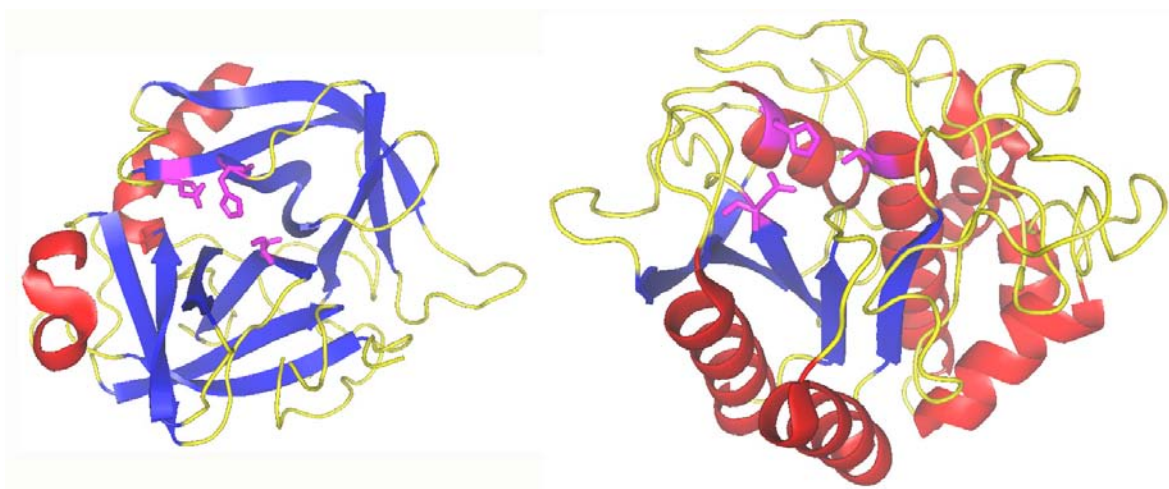


Figura 1-9: Arranjo característico das serinoproteinases. Os resíduos da tríade catalítica Asp102–Ser195–His57 estão representados em rosa. À direita está a quimotripsina e à esquerda está a subtilisina. Note como as estruturas são diferentes: quimotripsina possui maior quantidade de folhas- β s enquanto subtilisina maior conteúdo de α -hélices.

Três resíduos, os quais formam a chamada tríade catalítica, são essenciais no processo catalítico, são eles: His 57, Asp 102 e Ser 195 (seguindo-se a numeração do quimotripsinogênio). O primeiro passo na catálise é a formação de um intermediário acil-enzima entre o substrato e a Serina essencial. A formação desse intermediário covalente se dá através de um estado de transição intermediário negativamente carregado e então a ligação peptídica é clivada. Durante o segundo passo ou a deacilação, o intermediário acil-enzima é hidrolisado por uma molécula de água que desloca e substitui o peptídeo, que agora é o novo C-terminal do substrato clivado, e restaura a Ser-hidroxila da enzima. A deacilação, a qual envolve a formação de um estado de transição intermediário tetraédrico, acontece através do que é essencialmente o reverso da reação de acilação. A molécula de água é o nucleófilo que ataca o resíduo de serina. O resíduo de histidina atua como uma base geral e aceita o grupo OH da serina reativa. A seguir é mostrada a representação esquemática do mecanismo catalítico das serinoproteases.

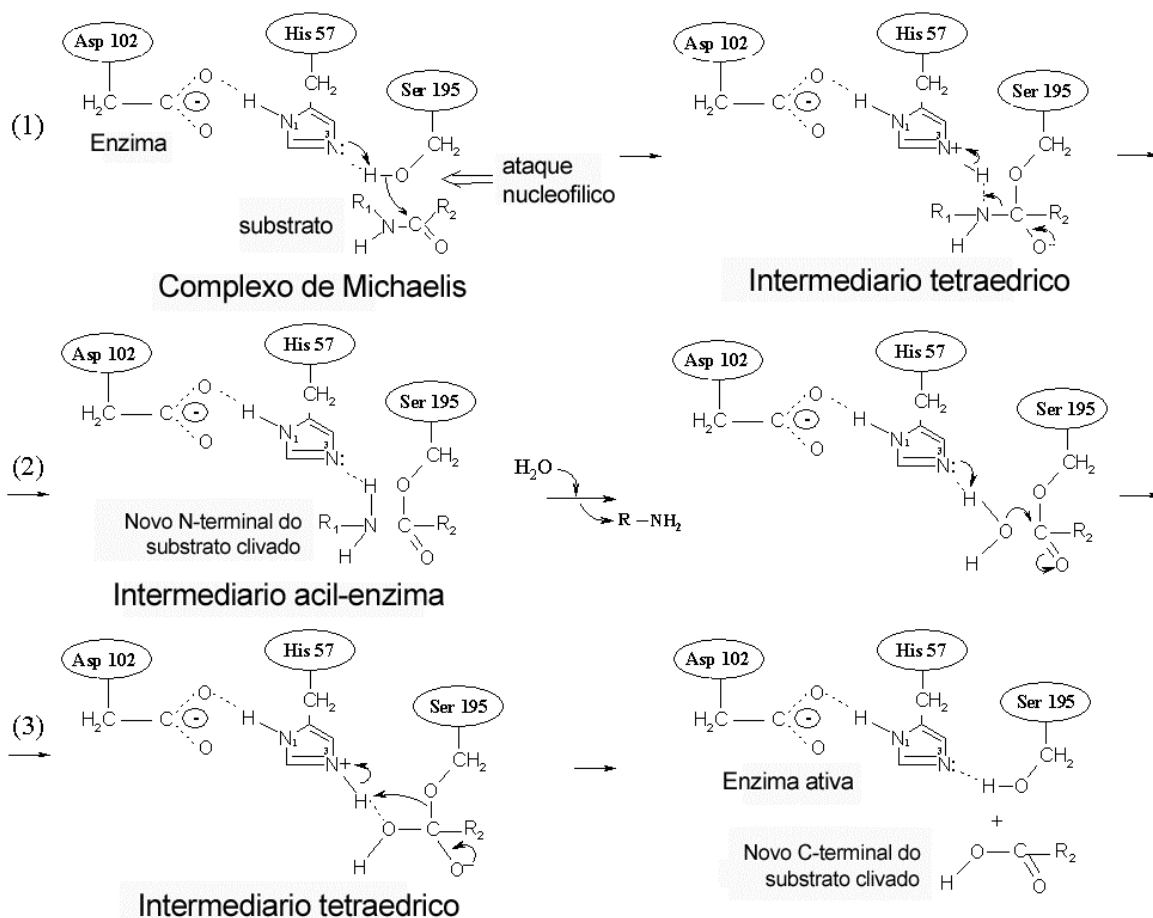


Figura 1-10: Representação esquemática do mecanismo catalítico das serinoproteínases.

Tabela 1-4: Atual classificação das serinoproteínases. (Tabela retirada do site <http://merops.sanger.ac.uk/indexes/families.html>)

FAMILY	SUBFAMILY	TYPE PEPTIDASE
S1	S1A	chymotrypsin A (<i>Bos taurus</i>)
	S1B	glutamyl endopeptidase I (<i>Staphylococcus aureus</i>)
	S1C	protease Do (<i>Escherichia coli</i>)
	S1D	lysyl endopeptidase (<i>Achromobacter lyticus</i>)
	S1E	streptogrisin A (<i>Streptomyces griseus</i>)
	S1F	astrovirus serine protease (Human astrovirus)
S3	S3	togavirin (Sindbis virus)
S6	S6	IgA1-specific serine endopeptidase (<i>Neisseria gonorrhoeae</i>)
S7	S7	Flavivirin (yellow fever virus)
S8	S8A	subtilisin Carlsberg (<i>Bacillus licheniformis</i>)
	S8B	kexin (<i>Saccharomyces cerevisiae</i>)
S9	S9A	prolyl oligopeptidase (<i>Sus scrofa</i>)
	S9B	dipeptidyl-peptidase IV (<i>Homo sapiens</i>)
	S9C	acylaminoacyl-peptidase (<i>Homo sapiens</i>)
	S9D	glutamyl endopeptidase (<i>Arabidopsis thaliana</i>)
S10	S10	carboxypeptidase Y (<i>Saccharomyces cerevisiae</i>)
S11	S11	D-Ala-D-Ala carboxypeptidase A (<i>Geobacillus stearothermophilus</i>)

Tabela 1-4: Atual classificação das serinoproteinases (continuação).

FAMÍLIA	SUBFAMÍLIA	PROTEINASE REPRESENTANTE (EXEMPLO)
S12	S12	D-Ala-D-Ala carboxypeptidase B (<i>Streptomyces sp.</i>)
S13	S13	D-Ala-D-Ala peptidase C (<i>Escherichia coli</i>)
S14	S14	endopeptidase Clp (<i>Escherichia coli</i>)
S15	S15	X-Pro dipeptidyl-peptidase (<i>Lactococcus lactis</i>)
S16	S16	lon protease (<i>Escherichia coli</i>)
S21	S21	cytomegalovirus assemblin (human cytomegalovirus)
S24	S24	repressor LexA (<i>Escherichia coli</i>)
S26	S26A	signal peptidase I (<i>Escherichia coli</i>)
	S26B	signalase (<i>Saccharomyces cerevisiae</i>)
S28	S28	lysosomal Pro-X carboxypeptidase (<i>Homo sapiens</i>)
S29	S29	hepacivirin (hepatitis C virus)
S30	S30	potyvirus P1 proteinase (plum pox virus)
S31	S31	pestivirus NS3 polyprotein peptidase (bovine viral diarrhea virus)
S32	S32	equine arteritis virus serine endopeptidase (equine arteritis virus)
S33	S33	prolyl aminopeptidase (<i>Neisseria gonorrhoeae</i>)
S35	S35	capillovirus serine protease (apple stem grooving virus)
S37	S37	PS-10 peptidase (<i>Streptomyces lividans</i>)
S39	S39A	sobemovirus proteinase (cocksfoot mottle virus)
	S39B	luteovirus proteinase (potato leaf roll luteovirus)
S41	S41A	C-terminal processing protease-1 (<i>Escherichia coli</i>)
	S41B	tricorn core protease (<i>Thermoplasma acidophilum</i>)
S43	S43	protein D2 porin (<i>Pseudomonas aeruginosa</i>)
S45	S45	penicillin G acylase precursor (<i>Escherichia coli</i>)
S46	S46	dipeptidyl-peptidase 7 (<i>Porphyromonas gingivalis</i>)
S48	S48	HetR endopeptidase (<i>Anabaena variabilis</i>)
S49	S49	protease IV (<i>Escherichia coli</i>)
S50	S50	infectious pancreatic necrosis birnavirus Vp4 protease (infectious pancreatic necrosis virus)
S51	S51	dipeptidase E (<i>Escherichia coli</i>)
S53	S53	sedolisin (<i>Pseudomonas sp. 101</i>)
S54	S54	Rhomboid-1 (<i>Drosophila melanogaster</i>)
S55	S55	SpoIVB peptidase (<i>Bacillus subtilis</i>)
S58	S58	aminopeptidase DmpA (<i>Ochrobactrum anthropi</i>)
S59	S59	nucleoporin 145 (<i>Homo sapiens</i>)
S60	S60	lactoferrin (<i>Homo sapiens</i>)

7. PROTEINASES DE MECANISMO DESCONHECIDO

Nessa classificação são consideradas as proteinases cujo mecanismo ainda é desconhecido. O fato da estrutura delas ainda não ser conhecida contribui bastante para o fato do seu mecanismo ainda não estar definido pois elas não podem ser enquadradas nas classes já conhecidas, muito embora para algumas os resultados prévios tenham mostrado comportamento diferente de qualquer proteinase conhecida. Na maioria desses casos é a simples falta de maiores informações sobre elas que as enquadram aqui.

Tabela 1-5: Proteinases de mecanismo catalítico desconhecido. (Tabela retirada do site <http://merops.sanger.ac.uk/indexes/families.html>)

FAMÍLIA	SUBFAMÍLIA	PROTEINASE REPRESENTANTE (EXEMPLO)
U4	U4	sporulation factor SpoIIIGA (<i>Bacillus subtilis</i>)
U6	U6	murein endopeptidase (<i>Escherichia coli</i>)
U9	U9	prohead proteinase (bacteriophage T4)
U32	U32	collagenase (<i>Porphyromonas gingivalis</i>)
U34	U34	dipeptidase A (<i>Lactobacillus helveticus</i>)
U35	U35	prohead protease (bacteriophage HK97)
U39	U39	hepatitis C virus endopeptidase 2 (hepatitis C virus)
U40	U40	protein P5 murein endopeptidase (bacteriophage phi-6)
U48	U48	prenyl protease 2 (<i>Saccharomyces cerevisiae</i>)
U49	U49	Lit protease (<i>Escherichia coli</i>)
U57	U57	yabG protein (<i>Bacillus subtilis</i>)
U61	U61	muramoyl-tetrapeptide carboxypeptidase (<i>Escherichia coli</i>)
U62	U62	microcin-processing peptidase 1 (<i>Escherichia coli</i>)
U64	U64	pseudomurein endoisopeptidase Pei (Methanobacterium phage psiM2)

CAPÍTULO 2

OS INIBIDORES DE PROTEINASES

1. INTRODUÇÃO

Como visto, as proteinases desempenham um amplo número de funções no organismo, mas, elas podem ser muito perigosas, por isso sua atividade é estritamente controlada para que sejam eficientes. Ter o sangue coagulado da cabeça aos pés por causa de um simples corte ou digerir o próprio estômago juntamente com a refeição podem ser exemplos de proteólise não controlada (LASKOWSKY Jr. *et al.*, 2000). Muitos mecanismos de controle das proteinases foram naturalmente desenvolvidos e dois predominam (LASKOWSKY & KATO, 1980).

O primeiro deles é que quase todas as proteinases são biosintetizadas como grandes precursores inativos chamados pró-proteínas ou zimogênios. Esses são armazenados e então ativados quando necessário. A ativação envolve a proteólise de uma ou várias ligações peptídicas da porção N-terminal em direção ao sítio catalítico da enzima. Essa localização do sítio de ativação assegura que o produto parcialmente sintetizado nunca seja ativo já que todas as proteínas são sintetizadas a partir do N-terminal.

Assim, outro mecanismo é necessário para controlar a enzima uma vez ativada. Isso é providenciado pela presença ubíqua de proteínas conhecidas como inibidores de

proteínases. Esses inibidores formam complexos completamente inativos ou parcialmente ativos com suas enzimas cognatas.

O termo *inibidor de proteinases* poderia a princípio denotar qualquer proteína que diminui a atividade enzimática de uma proteinase. Isso poderia trazer alguma confusão como, por exemplo, as proteinases que “inibem” outras proteinases digerindo-as mas, felizmente elas não são consideradas inibidores de proteinases. Entretanto dois grupos de proteínas completamente dissimilares são chamados inibidores de proteinases. O primeiro deles consiste das macroglobulinas que são proteínas de alta massa molecular aparentemente presentes no plasma de todos os mamíferos, freqüentemente em múltiplas formas e que não seguem um mecanismo padrão de associação com proteinases não sendo portanto específicos para uma determinada classe de proteinases. O segundo grupo é composto pelos inibidores classe específicos. (LASKOWSKY & KATO , 1980)

2. OS INIBIDORES PRESENTES NO PLASMA SANGÜÍNEO

Muitos processos proteolíticos como a coagulação sanguínea, dissolução de coágulos de sangue e a formação e destruição de hormônios ocorrem no plasma sanguíneo e requerem controle acurado. Por isso, não é surpreendente que o plasma sanguíneo contenha muitos tipos de inibidores de proteinases, alguns em concentrações bastante elevadas (LASKOWSKY & KATO, 1980).

Tabela.2-1: Principais inibidores de proteinases presentes no plasma sanguíneo humano.

Inibidor	Concentração (mg/100mL)	MM (kDa)	Cadeias polipeptídicas	Sítios Inibitórios
α_1 -proteinase	290 $\pm$ 45	52	1	1
α_1 -antiquimotripsina	49 $\pm$ 7	69	1	1
C ₁ -inativador	24 $\pm$ 3	70	1	1
α_2 -antiplasmina	7 $\pm$ 1	70	1	1
Antitrombina III	24 $\pm$ 2	65	1	1
Inter- α -tripsina	50	160	1	2
α_2 -macroglobulina	260 $\pm$ 70	720	4	1-2 ou mais

Dos inibidores listados na tabela 2-1, a cadeia do inibidor de inter- α -tripsina termina em dois domínios pertencentes à família BPTI-Kunitz e o sítio reativo se encontra nessa porção da cadeia. Por esse aspecto ele é discutido no capítulo 3 dedicado à família dos inibidores de serinoproteases. A α_2 -macroglobulina é uma armadilha geral para endopeptidases e os demais inibidores pertencem à família conhecida como serpinas (do inglês *serine proteinases inhibitors*) e ambos serão discutidos a seguir.

2.1. AS MACROGLOBULINAS

As macroglobulinas são exemplificadas pela α_2 macroglobulina humana (α_2 M) a qual tem uma massa molecular de 720kDa e é composta de quatro cadeias polipeptídicas idênticas. Ela dissocia em pares de cadeias após desnaturação e em cadeias simples após redução das pontes dissulfeto.

Quando α_2 M se combina com proteinases, apenas a atividade proteolítica sobre substratos protéicos grandes é diminuída ou eliminada. A atividade sobre substratos específicos e pequenos substratos sintéticos não é interrompida. Com efeito, os complexos α_2 M–proteinase formados podem ainda ser inibidos por pequenos inibidores de proteinases, como BPTI-Kunitz ou PSTI-Kazal por exemplo, mas não por inibidores maiores como STI-Kunitz por exemplo. Isso mostra claramente que nos complexos com α_2 M o sítio ativo das proteinases permanece aberto, apenas o acesso de substratos e inibidores a esse sítio é estericamente impedido. Esses fatos enfatizam o mecanismo de aprisionamento pelo qual as macroglobulinas agem. A proteinase hidrolisa uma ou mais ligações peptídicas particularmente susceptíveis da α_2 M o que induz uma mudança conformacional em α_2 M a qual aprisiona a molécula de enzima. Apenas moléculas de massa molecular muito elevada podem atuar por esse mecanismo.

Um outro aspecto marcante das macroglobulinas é a sua habilidade de se ligar a uma ampla variedade de proteinases pertencentes a todas as quatro classes mecanísticas. É relativamente raro encontrar endopeptidases que não formem complexo com macroglobulinas. Muitas pesquisas *in vivo* e *in vitro* comparam a importância da α_2 M com a de outros inibidores presentes no plasma sanguíneo (serpinas). Quando pequenas quantidades de enzima são misturadas com α_2 M e com um outro inibidor, complexos com

ambos são formados numa taxa que depende da constante de associação. Entretanto, o complexo com um inibidor típico, ainda que lentamente, dissocia enquanto o complexo com $\alpha_2\text{M}$ é irreversível. O resultado num experimento *in vitro* é que toda a proteinase é transferida para o complexo com $\alpha_2\text{M}$, a menos que haja tanta enzima presente que $\alpha_2\text{M}$ fique saturada. *In vivo* esse efeito é bastante acentuado sendo os complexos de $\alpha_2\text{M}$ com proteinases rapidamente formados: o tempo de meia-vida no homem é de 10 minutos. Essencialmente todas as proteinases injetadas são eliminadas pela via da $\alpha_2\text{M}$. Outros inibidores de serinoproteinases presentes no sangue estão envolvidos apenas transientemente na eliminação ou “limpeza” de proteinases.

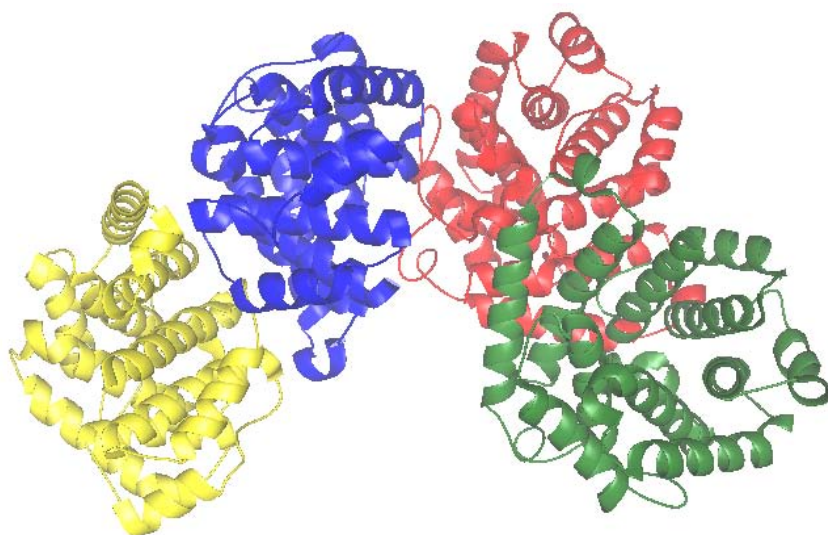


Figura 2-1: Estrutura 3D da α -2-macroglobulina (código PDB 1QSJ) mostrando as quatro cadeias idênticas que compõem a molécula.

Tudo isso mostra claramente que as macroglobulinas são de central importância no processamento de proteinases no sangue. Entretanto a sua interação com proteinases é tão diferente da de outros inibidores que, pelo menos do ponto de vista químico, a sua classificação como inibidor parece inadequada.

2.2. AS SERPINAS

Os inibidores chamados serpinas formam uma família de grandes (glico)proteínas homólogas. Assim como os inibidores canônicos, as serpinas parecem interagir com suas proteinases cognatas via um loop de ligação exposto. Os complexos resultantes são,

entretanto, apenas transientes e colapsam com a liberação de uma forma clivada de diferente estrutura e estabilidade. As serpinas exibem atividade inibitória contra uma ampla variedade de serinoproteinases. Entretanto, existem algumas proteínas que foram incluídas na família das serpinas por possuírem homologia, como é o caso da ovoalbumina, mas que não possuem propriedades inibitórias.

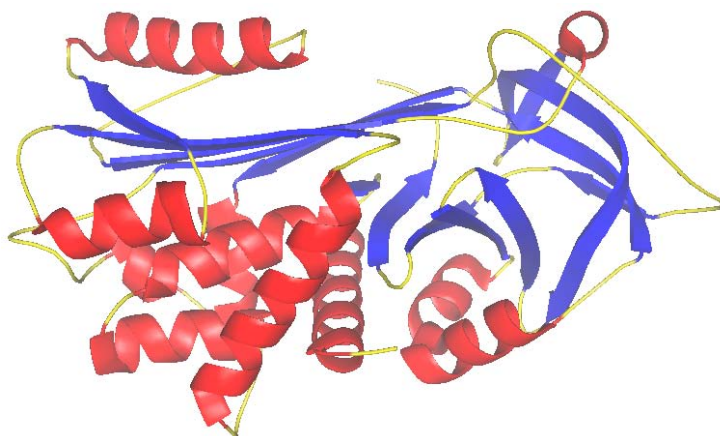


Figura.2-2: estrutura 3D da α_1 -antitripsina humana (código PDB 1QLP). Note a presença de dois domínios, um predominantemente alfa e o outro beta.

Com a determinação da sua estrutura 3D verificou-se que os serpinas possuem dois sítios de interação, sendo um deles responsável pelo reconhecimento da proteinase e o outro contém o loop reativo no qual ocorre a formação do complexo enzima–inibidor propriamente dito.

2.3. AS HIRUDINAS

Uma outra família de inibidores que não seguem o mecanismo padrão de associação descrito para os inibidores canônicos é a das hirudinas.

A saliva das sanguessugas contém uma variedade de inibidores de proteinases. A fim de manter o estado líquido do sangue ingerido, a maioria desses inibidores é dirigida contra proteases que estão envolvidas na coagulação sangüínea. A hirudina, purificada da saliva da sanguessuga *Hirudo medicinalis*, é um exemplo desses inibidores, sendo o mais potente inibidor natural de trombina conhecido. A constante de inibição de trombina por hirudina é da ordem de 10^{-14} M sendo o melhor valor de K_i descrito até agora. Entretanto,

como mencionado, a hirudina se liga à trombina seguindo um mecanismo de interação não canônico diferente do mecanismo padrão.



Figura 2-3.: alinhamento dos inibidores da família das hirudinas. As meias cistinas conservadas estão destacadas em preto.

A família das hirudinas é composta por inibidores homólogos de cerca de 7kDa com uma cadeia polipeptídica estabilizada por três pontes dissulfeto altamente conservadas. A estrutura desses inibidores está organizada em dois domínios, um na porção N-terminal (resíduos 3–48) formando um arranjo rígido estabilizado pelas três pontes dissulfeto, e outro na porção C-terminal (resíduos 49–65) formando uma cauda flexível.

Esse grupo de inibidores é atualmente produzido sinteticamente por várias empresas farmacêuticas e são usados na profilaxia e no tratamento de trombose e de outros distúrbios, sendo administrados por injeção subcutânea ou intramuscular ou em aplicações tópicas pois são absorvidos pela pele.

A estrutura das hirudinas consiste de um domínio globular N-terminal e um domínio C-terminal estendido (FOLKERS *et al.*, 1989) como mostrado na figura 2-4.

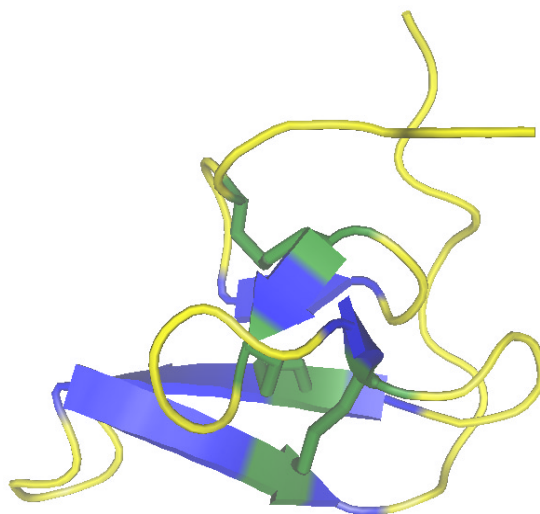


Figura 2-4: representação da estrutura 3D da hirudina de *Hirudo medicinalis* (código PDB 2HIR.). As pontes dissulfeto estão indicadas em verde .

3. INIBIDORES COM SÍTIO ATIVO CLASSE-ESPECÍFICOS

Outro grupo de inibidores de proteinases difere bastante das macroglobulinas. Para a esmagadora maioria deles as três afirmações a seguir são verdadeiras:

1. Nos complexos formados com as enzimas toda a atividade enzimática sobre substratos é totalmente abolida.
2. A inibição é estritamente competitiva.
3. Um sítio ativo inibitório particular pode inibir apenas proteinases pertencentes a uma das quatro classes mecanísticas.

Assim é conveniente dividir os inibidores em quatro classes de acordo com a proteinase inibida. São eles os inibidores de proteinases aspárticas, inibidores de proteinases cisteínicas, inibidores de metaloproteinases e os inibidores de serinoproteinases. Os três primeiros serão considerados em conjunto pois, apesar do crescimento exponencial de informações sobre eles, a quantidade de informações e estudos disponíveis sobre os inibidores de serinoproteinases ainda é esmagadoramente maior, e esses serão considerados separadamente.

4. INIBIDORES DE PROTEINASES ASPÁRTICAS, CISTEÍNICAS E METALO

Um número crescente de inibidores dessas proteinases tem sido reportados e são agrupados em famílias. O estabelecimento de uma família leva em consideração a homologia seqüencial entre os membros (o que implica na homologia estrutural entre eles), localização dos pares de meias-cistinas que formam as pontes dissulfeto e a composição e localização do sítio reativo. As famílias mais bem estabelecidas são apresentadas na tabela 2-2. Algumas das famílias descritas incluem proteínas que não são inibidores de proteinase, mas foram inseridas por apresentarem homologia seqüencial e estrutural bem como a conservação da quantidade e posição das pontes dissulfeto. Algumas dessas proteínas apresentam outras funções fisiológicas enquanto para outras ainda não se conhece a função.

Tabela 2-2: Famílias de inibidores de proteinases aspárticas, cisteínicas e metaloproteinases.

FAMÍLIA	MONÔMERO		ESTRUTURA 3D	
	MM _r (kDa)	½ CYS	LIVRE	COMPLEXO
INIBIDORES DE PROTEINASES ASPÁRTICAS				
Inibidores de catepsina	22–24	6	—	—
Inibidores de pepsina	22–24	4	1QS8	1F32
INIBIDORES DE CISTEINOPROTEINASES				
Estefinas	11	0	1STF	
Cistatinas	12	0	1EQK	
INIBIDORES DE METALOPROTEINASES				
TIMP	20–28	12	1BR9	1BQQ
Inibidores de	7	6	4CPA	—

4.1. FAMÍLIA DOS INIBIDORES DE CATEPSINA D

	10	20	30	40	50	60	70	80
1. SOLANUM TUBEROSUM	-----	ESPLPKPVLDTNGKELNPNSSYRIISIGRGALGGDVYLGKSPNSDAP	CPDGVF	-RYN				
2. SOLANUM NIGRUM	-----	PVLDTNGKELNPNSSYRIISIGRGALGGDVYLGKSPNSDAP	CPDGVF	-RYN				
3. LYCOPERSICON ESCULENTUM	-----	SPKPNPVLDTNGNELNPNSSYRIISTFWGALGGDVYLGKSPRSSAP	PLDGVF	-RYN				
4. NICOTIANA TABACUM	MKTNQFLPFLIFTISFNSFLSSAEAPPVVDIAGKKLRTGIDYYILPVVRGR	-GGGLTLDSTGNES	-CPLDAVVQEQ					
5. ARABIDOPSIS THALIANA	MSSLLYIFLLAVFISHRGVTTEAAVEP	-VKDINGKSLTGVNYYILPVIRGR	-GGGLTMSNLKTET	-CPTSVI	-QDQ			
6. NICOTIANA GLUTINOSA	----MKIISRIILLSCLLFLTLFQVKTEPVLDTNKQEIIRPGYTYILPAT	TGANGGGLTLAKGENGS	-CPLDVF	-QAQ				
7. CITRUS PARADISI	MKTSLITTLTSLILTLATQPLLGISSNSEPVLDIIGNKVKSTLEYKLI	PVLDGA	-GGFSTHLGTNGQ	-CPLDVI	-QLS			
8. ATROPA BELLADONNA	-----LVFSSSCAASLTNPPLPVVRDINGEILRSDTRYFVVSGLIGG	-SGGV	TQGPTVNGA	-CHDANFVCPS				
	90	100	110	120	130	140	150	160
1. SDVGPSGTPVRFIPLS	---GGIFEDQLLNIOFNIP	-----	TVKLCVSYTIWKVGNLNAYFRTMLLETGGTIGQADN	---SYFKIVKL				
2. SDVGPSGTPVRFIPLS	---GGIFEDQLLNIOFNIP	-----	TVKLCVSYTIWKVGNLNAYFRTMLLETGGTIGQADN	---SYFKIVKS				
3. SDVGTGTPVRFIPLS	---GGIFEDQLMNLQFNIA	-----	TVKLCVSYTIWKAGNLNAYRAMLETGGSIGQVDS	---SYFKIVKA				
4. QEIKN	-GLPLTFTFVNPKNKGVIR	-ESTDLNIKFSA	-----A	-SICVQTTLWKLD	DFDETTGKYFITIGNEGNGPRETISNWFKIEK			
5. FEVSQ	-GLPVKFSFYD	-KSRTIP	-VSTDVNIKFSPTS	SIWELANFDETTKQWFISTC	-----	GVEGNPGQKTVDNWFKIDK		
6. NSQSR	-GLPLKFLMVNSSAGLVIDENEDINIKFAA	-----	PRYVSI	CNKSTVWKIEDG	-----FVTGGIKGSGSENGTATSLFTIQK			
7. SPSE	-GDNLRLLPYD	-NYKIVKVSTDVNLRFV	-----	KVSVKECNEEPLWKVDSYDDP	-----LIVTGGSEGHGPAETLINWFKIEKA			
8. QVVHTPQDMGTSVYFKPKAPKQEEITEFTSLNIEFY	-----	LDNPTIC	KNNVWKVDGFP	GNEIPMFLSTNGVAGNPLN	---VSSWFQIEKV			
	190	200	210	220	230	240	250	260
1. ---SNFGYNLLSCPFTS	-IICLR	-CPEDQFC	AKVGVVIQN	-GKRRLLALVN	-ENPLDVLVFQEV	-----		
2. ---SKIGYNLLSCPFTS	-IICLR	-CPEDQFC	AKVGVVIQN	-GKRRLLALVN	-ENPLDVNFKEV	-----		
3. ---STFGYNLLYCP	ITRPVLCP	-CRGDDFC	AKVGVINQD	-GRRRLALVN	-ENPLGVYFKKV	-----		
4. ---FERDYKLYCP	---TVCNF	-CKVI	-CKDVGIPIQD	-GIRRLALS	-DVPFKVMFKKAQVVKD	-----		
5. ---FEKDYKIRFCP	---TVCNF	-CKVI	-CRDVGVFVQD	-GKRRLLALS	-DVPLKVMFKRAY	-----		
6. ---YEDAYALQYCP	---RAAGCSFI	CPRL	-GGYIGIAPAANGSRRLAVN	---RPIFKIEFKKASDTEVNFSQFDIYNRCGFSNTYMNKSY				
7. GNFSYGGVYNIHVCP	---SVCLS	-CPRR	-CNKIGVSTKD	-GVRRLLLVGDHEPALNVLFPPSEESSRLLQAANAISRFEEAAEASMTA				
8. ---HDNSYKLVFC	PYEEH	-ICSDI	IKVVD	-GQDHLAIRTNTFAVVFVKDPYIGIKSII	-----			

Figura 2-5: alinhamento entre algumas seqüências de membros da família de inibidores de catepsina D selecionados. Nenhum deles possui estrutura tridimensional determinada até o momento. As meias cistinas conservadas estão destacadas em preto.

Os inibidores de catepsina D foram primeiramente extraídos de batata (*Solanum tuberosum*) mas, como se pode ver no alinhamento da figura 2-5, ele já foi encontrado em várias outras plantas. Nenhuma estrutura 3D foi determinada para nenhum dos membros dessa família até o momento. Mas sabe-se que eles apresentam uma homologia

relativamente alta com os inibidores da família STI Kunitz (ver capítulo dedicado aos inibidores de serinoproteínases). Essa homologia é um forte indicativo de que as estruturas 3D dessas famílias sejam muito próximas. Além disso, esses também são inibidores efetivos de tripsina (HANNAPEL, 1993).

4.2. FAMÍLIA DOS INIBIDORES DE PEPSINA

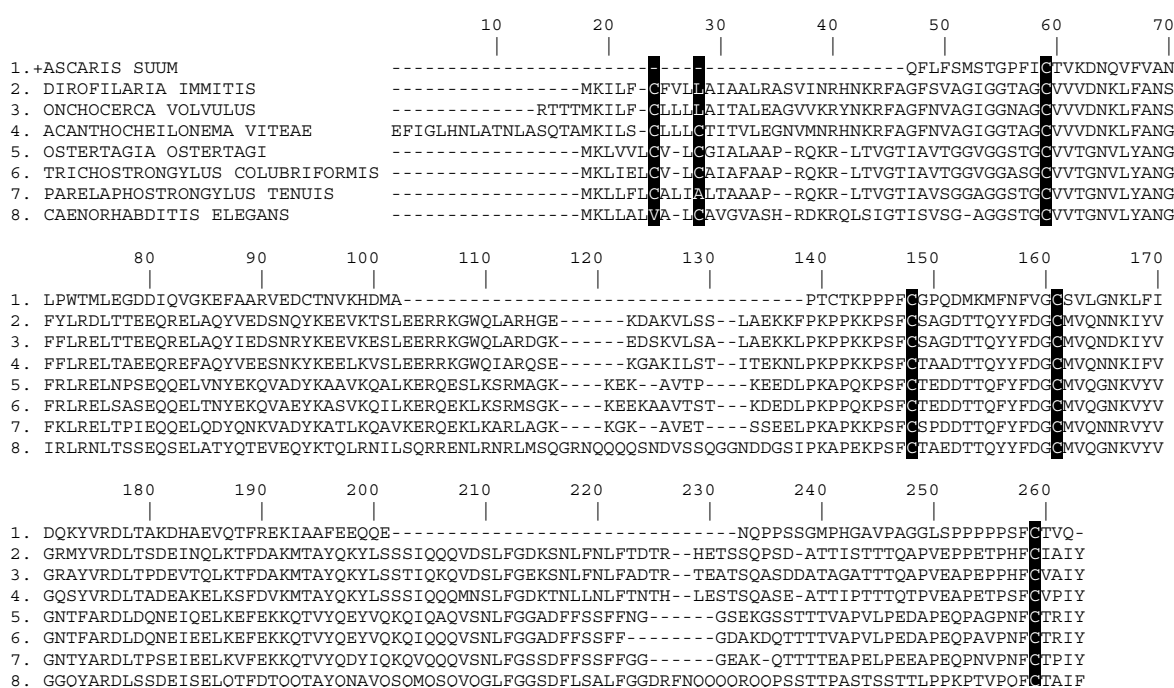


Figura 2-6: alinhamento entre algumas seqüências de membros da família inibidores de pepsina selecionados. Aquele marcado com + possui estrutura tridimensional determinada. As meias cistinas conservadas estão destacadas em preto.

Apesar das seqüências de vários membros dessa família serem conhecidas, como mostra o alinhamento da figura 2-6, apenas a estrutura do inibidor de pepsina de *Ascaris suum* foi recentemente determinada. A determinação da estrutura desse inibidor revelou um novo tipo de arranjo consistindo de dois domínios cada um, compreendendo uma folha- β flanqueada por uma α -hélice. A estrutura do complexo inibidor-pepsina também foi determinada mostrando que esse inibidor tem um novo modo de inibição: usando os resíduos do N-terminal para ocupar e bloquear os três primeiros sítios de ligação da pepsina. No complexo, a fita- β N-terminal do inibidor parecia com uma fita do sítio ativo da pepsina (resíduos 70-82) formando uma folha- β com oito fitas- β s que mantém as duas proteínas unidas (NG *et al.*, 2000).

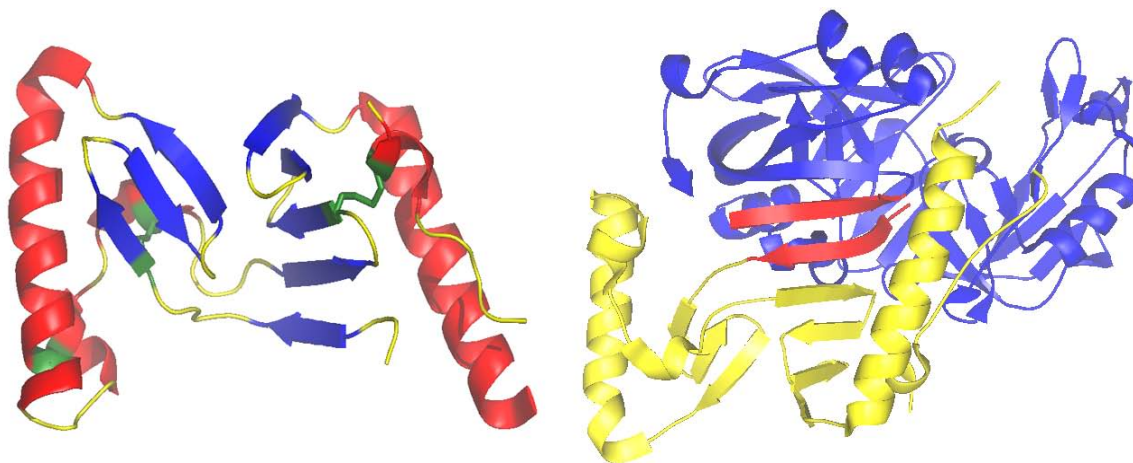


*Figura 2-7: À esquerda é mostrada a representação da estrutura 3D do inibidor de pepsina extraído do verme de porco *Ascaris suum* (código PDB 1F32). À direita é mostrado o complexo inibidor(amarelo)–pepsina(azul) com as fitas- β s que se unem para formar a folha- β destacadas em vermelho.*

4.3. FAMÍLIA DAS ESTEFINAS

Os inibidores de proteinases cisteínicas são amplamente distribuídos na natureza tendo sido purificados de mamíferos, aves, répteis, raízes e sementes e até mesmo de líquens. Eles são também conhecidos como a superfamília das cistatinas que é por sua vez dividida em três famílias: a família das estefinas, a família das cistatinas e a família dos cininogênicos. Uma característica marcante das três famílias é a ausência de pontes dissulfeto na estrutura da maioria dos seus membros.

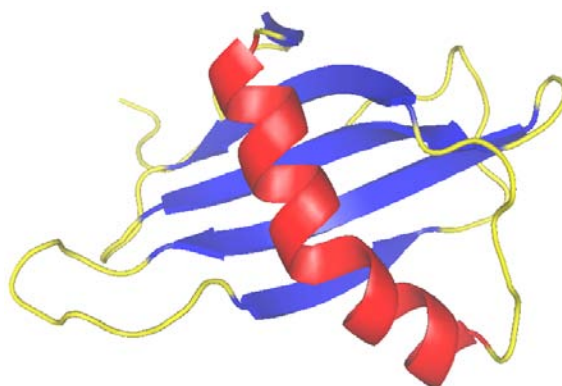


*Figura 2-8: representação da estrutura tridimensional do inibidor da estefina-B extraída de *Carica papaya* (código PDB 1STF).*

As estefinas formam a família dos menores inibidores naturais de proteinases cisteínicas com uma massa molecular relativa de cerca de 11kDa. A sua estrutura é bastante rígida a despeito da falta de pontes dissulfeto.

	10	20	30	40	50
1. HUMANA	GGPMDASV	-EEEGVRRALDF	AVGEYNKASND	MYHSRALQVVR	ARKQIVAG
2. RATTUS NORVEGICUS	AQEMDCND	---ESL	FQAVDTALKKYN	AGLKSGNQFVLYQ	VTEG-TKKDGS
3. MUS MUSCULUS	FGSINISN	---ANVKQCVWF	FAMKEYNKESED	KYVFLVDKILHAKL	QITDR
4. ONCHOCERCA VOLVULUS	GGWEDRDP	-KDEEILELLPS	ILMKVNEQSNDEY	HLMPIKLLKVSSQ	VVAG
5. SOLANUM TUBEROSUM	GGIINVNPNS	PEFQDLARFAVQ	DYNNNTQ--	NAHLEFVENLNVKE	QLVSG
6. DROSOPHILA MELANOGASTER	GGVSQLEGNSR	KEALELLD	ATLAQLATGD--	GPSYKAINVTSVTG	QVVAG
7. SARCOHAGA PEREGRINA	GCPSEVKGDKL	KQSEETLNKSL	SKLAAGD--	GPTYKLVKINSATT	QVVSG
8. TRIMERESURUS FLAVOVIDIS	RGDLECDD	---KEAKNWADDA	VRYINEHKLHG	HKQALNVIKNIC	VVPWNG

	60	70	80	90	100	110
1. V--	NYFLDVELGR	TTCTK---	TQPNLDNCP	PHDQPHLKR	KAFCSFQIY	AVPWQ--GTM-TLSKST
2. KT-	FYSFKYQI	KEGNC	SVQ--SGFAWQ	DCDFKDAE	-EAATGECTAT	LEKRRNN--KFS-IATQIC
3. M--	EYQIDVQI	SRNCKK---	PLNNTENCI	PQKPELEK	KMSCSFLV	GALPWNG-EFN-LLSKEC
4. V--	KYKMDVQ	VARSQCKK	SSNEKVDLT	KCKKLEG---	HPEKVM	TLVWEK
5. M--	MYITLA	ATDA-G----	NKKEYEAKI	WV---KEWED	FKKVIDF	KL VGN--DSA-KKLGGF
6. S-	LN	TYEVELD-----	NGSDKKQCT	VKIWT-QPWL	KENG	TNIIKCSG--DDG-ELDRTW
7. S--	KDVINAD	LKD-----	ENDKTKCD	ITIWS-QPWL	ENGIEVT	FNC
8. DLVAV	FLELN	LLETECHVL--	DPTFVEKCT	VRQQHN	HA	EMDCDAKIMFNVE---TFKRDV

Figura 2-9: alinhamento entre algumas seqüências de membros da família das estefinas.

4.4. FAMÍLIA CISTATINA

As sementes de arroz contêm um potente inibidor de cisteinoproteinases que é responsável pela hidrólise da reserva de glutelina durante a germinação (ABE *et al.*, 1985, 1987). Esse inibidor foi denominado oryzacistatina por suas propriedades similares às dos inibidores de origem animal da superfamília das cistatinas (BARRETT *et al.*, 1986). A seqüência de aminoácidos da oryzacistatina deduzida da seqüência nucleotídica do cDNA mostrou homologia significativa de 20–30% com diversas cistatinas animais. Um inibidor de cisteinoproteinases similar também foi isolado do endosperma de milho (ABE & WHITAKER, 1988).

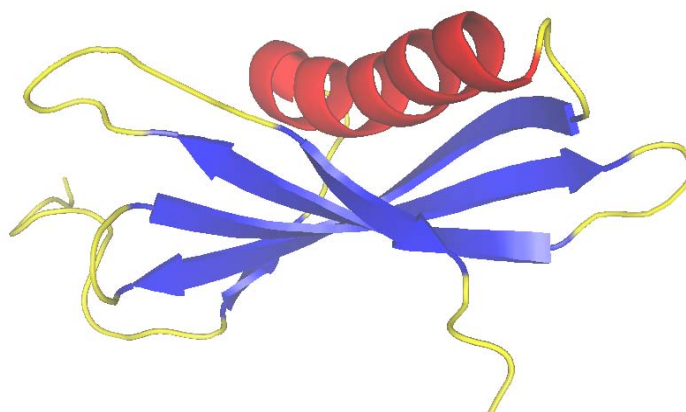


Figura 2-10: representação da estrutura tridimensional da oryzacistatina extraída de arroz *Oryza sativa japonica* (código PDB 1EQK).

		10	20	30	40	50
1. +ORYZA SATIVA	-MSSDGGPVLGGVEPV-GNENDLHLVDLARFAVTEHNKKANS--LLEFEK					
2. +GALLUS GALLUS	-----GAPVPVDENDEG--LQRALQFAMAEYNRASNDKYSSRVVR					
3. +HOMO SAPIENS A	-----MIPGGLSEAKPATP-EIQEIVDKVKPQLEKTNE--TYGKLE					
4. +HOMO SAPIENS B	-----MMSGAPSATQPATA-ETQHTADQVRSQLEEKYNK--KFPVFK					
5. +HOMO SAPIENS C	SSPGKPPRLVGGPMDASVEEAG--VRRALDFAVGEYNKASNDMYHSRALQ					
6. ZEA MAYS	SMADNTGTLAGGIKDVPGNENDLHLQELARFAVDEHNKKANA--LLGFKEK					
7. GLYCINE MAX	HHAPMAT--IGGLRDSQGSQNSVQTEALARFAVDEHNKKQNS--LLEFSR					
8. TRITICUM AESTIVUM	-----LARFAVSEHKNKTNA--LLEFEK					
9. LYCOPERSICON ESCULENTUM	-----GGIREAGGSENSLEINDLARFAVDEHNKKQNA--LLEFGK					
10. BRASSICA RAPA	----MAM--LGGVRDVPNSNSVEVESLARFAVDEHNKKENA--LLEFAR					
11. SORGHUM BICOLOR	-BGEESMALDGGIKDV PANENDLHLQELARFAVDEHNKKANA--LLGYEK					
12. SESAMUM INDICUM	----MAT--LGGVHDSNSNP--DTHSLARFAVDQHNTKENG--LLELVR					
13. IPOMOEIA BATATAS	NLIRMATTTLGGISDSASAENSVEIESLARFAVEHNKKENA--MIELVR					
14. ARABIDOPSIS THALIANA	----MAL--VGGVGDPANQNSGEVESLARFAVDEHNKKENA--LLEFAR					
15. SOLANUM TUBEROSUM	-----QTQNA--HLEFVE					
16. VIGNA UNGUICULATA	----MAA--LGGNRDVAGNQNSLEIDSLARFAVEHNKKQNA--LLEFGR					
17. COIX LACRYMA-JOBI	SMADDAGMLAGGIKDV PANENDLHLQELARFAVDEHNKKANA--LLGYEK					
18. CUCUMIS SATIVUS	----MSSSEIGGYVPCKDPNDP-HVKDIAEWAVEYNSQGH--HLLTVS					
19. CASTANEA SATIVA	----MAA--LVGGVSDVKGHENSLQIDDLARFAVDHNNKANT--LLQPKK					
20. MALUS DOMESTICA	----MAT--LGGVHESHGAQNSAEVDLARFAVQEHNNKENA--LLEFVS					
21. PYRUS COMMUNIS	----MAA--VGAVRDNQGVANSVETESLARYAVDEHNKKEND--LLEFVR					
22. CARICA PAPAYA	--MEPGIVIGGLQDVEGDANNLEYQELARFAVDEHNKKTNA--MLQPKR					
23. ONCORHYNCHUS KETA	AFTVANAGLIGGPM DANMNDQG--TRDALQFAVVEHNKKTNDMFVRQVAK					

	60	70	80	90	100	110	120
1. LVSVKQVQVAGTLYYFTIEVKEG--DAKK-LYEAKVWEKPMWDFKELQEFKPV--ASANA-----							
2. VISAKRQLVSGIKYILQVEIGRTTCPKSSGDLQSCFHDPEMAKYTTCTFVVSIPWLNQIKLLESKQ--							
3. AVQYKTQVAGTNYIYKVRAGDN--KYMHLKVKFSLPGQNEIDLVLTYGYQVDKNKDD--ELTGF--							
4. AVSFKSQVAGTNYFIKVHVGE--DFVHLRVFQSLPHENKPLTSLNYQTNKAKHD--ELTYF--							
5. VVRARKQIVAGVNYFLDVELGRITCTKTQPNLDNCPFHDPHLLKRAFCFSQIYAVPWQGTMTLSKSTCQDA							
6. LVKAKTQVAGTMYLTIIEVKDG--EVKK-LYEAKVWEKPMWDFKELQEFKPV--EGASA-----							
7. VVRTQEQVAGTLLHLLTLEAIEA--GEKK-LYEAKVWVKPWLNFKELEFQKPA--VPSFTSADLGVKKD							
8. VVRLKQVQVAGTMYIITIQVNEG--GAKK-MYEAKVWERPMDFKKLMEFRPAER--ASASA-----							
9. VVNKQVQVAGTMYIITILEATEG--GKKK-AYEAKVWVKPWQNFQVEDFKLIGD--AATA-----							
10. VVKAKEQVAGTMMHLLTLEAIEA--GKKK-LYEAKVWVKPWLNFKELEFQKPS--TTITPSDLGCKKG							
11. LVKAKTQVAGTMYLTIIEVKDG--EVKK-LYEAKVWEKPMWDFKELQEFKPV--EGASA-----							
12. VVEAREQVAGTLLHLLVLEVLDA--GKKK-LYEAKIWKPMWDFKELQEFKHVRD--VPSFTSSDLGAKTD							
13. VVKAEEQVAGTLLHLLTLEVIDA--GKRK-LYEAKVWLKPMWDFKELQGFNHIED--IPTLTSSDLGARRD							
14. VVKAKEQVAGTLLHLLTLEAIEA--GQKK-LYEAKVWVKPWLNFKELEFQKPA--APAITSSDLGCKQG							
15. VLNKQVQVAGTMMYIITLVATDD--GYKK-IYKTKIWKKEWENFKEVQEFKQIVY--ATK-----							
16. VVSAQQVQVSGTLYITILEAKDG--GQKK-VYEAKVWEKPMWDFKELQEFKHVGD--APA-----							
17. LVKAKTQVAGTMYLTIIEVKDG--EVKK-LYEAKVWEKPMWDFKELQEFKPV--EDASA-----							
18. ILKCESQVAGVNWRLVLCKDE--NNGEGNYETVWEKIENFRQLITFDHLLT-----							
19. VVNAKQVQVSGTIYILTLEVEDG--GKKK-VYEAKIWEKPMWDFKELQEFKQIVY--APTHSA-----							
20. VVKAKEQVAGTLLHLLTLEAIEA--GKKK-LYQAKVWVKPMWDFKELQEFKHAEDEETPSVTSSDLGVKQG							
21. VLDDKQVQVSGTMMYLIKIEATEG--GKKK-VYEAKVWVKPMWDFKELQEFKPVSD--S-----							
22. VVNKQAVVEGLKYCITLEAVDG--HGTK-VYEAELWLKLWENFRSLGFKLLGD--AH-----							
23. VVNAKQVQVSGMKYIFTVMGRTPCRK--GGVEKICSVHKDPQMAVPYKCTFEVWSRPWMSDIQMVKNQCES-							

Figura 2-11: alinhamento entre algumas seqüências de membros da família cistatina selecionados. Aqueles marcados (+) possuem estrutura tridimensional determinada.

4.5. FAMÍLIA TIMP

Os inibidores de metaloproteínas conhecidos como TIMP (de *Tissue Inhibitors of MetalloendoPeptidases*) são encontrados em tecidos e fluídos animais e desempenham um papel importante no controle de processos fisiológicos como artrite, gengivite e metástase tumoral (PAVLOFF *et al.*, 1992). Os TIMPs têm massa molecular entre 20–28kDa e contêm 12 meias-cistinas formando 6 pontes dissulfeto. Eles podem ser subdivididos em três grupos: TIMP–1 com cerca de 184 resíduos de aminoácidos, TIMP–2 com cerca de 194 resíduos e TIMP–3 com cerca de 188 resíduos. Apesar disso a estrutura global dos três

grupos é praticamente a mesma, sendo dividida em dois domínios: um na porção N-terminal no qual predominam as fitas- β s formando um barril- β e outro na porção C-terminal com predominância de α hélices embora quatro fitas- β s estejam presentes.

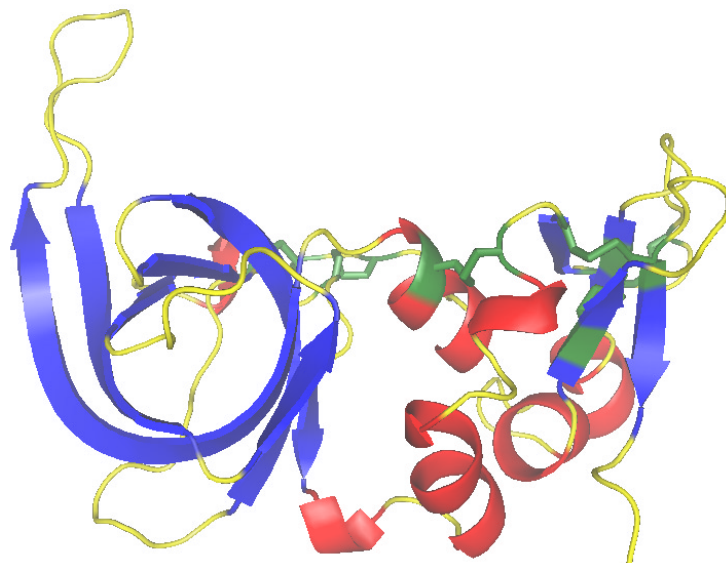


Figura 2-12: representação da estrutura tridimensional do inibidor da família TIMP-2 humano (código PDB 1BQQ. As pontes dissulfeto estão coloridas em verde).

	10	20	30	40	50	60		70	80	90	100	110	120	130	140	150
1. +HOMO SAPIENS	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
2. +BOS TAURUS	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
3. CRICETULUS LONGICAUDATUS	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
4. CANIS FAMILIARIS	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
5. EQUUS CABALLUS	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
6. MUS MUSCULUS	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
7. RATTUS NORVEGICUS	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
8. GALLUS GALLUS	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
9. XENOPUS LAEVIS	MSVCAL	TLILG	CFLFL	PLGDIS	KP----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
1.	GNDIYGN	PIKRIQ	YEIKQ	IKMF	KGP----	EKD--	IEFIYT	APSSAV	CG-	VSLDV	GGKKEY	LIAGKA	EGDG	GKMHTL	CD	FIVP
2.	GNDIYGN	PIKRIQ	YEIKQ	IKMF	KGP----	DQD--	IEFIYT	APSSAV	CG-	VSLDV	GGKKEY	LIAGKA	EGDG	GKMHTL	CD	FIVP
3.	GNDIYGN	PIKRIQ	YEIKQ	IKMF	KGP----	DKD--	IEFIYT	APSSAV	CG-	VSLDV	GGKKEY	LIAGKA	EGDG	GKMHTL	CD	FIVP
4.	GNDIYGN	PIKRIQ	YEIKQ	IKMF	KGP----	DKD--	IEFIYT	APSSAV	CG-	VSLDV	GGKKEY	LIAGKA	EGDG	GKMHTL	CD	FIVP
5.	G-----	PFGLT	VYTIK	QMKMY	RGF----	TKMPH	VQYI	ITEASE	SLCG-	LKLEVN	-KYQY	LLTGRVY	-DGKMY	TGLN	FVER	WDQL
6.	G-----	PFGLT	VYTIK	QMKMY	RGF----	SKMPH	VQYI	ITEASE	SLCG-	LKLEVN	-KYQY	LLTGRVY	-EGKMY	TGLN	FVER	WDHL
7.	G-----	PFGLT	VYTIK	QMKMY	RGF----	SKMPH	VQYI	ITEASE	SLCG-	LKLEVN	-KYQY	LLTGRVY	-EGKMY	TGLN	FVER	WDHL
8.	G-----	PFGLT	VYTIK	QMKMY	RGF----	QIMPH	VQYI	ITEASE	SLCG-	VKLEVN	-KYQY	LITGRVY	-EGKVY	TGLN	WVEK	WDRL
9.	G-----	PFGLT	VYTIK	QMKMY	RGF----	NKMPQ	VQYI	ITEASE	SLCG-	VKLEVN	-KYQY	LITGRVY	-EGKVY	TGLN	LIER	WEKL
1.	QKKS	LNHR	YQMG	CE-	CKIT	RC	PMIP	CYIS	SP	DB	CL	WMD	VVTE	KNING	HQAK	FFACI
2.	QKKS	LNHR	YQMG	CE-	CKIT	RC	PMIP	CYIS	SP	DB	CL	WMD	VVTE	KNING	HQAK	FFACI
3.	QKKS	LNHR	YQMG	CE-	CKIT	RC	PMIP	CYIS	SP	DB	CL	WMD	VVTE	KNING	HQAK	FFACI
4.	QKKS	LNHR	YQMG	CE-	CKIT	RC	PMIP	CYIS	SP	DB	CL	WMD	VVTE	KNING	HQAK	FFACI
5.	QRKGL	NYRY	HLGN	-CKIK	SCY	YLP	CFVT	SKNE	CL	WTD	MLSN	FGYP	PGYQ	SKHYA	CI	RQKGGY
6.	QRKGL	NYRY	HLGN	-CKIK	SCY	YLP	CFVT	SKNE	CL	WTD	MLSN	FGYP	PGYQ	SKHYA	CI	RQKGGY
7.	QRKGL	NYRY	HLGN	-CKIK	SCY	YLP	CFVT	SKNE	CL	WTD	MLSN	FGYP	PGYQ	SKHYA	CI	RQKGGY
8.	QRKGL	NYRY	HLGN	-CKIK	SCY	YLP	CFVT	SKNE	CL	WTD	MLSN	FGYP	PGYQ	SKHYA	CI	RQKGGY
9.	QRKGL	NYRY	PLGT	-CKIK	PCY	YLP	CFIT	SKNE	CL	WTD	MLSN	FGYP	PGYQ	SKNYA	CI	KQKEGY

Figura 2-13: alinhamento entre algumas seqüências de membros da família TIMP selecionados. Aqueles marcados (+) possuem estrutura tridimensional determinada. As meias cistinas conservadas estão destacadas em preto.

4.6. FAMÍLIA METALOCARBOXIPEPTIDASE

Essa família é formada pelos inibidores de metalocarboxipeptidases que são na sua maioria de origem vegetal, embora um dos membros, mostrado no alinhamento da figura 2-14, seja o inibidor extraído da sanguessuga *Hirudo medicinalis*. O primeiro membro da família foi o inibidor extraído de batatas *Solanum tuberosum*.

Eles são bem menores que a família anterior de inibidores de metaloproteinases, sendo compostos por cerca de 40 resíduos de aminoácidos e contêm três pontes dissulfeto que mantêm a rigidez e estabilidade de uma estrutura que possui apenas uma folha- β formada por duas fitas- β s.

		10	20	30	40
1. +HIRUDO MEDICINALIS	-----	MFLLVFLCCLHLV	ISSHTPDES	---	FLCYQPDQ
2. SOLANUM TUBEROSUM	-----	-----	-----	-----	QQHADP--
3. LYCOPERSICON ESCULENTUM	-----	MAQKFTILFTIL-	LVVIAAQDVMAQ	DATLTKLFQ	QYDP--
4. SOLANUM BREVIDENS	MAAKNSEM	KFAIFFVDL-	LTTTSLMALRDM	PPQETLLKMK	LFPTN
5. HYOSCYAMUS NIGER	MASFQVFL	FIALFLAISVNF	SQVMALRDL	PDDIAAMKGL	LPLG

	50	60	70	80	90	100
1.	---VCCFIC	RGAPLPSEGE	ENPHPTAPWC	REGAVEWVPY	STG-----	QCRTTCIPYVE---
2.	---ICNKPC	K-----	THDDCSGAWFC	QACWN-----	SARTCG-----	PYVG-----
3.	---VCHKPCS	-----	TQDDCSGGTFC	QACWR-----	FAGTCG-----	PYVGRAMAIGV---
4.	ILGTCTNY	YCT-----	SNSDQIGITFC	QWCKV-----	KTNEYD-----	GSASRTCSVLP---
5.	DIVTCTFK	YCN-----	VESDQSDGWL	CYNQVPSAFEG	WRSQCDKLT	VTGEGYFGSILRAKYNKI

Figura 2-14: alinhamento entre algumas seqüências de membros da família metalocarboxipeptidase selecionados. Aqueles marcados (+) possuem estrutura tridimensional determinada. As meias cistinas conservadas estão destacadas em preto.

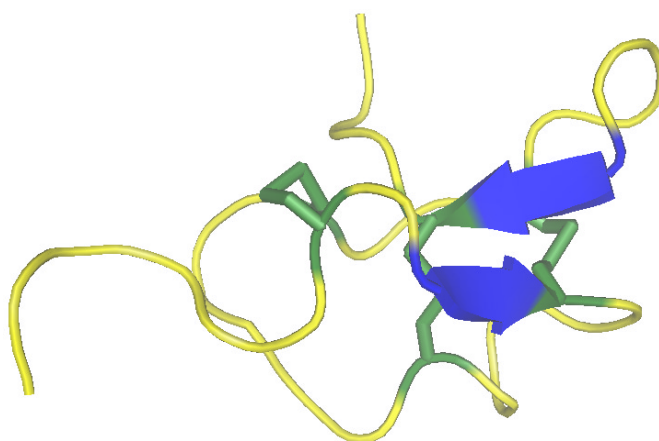


Figura 2-15: representação da estrutura tridimensional do inibidor de metalocarboxipeptidase de batata *Solanum tuberosum* (código PDB 4CPA). As pontes dissulfeto estão em verde.

CAPÍTULO 3

OS INIBIDORES DE SERINOPROTEINASES

1. INTRODUÇÃO

O número de inibidores de serinoproteinases excede de longe o número de inibidores das outras três classes descritos. Essa abundância relativa não reflete a distribuição de inibidores na natureza mas apenas o interesse e a conveniência dos pesquisadores.

Em contraste com os outros inibidores, o mecanismo de interação dos inibidores de serinoproteinases com suas enzimas cognatas é consideravelmente bem detalhado. Dados de seqüenciamento e cristalografia de raios-X mostram que esses inibidores não são todos homólogos consistindo de muitas (pelo menos 16) famílias. Surpreendentemente, entretanto, a grande maioria dos inibidores de serinoproteinases interage seguindo o “mecanismo padrão” que será descrito. Esta situação é análoga à existência de pelo menos duas famílias de serinoproteinases não homólogas, a família da subtilisina e a da quimotripsina, as quais não dividem uma estrutura tridimensional comum mas hidrolisam seus substratos de acordo com o mesmo mecanismo. Como já foi descrito, existem apenas

dois tipos de inibidores de serinoproteinases que não seguem o mecanismo padrão, a família dos bem pequenos e muito estáveis inibidores de trombina de sanguessugas chamados de hirudinas e a família das serpinas muito maiores e relativamente instáveis inibidores encontrados no plasma de mamíferos.

2. O INÍCIO DOS ESTUDOS SOBRE INIBIDORES

Talvez o interesse pelos inibidores de proteinases tenha tomado impulso em 1936, quando KUNITZ e NORTHROP isolaram e cristalizaram o inibidor de tripsina bovina extraído do pâncreas bovino, que recebeu o nome de BPTI – de *Bovine Pancreatic Trypsin Inhibitor*. O que logo se seguiu, em 1945, pela cristalização do inibidor de tripsina extraído de soja, que recebeu o nome de STI – de *Soybean Trypsin Inhibitor*. Os experimentos de Kunitz deixaram uma série de definições e padronizações na pesquisa sobre inibidores.

Entretanto, surgiram dois grandes problemas do legado deixado por Kunitz. Um foi a crença de que a maioria dos inibidores de proteinases e, assim, a maioria dos inibidores é inibidores de tripsina. Muitos pesquisadores testaram numerosos materiais biológicos quanto à inibição de tripsina e então, previsivelmente, foram descobertos mais e mais inibidores de tripsina. Isso levou à crença ilógica e errônea de que inibidores de tripsina são esmagadoramente dominantes na natureza. O outro problema deixado pelo legado de Kunitz foi que tanto a tripsina quanto seus muitos inibidores eram verdadeiras ‘caixas pretas’. O que foi piorando pois, enquanto os pesquisadores ficavam intrigados por como a tripsina se combina com seus inibidores, os estudos de modificações químicas e dicroísmo circular davam resultados complexos (LASKOWSKY Jr *et al.*, 2000).

Em 1954, no laboratório de NORD foi feita a sugestão correta de que inibidores são análogos aos substratos das proteinases e que os complexos enzima–inibidor formados seriam especialmente estáveis (SRIRAM *et al.*, 1954). Entretanto, os autores não levaram essa idéia adiante para encontrar o sítio ativo ou para conseguir alguma evidência experimental forte. O desenvolvimento da tecnologia de seqüenciamento de proteínas levou KASSEL e LASKOWISKY a seqüenciar BPTI. Logo em seguida foi feito o seqüenciamento do ‘inativador de calicreína bovina’. E então, após alguns pequenos desvios na seqüência do inativador de calicreína serem corrigidos, percebeu-se que os dois eram, na verdade, a mesma proteína (LASKOWSKY Jr. *et al.*, 2000).

A estrutura tridimensional de BPTI foi determinada pouco tempo depois da cristalografia de proteínas ser possível. Apesar de se conhecer a sequência e a estrutura tridimensional do inibidor livre, o mecanismo de combinação ainda se mostrava difícil de deduzir. Isto mudou completamente quando as estruturas tridimensionais dos complexos de tripsina–inibidor foram determinadas.

3. DESVENDANDO O MECANISMO DE AÇÃO DOS INIBIDORES

Muitos estudos de hidrólise foram feitos utilizando-se inibidores e proteinases. Medindo-se a liberação de íons hidrogênio durante a combinação de STI e tripsina, foi observado que uma grande e rápida liberação de prótons é seguida por um pequeno e lento consumo de prótons. A explicação apresentada por FINKENSTADT e LASKOWSKY (1965) foi que o complexo enzima–inibidor poderia dissociar tanto para enzima e inibidor virgem quanto para enzima e inibidor modificado. No STI modificado chamado STI* a ligação peptídica do sítio ativo Arg63–Ile64 (P1–P1') é especificamente hidrolisada. Os dois fragmentos resultantes são mantidos unidos por uma ponte dissulfeto e por uma ampla rede interna de pontes de hidrogênio que se mantém intacta mesmo após a clivagem do inibidor.

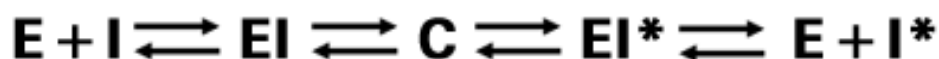


Figura 3-1: reação que ilustra o mecanismo geral seguido pelos inibidores canônicos.

A presença de Arg na posição P1 é consistente com a especificidade da tripsina, mas observe que, dentre os muitos resíduos de Lys e Arg em STI, apenas um, Arg63, é consistentemente selecionado. Em muitos inibidores de tripsina, STI por exemplo, o resíduo na posição P1 é uma Arg, enquanto em muitos outros, por exemplo BPTI, é uma Lys. Isso explica os resultados aparentemente anômalos previamente obtidos por modificação química que tanto intrigaram os pesquisadores: a acetilação dos inibidores cujo resíduo é Lys extinguiu completamente sua atividade inibitória enquanto a acetilação dos inibidores cujo resíduo é Arg não tinha qualquer efeito sobre sua atividade. Isso acontece porque o grupo NH₂ da Lys é bloqueado por acetilação o que impede a interação

iônica com o Asp102 da tripsina necessária para a associação enzima–inibidor e, no caso da Arg, apenas um dos grupos NH_2 é bloqueado, permanecendo o outro livre para a interação. Isso foi endossado por experimentos de mutação enzimática, nos quais o resíduo Lys na posição P1 do inibidor era substituído por Arg, fazendo com que a atividade sobre tripsina não fosse afetada. Essa talvez tenha sido a primeira mutagênese sítio específica realizada (LASKOWSKY Jr *et al.*, 2000).

Como se viu, uma grande quantidade de informações têm sido acumulada nas últimas décadas sobre inibidores de serinoproteinases e seu mecanismo de ação. A fim de se compreender o mecanismo pelo qual os inibidores influenciam a atividade proteolítica, é preciso relembrar o mecanismo de ação das serinoproteinases. Como já foi exposto, o mecanismo padrão pelo qual as serinoproteinases clivam peptídeos é acompanhado em três passos. Primeiro a enzima se liga ao substrato para formar o complexo de Michaelis no qual o nucleófilo (a Ser195) ataca a carbonila da ligação “scissile” do peptídeo para formar um complexo de transição conhecido como intermediário tetraédrico. Segundo, o intermediário tetraédrico se compõe no intermediário acil–enzima com o grupo abandonador não covalentemente ligado. Então o grupo abandonador, o novo N–terminal do peptídeo substrato clivado, é deslocado da enzima e substituído por uma molécula de água. No terceiro e último passo, o intermediário acil–enzima é deacilado pelo que é essencialmente o reverso do primeiro passo, com a molécula de água sendo o nucleófilo atacante e a serina catalítica o grupo abandonador, seguido pela saída do novo C–terminal do peptídeo clivado e regenerando a enzima ativa (VOET & VOET, 1995).

Um inibidor padrão clássico é hidrolisado cerca de 10^7 vezes mais lentamente que um bom substrato sendo que o passo determinante é a deacilação do intermediário acil–enzima. Primeiro o equilíbrio entre o complexo de Michaelis e o intermediário acil–enzima é rapidamente estabelecido. A ligação do inibidor clivado no sítio ativo da enzima é forte e orientada, prevenindo a hidrólise do intermediário acil–enzima e favorecendo a reação inversa com a religação do grupo abandonador o que diminui drasticamente a velocidade da reação (RADISKY & KOSHLAND, 2002).

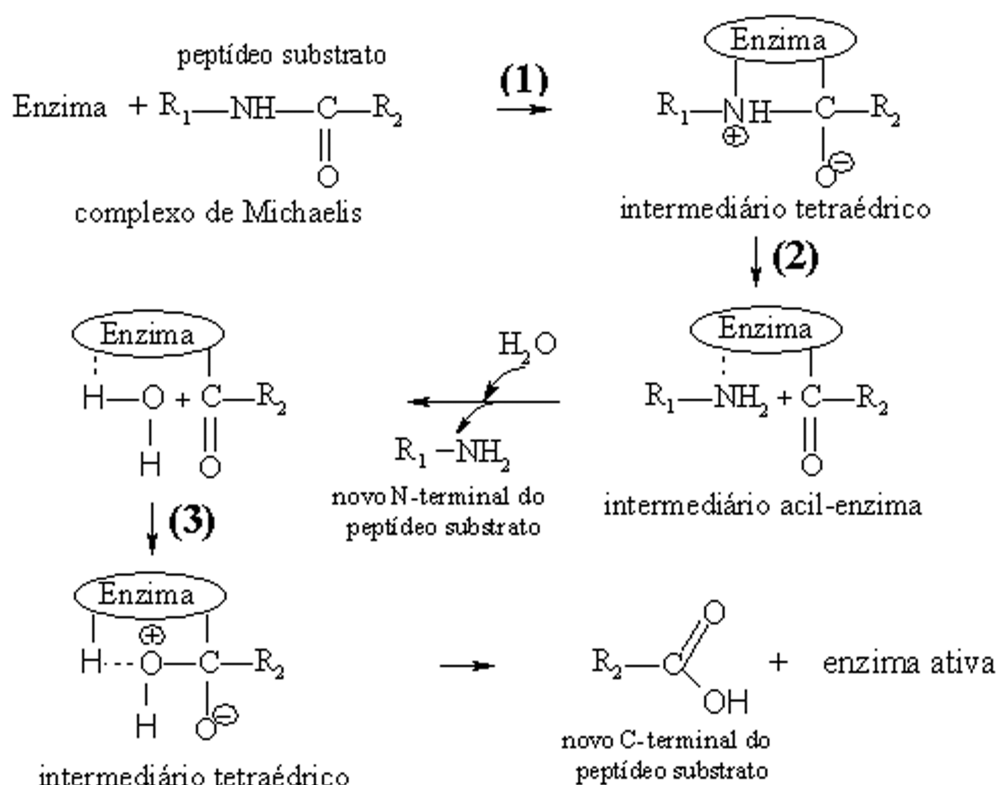


Figura 3-2: Representação esquemática do mecanismo geral de ação das serinoproteinases.

A conformação canônica do loop reativo é muito importante para um posicionamento adequado e uma adaptação efetiva ao sítio ativo da enzima proteolítica o que resulta na formação do intermediário acil-enzima. O resíduo P1 desempenha um papel chave durante o processo, definindo a especificidade do inibidor. Por exemplo, a presença de uma arginina ou uma lisina na posição P1 permite uma interação iônica com o Asp189 no sítio de ligação da tripsina. A geometria do grupo carbonila do resíduo na posição P1 define a orientação adequada para o ataque nucleofílico da serina catalítica, a qual primeiro forma o acil-enzima e então faz o ataque nucleofílico ao grupo abandonador, favorecendo a religação. Além disso, o posicionamento e a orientação do novo N-terminal do inibidor clivado são também cruciais para um ataque nucleofílico efetivo. Isso é providenciado por uma extensa rede de interações intra e intermoleculares entre inibidor e enzima.

4. O ENIGMA DA PROTEÓLISE LIMITADA

A determinação das estruturas do inibidor e do seu complexo com a proteinase que ele inibe é crucial para compreender a sua performance. Por exemplo, porque os inibidores

não são completamente clivados na presença da proteinase como a maioria das proteínas? Tripsina, por exemplo, deveria clivar cada uma das ligações Lisina–X e Arginina–X (com a exceção de Prolina em X) e, nesse caso, cerca de 5–10% das ligações peptídicas em uma proteína típica seria susceptível ao ataque proteolítico. Entretanto, os inibidores de proteinases fornecem um paradigma uma vez que apenas a ligação peptídica ‘scissile’ P1-P1' no loop reativo é proteoliticamente atacada resultando na formação do complexo enzima–inibidor no qual o loop reativo do inibidor está ligado ao sítio ativo da enzima numa conformação de “substrato perfeito”, enquanto, baseado em sua seqüência, deveria ser esperado que eles fossem rapidamente proteolisados em muitos outros sítios de clivagem (HUBBARD *et al.*, 1991).

É uma combinação de fatores estruturais e termodinâmicos que ajuda a entender o quebra-cabeça biológico da proteólise limitada dos inibidores. Sítios proteolíticos são encontrados em regiões expostas, protuberantes e flexíveis de uma proteína e são tipicamente loops ou turns, raramente em ou próximos a hélices mas, aparentemente nunca em estruturas betas. Eles se expõem na superfície da proteína e são encontrados em regiões onde o empacotamento local não restringe o desenovelamento local que é necessário para o reconhecimento e a clivagem pelas proteinases. A chave determinante para uma região ser considerada um sítio proteolítico é a habilidade para desenovelar localmente e se adaptar ao sítio ativo da enzima. Existem três importantes fatores para um sítio proteolítico: exposição, flexibilidade e interações locais. A exposição é caracterizada pela acessibilidade e embora não seja um requerimento absoluto, um sítio proteolítico estará mais provavelmente situado numa região próxima à superfície da proteína na qual o desenovelamento local é mais facilmente acompanhado. A flexibilidade é crítica para o desenovelamento local e adaptação ao sítio ativo da enzima. As interações locais dependem da estrutura secundária e das pontes de hidrogênio. Um bom candidato para desenovelamento local e adaptação não deve ser enrijecido por interações como pontes dissulfeto ou pontes de hidrogênio, comuns em estruturas secundárias regulares, particularmente folhas- β s (NOVOTNY & BRUCCOLERI, 1987; HUBBARD, 1994, 1998).

Essas observações ajudam a explicar alguns aspectos dos inibidores de proteinases. Eles são moléculas pequenas, os menores são os inibidores de serinoproteinases da família abóbora com cerca de 3kDa e os maiores são os da família Grasshopper cujo domínio inibitório tem 44kDa. A grande maioria dos inibidores possui pontes dissulfeto, alguns em

grande quantidade como os da família Bowman–Birk com 7 pontes dissulfeto numa molécula de cerca de 8kDa. Além disso, os inibidores de proteinases possuem uma estrutura secundária bem definida, sendo que a maioria possui grande quantidade de folhas- β s e as regiões que formam os loops são normalmente enterrados na estrutura em regiões de difícil acesso. Apenas o loop reativo é mais exposto e flexível. Essas características estruturais fazem dos inibidores moléculas compactas, rígidas e estáveis o que permite que a sua integridade seja mantida mesmo após a clivagem pela proteinase.

Por exemplo, no caso do inibidor de quimotripsina de cevada *Hordeum vulgare* (CI2) foi demonstrado que a rede de pontes de hidrogênio permanece intacta no inibidor clivado no loop ativo e estabiliza as partes resultantes do inibidor mantendo-as unidas (SHAW et al., 1995). Além disso, quando o inibidor e enzima estão ligados, uma superfície de contatos é estabelecida entre eles envolvendo resíduos localizados em diferentes partes das moléculas com a maioria dos contatos relacionando os resíduos dos sítios reativos. Esses numerosos contatos mantêm o inibidor clivado numa orientação ótima para o ataque nucleofílico do novo N-terminal sobre o novo C-terminal favorecendo a religação. A posição e orientação das partes clivadas também impedem estericamente que a molécula de água se aproxime da histidina para a ativação nucleofílica necessária.

Em adição aos aspectos estruturais, a formação do complexo enzima–inibidor também é um evento termodinamicamente favorável. Por exemplo, o ganho em energia livre de Gibbs para BPTI tem um valor típico entre -8kcal/mol (para o complexo BPTI–Tripsinogênio) e -18kcal/mol (para o complexo BPTI–Tripsina) (KRYSTEK et al., 1993). Esse ganho energético é explicado em parte pela formação de uma folha- β no complexo inibidor-enzima sendo uma das fitas- β é fornecida pelo loop reativo do inibidor e a outra pela enzima.

É a contribuição de cada um desses fatores que determina a eficiência de inibição do inibidor de proteinases.

5. A CONFORMAÇÃO CANÔNICA PARA O LOOP REATIVO DOS INIBIDORES

Estudos estruturais de diversos inibidores livres e de numerosos complexos enzima–inibidor confirmaram e estenderam as primeiras observações de que todos os inibidores têm um grande loop exposto ao redor do resíduo P1. O resíduo P1 é hiper-exposto nos

inibidores livres. Esse loop atua como uma das duas fitas de uma folha- β distorcida que é formada nos complexos com enzimas da família da quimotripsina. A segunda fita é fornecida pela enzima. Na interação com enzimas da família da subtilisina, a enzima contribui com duas fitas e a folha- β distorcida resultante é composta de três fitas. Além dos resíduos próximos à ligação peptídica do sítio ativo que compõem o loop reativo, outros resíduos de diferentes partes do inibidor também fazem contatos com a enzima tanto com a cadeia principal quanto com as cadeias laterais. Embora essas interações não sejam tão conservadas em todos os inibidores quanto aquelas feitas pelos resíduos do loop reativo, elas são compartilhadas por inibidores dentro das diferentes famílias.

Os resíduos de P_4 a P_3' de todos os inibidores exibem os mesmos ângulos de RAMACHANDRAM –veja alguns exemplos na Tabela 2-3 (LASKOWSKY Jr. *et al.*, 2000). Os ângulos apresentados para P_1 são aproximadamente aqueles de uma hélice 3_{10} e os apresentados pelos demais resíduos são aproximadamente aqueles das fitas- β s. É surpreendente que os ângulos de Ramachandram apresentados pelo loop reativo dos inibidores da maioria das 16 famílias sejam tão próximos a despeito do fato de que cada uma dessas 16 famílias tem sua própria estrutura tridimensional global. Essas diferenças na estrutura global mas, as surpreendentes similaridades na estrutura local onde realmente elas são importantes para a atividade inibitória ajudam a entender as dificuldades dos estudos estruturais prévios, feitos com dicroísmo circular por exemplo, em explicar o comportamento dos vários inibidores de tripsina.

Os ângulos de Ramachandran não são apenas comuns a muitos inibidores livres, eles permanecem os mesmos nos complexos. Observe na Tabela 2-3 que os inibidores formam complexos com muitas enzimas diferentes e os ângulos de Ramachandran do loop permanecem os mesmos. O mesmo pode ser visto na figura 3-3, na qual foram superpostos os loops reativos de 23 diferentes inibidores complexados com proteinases (RADISKY & KOSHLAND, 2002). Essa é a maior manifestação de que o mecanismo de associação de enzimas e inibidores é predominantemente do tipo ‘chave-fechadura’ de Fischer. Os inibidores que exibem os ângulos de Ramachandram esperados tanto em sua forma livre quanto em seus complexos foram chamados *canônicos* por BODE e HUBER (1992).

Tabela 2-3: Os ângulos ϕ e ψ da cadeia principal dos resíduos do loop reativo dos inibidores.

Inibidor	Kazal				BPTI				Batata I		
PDB	2ovo	1cho	3sgb	1ppf	1bpi	2pte	2kai	1cbw	1mit	1acb	1cse
Enzima	livre	quimo	SGPB	HLE	livre	tripsina	calicreína	quimo	livre	quimo	subtilisina
P4 ϕ	-149	-129	-158	-127	81	79	77	99	-71	-81	-71
P4 ψ	163	136	157	148	-178	175	179	172	153	156	140
P3 ϕ	-131	-131	-126	-119	-93	-77	-64	-84	-132	-144	-139
P3 ψ	155	150	147	143	-3	-29	-32	-16	163	166	168
P2 ϕ	-87	-68	-69	-73	-87	-70	-91	-69	-74	-79	-62
P2 ψ	174	160	162	156	162	156	168	154	150	153	143
P1 ϕ	-96	-107	-119	-101	-115	-117	-112	-103	-82	-97	-115
P1 ψ	9	32	45	31	19	39	46	21	12	40	45
P1' ϕ	-58	-74	-84	-84	-64	-88	-88	-81	-99	-90	-97
P1' ψ	139	159	155	155	162	164	167	-169	159	166	169
P2' ϕ	-99	-113	-100	-102	-121	-113	-115	-150	-111	-131	-117
P2' ψ	93	107	114	107	90	79	84	79	128	121	110
P3' ϕ	-130	-142	-148	-138	-106	-106	-110	-104	-129	-129	-121
P3' ψ	69	76	92	82	118	122	108	134	123	129	112

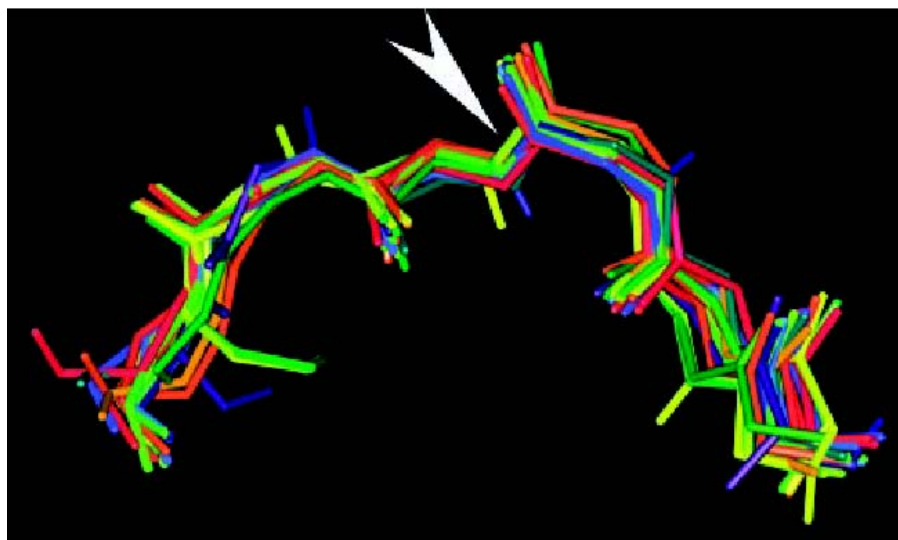


Figura 3-3: Superposição dos átomos das cadeias principais dos 6 resíduos do loop reativo de 23 diferentes inibidores de proteinases em complexo com proteinases. As estruturas são de 12 diferentes famílias de inibidores não relacionadas em seqüência e em arranjo, em complexo com 9 serinoproteinases incluindo membros tanto da família da quimotripsina quanto da subtilisina. A seta branca indica o sítio de clivagem entre P1–P1'.

Apesar das palavras *mecanismo padrão* e *canônico* não serem sinônimos, todos os inibidores com *mecanismo padrão* para os quais as estruturas tridimensionais foram determinadas são *canônicos*. Mas o recíproco não é verdadeiro, muitas seqüências canônicas que nada tem a ver com inibidores têm sido encontradas em proteínas (APOSTOLUK & OTLEWSKI, 1998).

Em um artigo recente (JACKSON, 1999) foram comparadas as importâncias relativas das interações entre cadeia-principal–cadeia-principal com as interações entre cadeia-lateral–cadeia-lateral em várias interações proteína–proteína. Apontou-se que as interações entre as cadeias-principais são mais importantes para as interações entre proteinase–inibidor do que para as interações entre antígeno–anticorpo. Essa diferença se deve em parte ao longo tempo de evolução usado para otimizar a associação proteinase–inibidor. A grande importância das interações entre cadeia-principal–cadeia-principal dentre os inibidores canônicos com mecanismo padrão pode ser responsável pelas regras peculiares de seu comportamento e da sua estrita aderência a elas, o que não é muito comum dentre as proteínas.

6. AS DIFERENTES FAMÍLIAS DE INIBIDORES DE SERINOPROTEINASES

Como já foi citado, existem pelo menos 16 diferentes famílias de inibidores de serinoproteinases reconhecidas até agora. É muito provável que muitas novas famílias sejam descritas no futuro mas também muitos inibidores novos que são descobertos agora – o que tem aumentado exponencialmente devido aos estudos de seqüenciamento de genomas – sejam enquadrados nessas famílias já estabelecidas. As principais famílias de inibidores de serinoproteinases são listadas na tabela 2-4. Tentou-se dividir essas famílias em inibidores de origem animal, de plantas e de microorganismos e apesar de se ter conseguido bastante êxito, essa divisão não é perfeita (LASKOWSKY Jr *et al.*, 2000). Uma exceção notável é a *eglin c*, a qual é um inibidor de origem animal extraído de sanguessuga *Hirudo medicinalis* e é um membro da família Batata I. Existem muitas outras exceções.

Para cada uma dessas famílias é apresentado a seguir um pequeno histórico bem como uma descrição das características e peculiaridades de seus membros. Devido à grande abundância de seqüências disponíveis para algumas das famílias, são escolhidas apenas algumas dentre aquelas pertencentes a cada família com as quais é feito um alinhamento a

fim de se mostrar os resíduos conservados, particularmente as meias-cistinas e o loop reativo. Também, quando possível, é apresentada a estrutura 3D de um dos membros que caracterize a família.

Como já foi advertido, algumas das famílias descritas são famílias de proteínas com membros que não possuem atividade inibitória e muitos dos quais apresentam outras funções.

Tabela 2-4: As famílias de inibidores de serinoproteínases.

FAMÍLIA	MONÔMERO		ESTRUTURA 3D (exemplo do PDB)	
	MM _r (kDa)	½ CYS	LIVRE	COMPLEXO
BPTI Kunitz ^a	6	6	1BPI	2PTC
STI Kunitz ^b	21–22	4	1AVU	1AVW
BBI ^b	8–9	14	1PI2	1SMF
Batata I ^b	8–9	0–2	1MIT	1ACB
Batata II ^b	6	8	1TIH	4SGB
Abóbora ^b	3	6	2CTI	1PPE
Cereal ^b	12–13	10	1B1U	NÃO
Ragi I-2 ^b	12–13	7–8	1AFH	1MZM
Thaumatina ^b	22–23	16	1DU5	–
Kazal ^a	6	6	2OVO	1CHO
SSI ^c	12–14	2-4	2SSI	2SIC
Antistasina ^a	8–10	20	1SKZ	1HIA
Grasshopper ^a	44	6	1PMC	–
Chelonianina ^a	6	6	2REL	1FLE
Ascaris ^a	6	10	1ATA	1EAI
Ecotina ^c	12–14	2	1ECY	1AZZ

Inibidores de origem ^aanimal, de ^bplantas e de ^cmicroorganismos.

6.1. FAMÍLIA BPTI KUNITZ

A família BPTI Kunitz é composta de pequenas proteínas com massa molecular em torno de 6kDa (cerca de 55–60 aminoácidos) com representantes em espécies de vertebrados e invertebrados, incluindo humanos, bovinos, ratos, camundongos, peixes,

cavalos, sapos, vermes e bicho-da-seda. Essa ubiquidade sugere o papel fundamental dessa família de proteínas nos animais.

Apesar dos membros da família BPTI Kunitz compartilharem um alto grau de similaridade seqüencial e estrutural, o que faz com que todos apresentem o mesmo arranjo global característico, eles exibem uma faixa enorme de funcionalidades e especificidades. Os membros dessa família de proteínas são encontrados como cadeia simples independente como é o caso do próprio BPTI, e também como domínio(s) numa estrutura global maior na qual o domínio(s) BPTI é covalentemente ligado a outros da proteína como, por exemplo, no colágeno e na β -bungarotoxina (GREENSPAN, 1993; KOHFELDT et al., 1996; KONDO et al., 1982) e em glicoproteínas (BRINKMANN, 1997). O arranjo do tipo BPTI Kunitz é altamente adaptável e tem sido encontrado em proteínas que desempenham muitas funções diferentes na natureza, desde a inibição de proteinases até poderosas neurotoxinas (SKARZYNSKI, 1992; MARSHALL & HARAVEY, 1992).

Dentre as proteínas que possuem domínios BPTI Kunitz estão as chamadas bikuninas, que compreendem os inter- α inibidores de tripsina ($I\alpha I$) que possuem o domínio BPTI Kunitz que é a parte inibitória do $I\alpha I$. As bikuninas possuem dois domínios tipo BPTI Kunitz: o domínio N-terminal (domínio 1) tem um sítio de ligação neutro e específico para elastases enquanto o domínio C-terminal (domínio 2) tem um sítio de ligação básico sendo específico para as serinoproteinases tripsina e plasmina. Um impedimento estérico que um domínio impõe ao outro previne que ambos se liguem simultaneamente a enzimas e também restringe o tamanho do alvo que será ligado. Por essa versatilidade na inibição, as bikuninas são atualmente prescritas no Japão para o tratamento de pancreatite aguda e choque hemorrágico. Sabe-se também que as bikuninas estão envolvidas no controle do crescimento celular e os $I\alpha I$ participam em processos na matriz extracelular como morfogênese e diferenciação de tecidos (LINDQVIST & AKERSTRÖM, 1996; ITOH et al., 1994).

Tabela 2-5: Família Kunitz (BPTI)

INIBIDOR	ATIVIDADE INIBITÓRIA PRINCIPAL	OUTRAS FUNÇÕES CONHECIDAS
BPTI	Forte inibidor de tripsina e caliceína. Também inibe muitas outras serinoproteínas	BPTI e seus fragmentos 1-15, 18-39, e 40-58 possuem atividade bactericida contra muitas bactérias gram positivas e gram negativas. Inibe a cisteinoproteínas de adenovírus com uma constante de ligação de $5 \times 10^5 \text{ M}^{-1}$ e um sítio de inibição diferente do que liga tripsina. Inibe o canal $\text{K}^+/\text{Ca}^{++}$ ativado. A subunidade α do canal K_{Ca} contém um domínio inibidor de proteinase com sítio de ligação conservado.
APPI Isoformas da proteína precursora de amiloide (APP) que contém um domínio tipo BPTI-Kunitz.	APPI é um forte inibidor de tripsina e inibe muitas outras serinoproteínas.	Acredita-se que a função das APPs como receptores nucleares está envolvida na via de sinalização intracelular através da proteína de ligação (Go) de GTP. Na doença de Alzheimer, um fragmento de APP, a proteína β amiloide é depositada em vários tecidos cerebrais.
Calcicludina (encontrada no veneno de algumas cobras)	Atividade inibitória desconhecida	Bloqueiam especificamente a maioria dos canais de Ca^{++} .
Dendrotoxinas (encontradas no veneno de muitas cobras) Exemplos: toxina I, toxina K, toxina E.	Inibe tripsina e caliceína.	Bloqueiam alguns canais neuronais de K^+ e aumentam a liberação de neurotransmissores.
Cadeia B das bungarotoxinas β_1 e β_2 (encontradas no veneno de <i>Bungrus multicinctus</i>)	Atividade inibitória desconhecida	A bungarotoxina tem duas cadeias (A e B) que bloqueiam a transmissão neuromuscular, sendo a cadeia A uma fosfolipase.
Bikunina	Inibe tripsina, quimotripsina, PPE e HLE.	Bikunina é co-transportada com α_1 microglobulina a qual pertence à superfamília das lipocalinas. Esse precursor passa por um processamento pós-translacional para produzir α_1 microglobulina e bikunina. Bikunina então se associa covalentemente com uma ou duas cadeias da família multioxidase formando $\text{I}\alpha\text{I}$.
Os colágenos tipo VII e tipo VI (cadeia 3) contêm um domínio BPTIKunitz nas porções C-terminais.	Nenhuma atividade inibitória relatada. Entretanto, uma mutação simples em P1' (DA) confere uma fraca inibição de tripsina e uma tripla mutação (P3TP; P1'DA; P2'FR) confere uma forte inibição contra tripsina.	O tipo VII é um componente das fibrilas de ancoramento enquanto o tipo VI atua como uma proteína de ligação celular.
BPTI/Bikunina/Tripstatina	As mesmas descritas para BPTI e Bikunina	Inibem a ligação da HIV glicoproteína gp120 com a triptase TL.

BPTI (protease inhibitor)	
Bos taurus	CLEPPYTGECK--ARIIRYFYNAKAGLCQTFVYGCRAKRNNFKSAEDCMRTC
Mus musculus	CLASYKVGRCR--GSFPRWYDPEQICKSFTFCGLGNKNNYLREEECMLAC
Anemonia sulcata	CELPKVVGECR--ARFPRYYNSSSKRCEKFIYGC GGNANNFHTLECEKVC
Anthopleura aff. Xanthogrammica	CLLPKKVGEPCR--AAVPRFYNSDSGKCEGFYGC GGNANNFHTLECEKVC
Bombyx mori	CLLPKIKTGECK--GSFPRYAYDSEDKQVEFIYGC GGNANNFHTLECEKVC
Helix pomatia	CNLPAGTGECK--ASFRRYYNSKSGCQPFYGC GGNANNFHTLECEKVC
Sarcophaga bullata	CLQPKVEGECRK-SDFVF--FYNADTKACEEFYGC GGNANNFHTLECEKVC
Boophilus microplus	CVPTADPGECCKG-FMPMW--WYNIFTSCCEEFYGC GGNANNFHTLECEKVC
Radianthus macrodactylus	CLEPKVVGECR--AYFPRFYFDSGTGKCTPFYGC GGNANNFHTLECEKVC
Stoichactis helianthus	CLEPKRVGECR--GYFPRFYFDSGTGKCTPFYGC GGNANNFHTLECEKVC
Melithaea caledonica	CSLKKVVGECR--GAFRRYYFDSVSGKCEEFYGC GGNANNFHTLECEKVC
APPI (Alzheimer's Disease Amyloid A4 Protein)	
Homo sapiens	CSEQAETGEPCRA-MISRW--YFDVTEGKCAPFFYGC GGNANNFHTLECEKVC
Mus musculus	CSEQAETGEPCRA-MISRW--YFDVTEGKCAPFFYGC GGNANNFHTLECEKVC
Cavia porcellus	CSEQAETGEPCRS-MISRW--YFDVTEGKCAPFFYGC GGNANNFHTLECEKVC
Macaca mulatta	CSEQAETGEPCRA-MISRW--YFDVTEGKCAPFFYGC GGNANNFHTLECEKVC
Macaca fascicularis	CSEQAETGEPCRA-MISRW--YFDVTEGKCAPFFYGC GGNANNFHTLECEKVC
Sus scrofa	CSEQAETGEPCRA-MISRW--YFDVTEGKCAPFFYGC GGNANNFHTLECEKVC
Rattus norvegicus	CSEQAETGEPCRA-MISRW--YFDVTEGKCAPFFYGC GGNANNFHTLECEKVC
Saimiri sciureus	CSEQAETGEPCRA-MISRW--YFDVTEGKCAPFFYGC GGNANNFHTLECEKVC
Calciclutina	
Dendroaspis angusticeps	CKEPVRIGSCK--KQFSSFYFKWAKKCLPFLFC GGNANRFQTIGCEKVC
Kaliclutine	
Anemonia sulcata	CLLPMDVGRCR--ASHPRYYNSSSKRCEKFIYGC GGNANNFHTLECEKVC
Dendrotoxin	
Dendroaspis polylepis polylepis	CKLPAEPGECK--ASIPAFYYNWAACKCLFHYGC GGNANRFSTIECEKVC
Dendroaspis angusticeps	CKLPVRYGECR--KKIPSFYKWKAKCLPFDYGC GGNANRFSTIECEKVC
Araneus ventricosus	CLLPKVTGECK--ASLTRYYYDKDTKQVEFIYGC GGNANRFSTIECEKVC
Inibidores encontrados no veneno de cobra	
Naja naja	CELPAAGLCK--AHKPAFYNNKDSHRCQKFIYGC GGNANRFSTIECEKVC
Ophiophagus hannah	CELPPPEGLCNAR-KTFF--YYSLSHSAQKFIYGC GGNANRFSTIECEKVC
Naja nivea	CELPAETGLCK--ARIRSFHYNRAAQCLEFIYGC GGNANRFSTIECEKVC
Hemachatus haemachatus	CELPAETGLCK--AYIRSFHYNLAQAQCLQFIYGC GGNANRFSTIECEKVC
Vipera ammodytes ammodytes	CYLPADPGRCCL--AYMPRFYYNPASNKCEKFIYGC GGNANRFSTIECEKVC
Eristocophis macmahoni	CYLPDDPGVCK--AHIPRFYYNPASNKCKNFIYGC GGNANRFSTIECEKVC
Daboia russelli siamensis	CNLAPESGRCRG-HLRR--YYNLESNKCKVFIYGC GGNANRFSTIECEKVC
Bungarotoxin	
Bungarus multicinctus	CDKPPDKGNCG--PVRRAFYYDTRLKTKAFQYGC GGNANRFSTIECEKVC
Bungarus candidus	CDKPPDKGNCG--SVRRAFYYDTRLKTKAFQYGC GGNANRFSTIECEKVC
Bikunina	
Cavia porcellus	CQLGHSAGFCGLGFKRYF--YNGSSMAEIFFHYGC GGNANRFSTIECEKVC
Mus musculus	CQLNYSSEGFCLG-MQERY--YNGASMAEETFQYGC GGNANRFSTIECEKVC
Xenopus laevis	CRLAPASGFCCL--GNHNRYYNYSSTMAEETFQYGC GGNANRFSTIECEKVC
Tissue Factor Pathway Inhibitor	
Canis familiaris	CLTPADRGLCH--ANESRFYYNSVIGKCRPFKYGC GGNANRFSTIECEKVC
Macaca mulatta	CLAPADRGLCR--ANENRFYYNSVIGKCRPFKYGC GGNANRFSTIECEKVC
Oryctolagus cuniculus	CLPPADRGLCQ--ANEIRFFYNAIIGKCRPFKYGC GGNANRFSTIECEKVC
Rattus norvegicus	CLPADSGLCK--ASEKRFYYNPAIGKCRQFNKYGC GGNANRFSTIECEKVC
Mus musculus	CLQPADSGLCK--ASERRFYNSATGKCHRFNYGC GGNANRFSTIECEKVC
Brachydanio rerio	CHLEDEPGPCR--GLVPRYFFDQKSQECQFFYGC GGNANRFSTIECEKVC
Homo sapiens	CLLPDLDYGEPCR--ALLLRYYYDRYTQSCRQFLYGC GGNANRFSTIECEKVC
Inibidor-Inter-Alfa (Trypsin Inhibitor)	
Ovis aries	CQLGYSQGFCLGMFKRYF--YNGTSMACETFYGC GGNANRFSTIECEKVC
Equus caballus	CQLDHAQGFCLG-MISRY--FYNGTSMACETFYGC GGNANRFSTIECEKVC
Colágenos	
Homo sapiens	CKLPKDEGTCDRDFILKWY--YDPNTKSCARFWYGC GGNANRFSTIECEKVC
Gallus gallus	CLLQKEEGTCDRDFVLKWH--YDLKTKSCARFWYGC GGNANRFSTIECEKVC

Figura 3-4: Alinhamento de algumas das seqüências dos diversos membros da família BPTI Kunitz agrupados em suas respectivas subfamílias.

Três proteínas, chamadas caliclodinas, foram isoladas da anêmona marinha *Anemonia sulcata* (SCHWEITZ *et al.*, 1995) e pertencem à família BPTI Kunitz, tendo bastante homologia com o inibidor de calicreína extraído da mesma espécie. As anêmonas marinhas produzem toxinas com as quais paralisam as suas presas, que são as proteínas bloqueadoras dos canais de Na^+ . As caliclodinas inibem competitivamente a ligação da dendrotoxina às membranas do cérebro de rato e bloqueiam os canais de K^+ voltagem-sensíveis. Também foi descoberto que elas inibem tripsina, o que não é usual dentre as toxinas pertencentes à família BPTI Kunitz. As anêmonas marinhas apareceram por volta de 800 milhões de anos, o que ressalta a natureza fundamental e ancestral desse arranjo BPTI Kunitz (KONDO *et al.*, 1982).

A família BPTI Kunitz também tem muitos membros presentes nos venenos de cobras, alguns dos quais não são tóxicos, atuando primariamente como inibidores de serinoproteinases, enquanto muitos outros são neurotoxinas altamente específicas e potentes. As cobras surgiram há cerca de 100 milhões de anos e, desde então, algumas famílias de cobras desenvolveram independentemente sistemas de veneno incorporando glândulas que derivaram das glândulas salivares. Refletindo a via evolutiva que seguem as glândulas de veneno a partir de órgãos associados à digestão, o veneno de cobra contém numerosos compostos protéicos que compartilham uma história evolutiva com enzimas digestivas e seus inibidores, por isso é natural se esperar que as semelhanças estruturais também sejam compartilhadas (SKARZYNSKI, 1992; MARSHALL & HARAVEY, 1992).

Na categoria de neurotoxinas estão incluídas as dendrotoxinas e as caliclodinas. As dendrotoxinas são divididas em formas curtas, cujos representantes são a α -dendrotoxina e a toxina K com cerca de 57 aminoácidos, e longas, como as toxinas I e δ -dendrotoxinas com cerca de 60 aminoácidos. Algumas espécies possuem dendrotoxinas das duas formas, por exemplo, *D. polylepis* e *D. angusticeps*, o que, aliado com a alta homologia seqüencial e estrutural, dá uma pista de que ambas dividem uma seqüência ancestral comum. As dendrotoxinas mostram muito pouca atividade antiproteinases, mas são bloqueadores primários farmacologicamente ativos dos canais neuronais de K^+ voltagem-bloqueados e também facilitam a liberação de acetilcolina das membranas pré-sinápticas das sinapses colinérgicas. As caliclodinas (CaC) são estruturalmente homólogas às dendrotoxinas, mas

são potentes bloqueadores dos canais de Ca^{2+} ativados por alta voltagem não apresentando qualquer efeito em canais de K^{+} voltagem–bloqueados.

O arranjo global característico que forma a estrutura dos membros da família BPTI Kunitz, mostrado na Figura 3-5, é composto por uma folha- β antiparalela central a qual é fechada em um dos lados por uma volta- β formando um grampo- β e no outro lado por uma ponte dissulfeto. A seção da cadeia próxima ao N-terminal forma uma α -hélice curta e algumas vezes distorcida enquanto a seção próxima ao C-terminal forma uma α -hélice mais regular. A α -hélice C-terminal é covalentemente ligada tanto à folha- β central quanto à α -hélice N-terminal por pontes dissulfeto em pH fisiológico. As três pontes dissulfeto na estrutura são altamente conservadas, como se pode verificar no alinhamento mostrado na Figura 3-4, provavelmente para manter e reforçar a rigidez e estabilidade estrutural, o que é confirmado pela sua resistência extremamente forte à desnaturação por temperatura – o ponto de fusão de BPTI é de 95°C (BODE & HUBER, 1992) – por uréia ou por cloreto de guanidina – BPTI é essencialmente nativo numa solução 6M de cloreto de guanidina (BODE & HUBER, 1992). Quando essas três pontes dissulfeto são reduzidas, a proteína perde o seu arranjo estrutural e fica inativa.

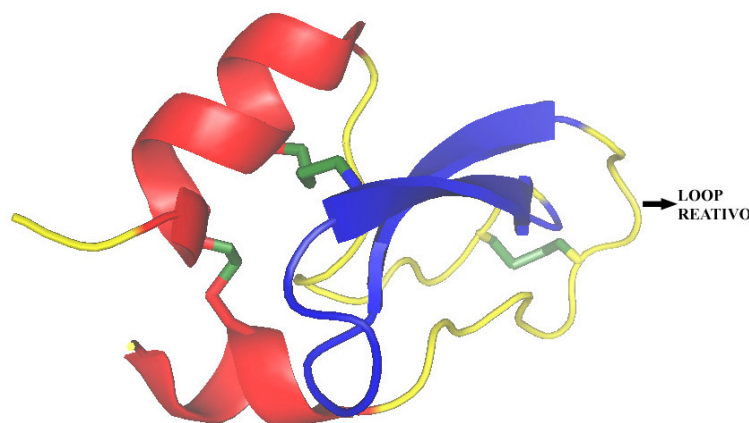


Figura 3-5: representação da estrutura 3D de BPTI (código PDB 1BPI). As pontes dissulfeto estão em verde e o loop reativo é indicado pela seta.

6.2. FAMÍLIA STI KUNITZ

Os membros dessa família têm uma massa molecular de cerca de 21kDa (aproximadamente 180 aminoácidos) com quatro meias-cistinas formando duas pontes dissulfeto. O inibidor extraído de soja estudado por Kunitz que deu nome à família possui

apenas um sítio de inibição para tripsina mas, existem outros membros que inibem quimotripsina, além de outros que têm dois sítios de inibição.

Tabela 2-6: Família Kunitz STI de inibidores

INIBIDOR	ATIVIDADE INIBITÓRIA PRINCIPAL	OUTRAS FUNÇÕES CONHECIDAS
Nodulina (encontrada nos nódulos de feijão alado quando a senescência começa)	Tripsina e outras proteinases endógenas	Interage com bactérias e se acredita estar envolvido na senescência nodular
Miraculina (encontrada em frutinhas vermelhas de <i>Richadella dulcifica</i>)	Função inibitória desconhecida, ~40% identidade seqüencial com STI.	Mascara o sabor azedo do sabor doce
Inibidor de amilase/subtilisina encontrado no trigo (WASI)	Subtilisina	Inibe α -amilase
Inibidor de amilase/subtilisina encontrado no arroz (RASI)	Subtilisina	Inibe α -amilase
Inibidor de amilase/subtilisina encontrado na cevada (BASI)	Subtilisina	Inibe α -amilase
Inibidor de catepsina D (PDI) (encontrado em batata)	Tripsina	Inibe catepsina D, uma aspartatoproteinase.
Novo inibidor de catepsina D (NID)	Tripsina	Inibe catepsina D
Asp proteinase induzida por injúria (CDI)	Tripsina	Inibe catepsina D
Inibidor de cisteinoproteinase de batata (PCPI)	Função inibitória desconhecida, ~23% identidade seqüencial com STI	Inibe catepsina L, uma cisteinoproteinase
Albumina de feijão alado (WBA-1)	Atividade inibitória desconhecida	Presumivelmente uma proteína de reserva

Os inibidores da família STI Kunitz são encontrados principalmente nas Leguminosas. Apesar da grande maioria desses inibidores ser composta de uma única cadeia polipeptídica, foram isolados alguns inibidores compostos de duas cadeias, uma cadeia maior com cerca de 12kDa e uma cadeia menor com cerca de 9kDa e com apenas uma ponte dissulfeto. Esses inibidores foram primeiramente isolados das sementes pertencentes às Mimosoideae (*Acácia*, *Adenantha* e *Albizia*), e, mais recentemente, das sementes de Caesalpinioideae (*Copaífera*) e de Papilionoideae (*Swartzia*). Essas duas cadeias são o resultado de um provável processamento proteolítico de um precursor com uma cadeia polipeptídica única, mas as razões pelas quais esta clivagem é restrita à essas espécies ainda são controversas, sendo sugerido por alguns autores que seja por razões evolutivas. A família STI Kunitz não está restrita às Leguminosae, certos inibidores bifuncionais de α -amilase/proteinase de cereal (Gramineae) exibem um grau de homologia

de 20–30% com os inibidores dessa família e existem também os inibidores de α -amilase/subtilisina de cevada, trigo e arroz. Diferentes dos inibidores tipo (STI) Kunitz de legume, os quais parecem não afetar as proteinases endógenas, os inibidores tipo (STI) Kunitz de cereal parecem ser importantes na regulação da atividade de uma α -amilase específica a qual mobiliza as reservas de amido das sementes durante a germinação. (MANEN *et al.*, 1991; THEERASILP *et al.*, 1989; MUNDY *et al.*, 1984; OHTSUBO & RICHARDSON, 1992; VALLEE *et al.*, 1998; MARES *et al.*, 1989; RITONJA *et al.*, 1990; HILDMANN *et al.*, 1990; KRIZAJ *et al.*, 1993; KORTT *et al.*, 1989)

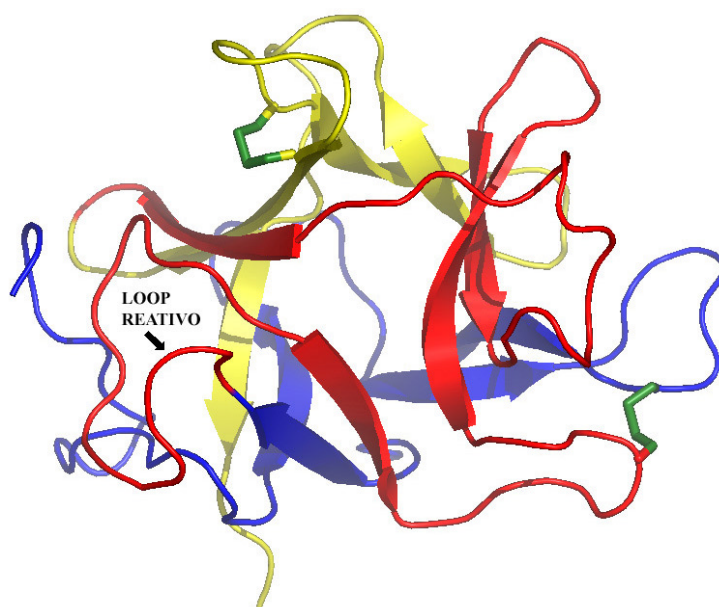


Figura 3-6: representação da estrutura tridimensional de STI (código PDB 1AVU) colorida de forma a mostrar as três unidades de repetição que formam o arranjo global conhecido como β -trefoil; as pontes dissulfeto são coloridas em verde; o loop reativo que contém a ligação peptídica é indicado pela seta.

A estrutura tridimensional da família STI Kunitz já é bem conhecida e a estrutura do complexo STI–Tripsina já foi determinada. A Figura 3-6 mostra o arranjo típico dessa família. A estrutura é composta por 12 fitas- β s sendo que 6 delas formam um barril- β e as outras 6 formam uma tampa para esse barril compondo um arranjo global que é conhecido como ‘ β -trefoil’.

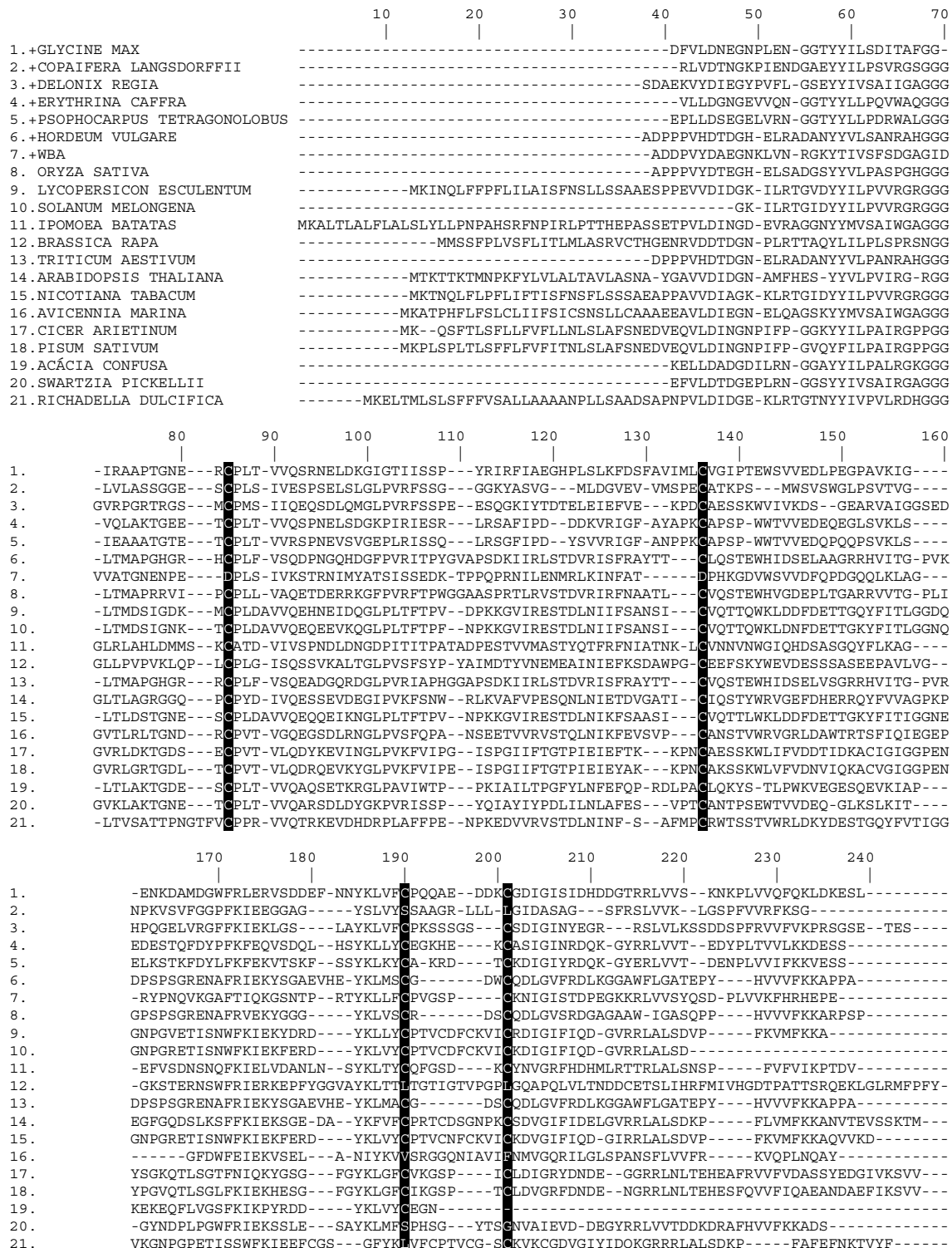


Figura 3-7: alinhamento entre algumas seqüências de membros da família STI Kunitz selecionados. Aqueles marcados (+) possuem estrutura tridimensional determinada. As meias cistinas conservadas estão destacadas em preto.

Esse arranjo global é também observado em outras proteínas funcionalmente não relacionadas com a família STI Kunitz. Essas proteínas incluem os fatores de crescimento (ZHU *et al.*, 1991; ZHANG *et al.*, 1991; ERIKSSON *et al.*, 1993; BELLOSTA *et al.*, 2001; PLOTNIKOV *et al.*, 2001), as interleucinas-1 α e 1- β (GRAVES *et al.*, 1990; PRIESTLE *et al.*, 1988), o antagonista do receptor de interleucina (SCHREUDER *et al.*, 1997), as toxinas como ricina (RUTENBER *et al.*, 1991), as aglutininas vegetais (TRANSUE *et al.*, 1997), a xilanase 10A (NOTENBOOM *et al.*, 2002), o domínio rico em cisteínas N-terminal do receptor de manose (LIU *et al.*, 2000) e as proteínas que ligam actina como hisactofilina humana (*et al.*, 1992) e fascina humana (ONO *et al.*, 1997). Essas famílias possuem funções distintas, diferentes localizações celulares, compartilham pouca ou nenhuma similaridade seqüencial detectável e têm diferentes ligantes e modos de ligação. Estudos filogenéticos sugerem que a causa da conservação estrutural do arranjo β -trefoil dentre essas famílias não relacionadas pode ser uma consequência da possível dupla duplicação gênica na superfamília dessas proteínas e também providenciam evidência estatística que essas proteínas com β -trefoil-fold todas vieram de um ancestral comum por evolução divergente (PONTING & RUSSELL, 2000).

Devido à homologia seqüencial e estrutural existente entre os inibidores do tipo (STI) Kunitz e uma albumina (WBA) extraída das sementes de *Psophocarpus tetragonolobus* (KORTT *et al.*, 1989) essa foi incluída na família STI Kunitz.

6.3. FAMÍLIA BOWMAN-BIRK (BBI)

Esse tipo de inibidor de proteinase é amplamente distribuído em sementes de Leguminosae e outras seqüências de aminoácidos homólogas foram encontradas no feijão comum *Phaseolus vulgaris*, no feijão *Vigna angularis* e *Macrotyloma axillare*, no amendoim *Arachis hypogaea* e em *Vigna unguiculada*. Também o inibidor induzido por injúria mecânica das folhas de alfafa *Medicago sativa* tem uma seqüência de aminoácidos a qual pertence à família Bowman-Birk. Além de leguminosas, a família Bowman-Birk também inclui o inibidor da sulfidril proteinase bromelaína isolado do abacaxi *Ananás sativus*, inibidores de tripsina de germe de trigo *Triticum aestivum*, arroz *Oryza sativa*, sementes de *Coix lachryma jobi*, sementes de capim *Setaria itálica* e cevada *Hordeum vulgare*. A família Bowman-Birk possui também similaridade seqüencial com as proteínas

de armazenamento 2S de *Ricinus communis* e do óleo das sementes de *Brassica napus*, as quais por sua vez têm homologia com os inibidores da superfamília cereal e com as proteínas/albuminas solúveis em clorofórmio-metanol (CM) extraídas dos endospermas de vários cereais. A observação dessa homologia seqüencial, ainda que fraca, leva à hipótese de uma origem evolucionária comum para segmentos curtos em várias proteínas de sementes. (ODANI, 1986; BIRK & APPLEBAUM, 1985; TSUNOGAE *et al.*, 1986; PRAKASH *et al.*, 1996)

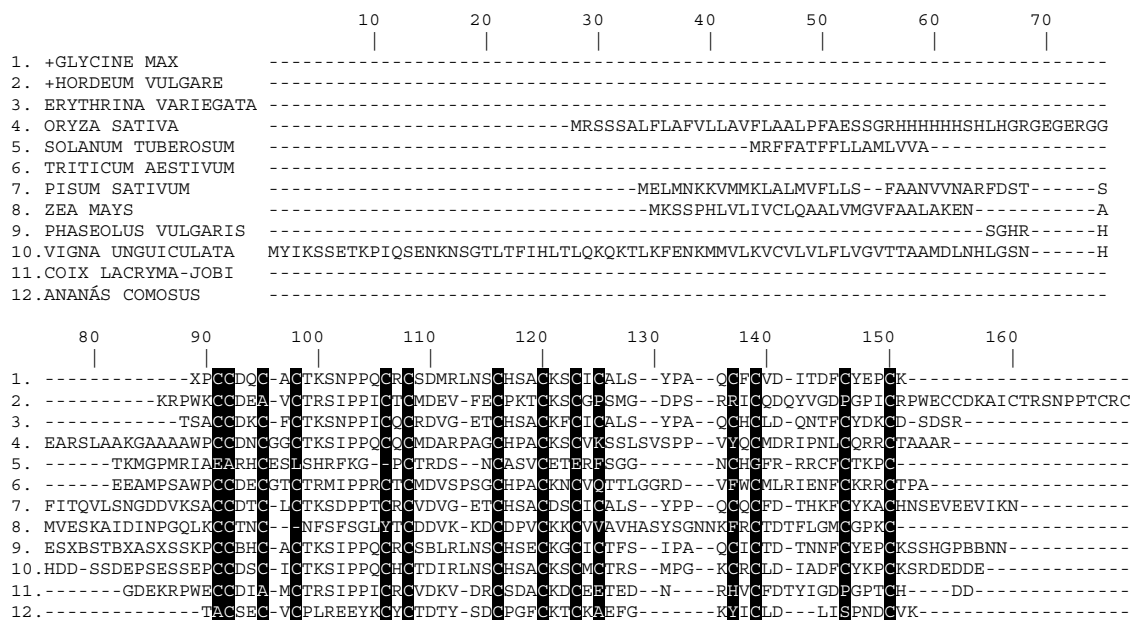


Figura 3-8: alinhamento entre algumas seqüências de membros da família Bowman-Birk selecionados. Aqueles marcados (+) possuem estrutura tridimensional determinada. As meias cistinas conservadas estão destacadas em preto.

As proteínas dessa família são caracterizadas por uma massa molecular de cerca de 8–9kDa e um alto conteúdo de meias-cistinas formando sete pontes dissulfeto. Eles possuem geralmente dois sítios de inibição e são capazes de formar complexos 1:1 com diferentes proteinases – usualmente tripsina e quimotripsina, raramente elastase – em sítios ativos independentes, cada qual numa região com seqüência repetida ou homologia interna. Essas regiões repetitivas ou domínios são o resultado de duplicação gênica. Interessantemente, o inibidor tipo Bowman-Birk de soja foi clivado por meios químicos e enzimáticos separando-se os seus dois domínios sem que nenhum deles perdesse sua atividade inibitória. (ODANI *et al.*, 1986)

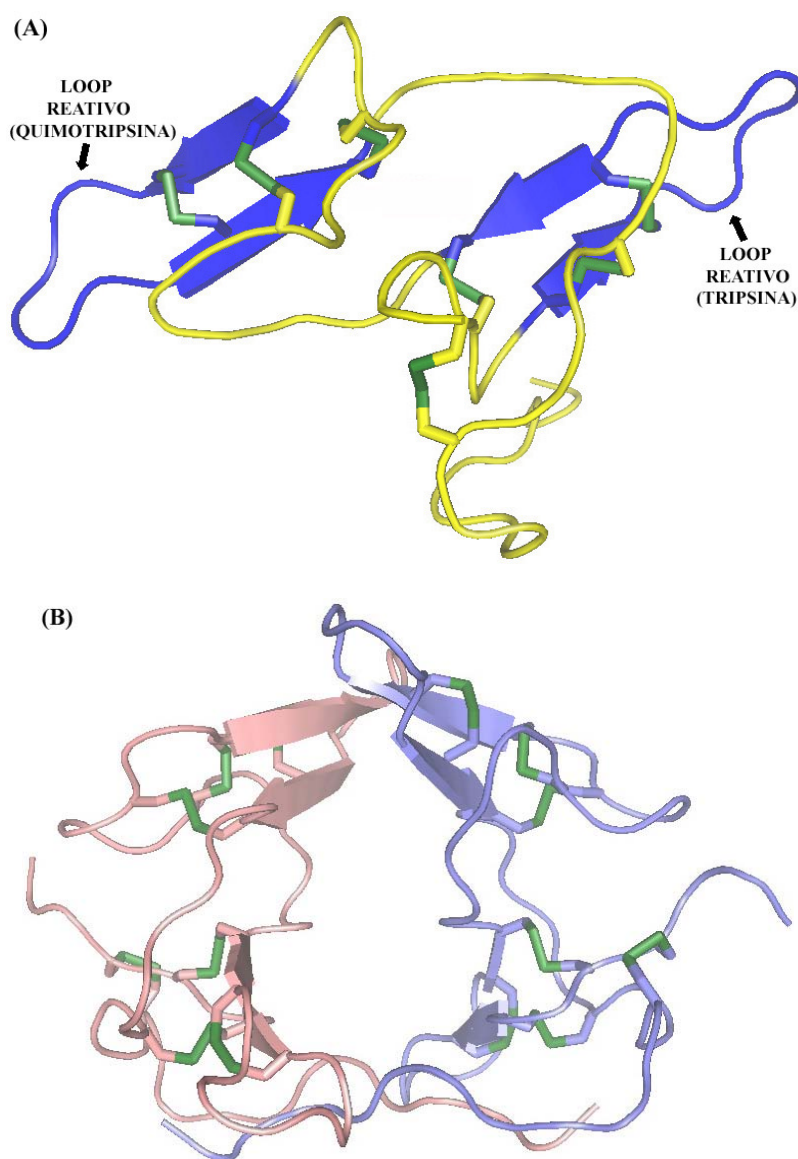


Figura 3-9: (A) representação da estrutura tridimensional de BBI (código PDB 1BBI) colorida de acordo com os elementos secundários estando as fitas- β s pregueadas em azul, os loops em amarelo e as pontes dissulfeto em verde; ambos os loops reativos que inibem tripsina e quimotripsina são indicados pelas setas; (B) estrutura do dímero formado pela junção de duas moléculas como mostradas em (A) sendo uma colorida em rosa e a outra em azul, as pontes dissulfeto estão em verde.

6.4. FAMÍLIA BATATA I

A família de inibidores do tipo batata I inclui todas aquelas proteínas que mostram uma grande homologia com o inibidor CI-1 de quimotripsina isolado de tubérculos de batata. Esse forte inibidor de quimotripsina, subtilisina e tripsina é um pentâmero cujos protômeros de 8kDa existem em pelo menos dez formas diferentes e são codificados por

mais de seis genes diferentes. As seqüências de aminoácidos desses protômeros consistem de cerca de 70 resíduos e contêm apenas uma ponte dissulfeto. A ligação peptídica do sítio ativo está localizada em Met/Leu(47)–Asp(48). (RICHARDSON, 1991)

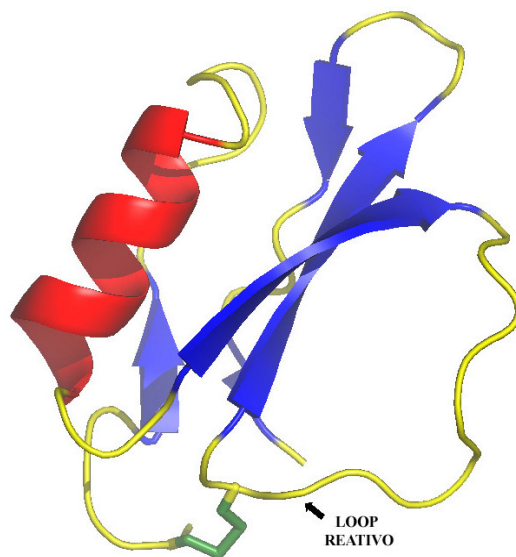


Figura 3-10: representação da estrutura tridimensional do inibidor tipo batata I extraído de *Cucúrbita maxima* (código PDB 1MIT). A única ponte dissulfeto é mostrada em verde e o loop reativo que contém a ligação peptídica é indicado pela seta.

A seqüência de aminoácidos de uma proteína induzida por injúria mecânica nas folhas de tomate foi deduzida a partir do cDNA e é 71% homóloga com o inibidor tipo batata I. Os inibidores de subtilisina de sementes de cevada e de feijão de fava têm 45% e 42% de homologia, respectivamente, com o inibidor de batata I, mas não apresentam nenhuma ponte dissulfeto. Outros inibidores de proteinase de 8kDa isolados das sementes de *Fagopyrum esculentum* também apresentam homologia com a família batata I, assim como um pequeno inibidor de tripsina com 62 aminoácidos das sementes de *Momordica charantia*. (RICHARDSON, 1991)

Todos os inibidores de plantas dessa família mostram uma homologia surpreendentemente forte (aproximadamente 36%) com um inibidor da sanguessuga *Hirudo medicinalis*. Dessa forma a família batata I é uma das mais amplamente distribuídas dentre as famílias de inibidores enzimáticos conhecidas, com exemplares presentes em Solanaceae, Gramineae, Leguminoseae, Polygonaceae, Cucurbitaceae e em animal. (RICHARDSON, 1991)

1. Solanum tuberosum	KLSPWPELVGVP	PAHYAKGII	EKENSLITNVQ	ILLNGSPVTMDYR	CNRVRLFDN	ILGDVVQ	-IPRVA
2.+Cucurbita maxima	KSSWPHLCV	GGSVAKAI	IERQNPVKA	VILEE-GTPVTKD	FRCNVRV	IWNKRLVVS	-PPRIG
3.+Linum usitatissimum	KNAPWPELVG	KSGSNMAA	ATVERENRN	VHAIVLKE-GS	AMTKDFRC	DRVWVIVNDH	GVVTS-VPHIT
4.+Hirudo medicinalis	LKSFPPEVVG	KTVQDARE	YFTLHYPO	YDVYFLPE-G	SPVTLDLR	YNRVRFYNPG	TNNVNHVPHVG
5.+Hordeum vulgare	KTEWPPELVG	KSVEEAKK	VIILQDKP	AAQIIVLPV-G	TIIVTMEY	RIDRVRLFVDR	LDNIAQ-VPRVG
6. Momordica charantia	KSSWPQLVG	STGAAAKA	VIERENPR	VRAVIKIV-G	SGATKDFRC	DRVRVWVTER	GIVAR-PPTIG
7. Sambucus nigra	KNTWPPELVG	ARGEAAAT	VETENPS	VTAVIPE-GS	IVTTDERC	DRVRVWVDEN	GIVTR-VPVIG
8. Zinnia elegans	KRSWPPELVG	ARTTAARF	ERENPRV	NAIVVVEGR	DMVTADFR	CDRVWVWVNS	NGVVLRTPSIG
9. Citrus paradisi	KSSWPPELVG	KNGEVA	AAAIIEKEN	PCVHAI	VLLKE-GTP	VTKDYRIDRV	RVWVWVVKGVIR-VPRIG
10.Linum usitatissimum	KNAPWPELVG	KSGSNMAA	ATVERENRN	VHAIVLKE-GS	AMTKDFRC	DRVWVIVNDH	GVVTS-VPHIT
11.Fagopyrum esculentum	KQEWPELVG	ERGSKAAK	IIENENED	VRAIVLPE-GS	AVPRDLLC	DRVDVFDVDER	GVVVD-TPVVM
12.Lycopersicon esculentum	KQMWPELVG	VPTKLAK	EIIKENPS	ITNIPILL	SGSPITLD	YLC	DRVRLFDN
13.Lycopersicon peruvianum	KQFWPELVG	VPAHYAKG	IIKENPS	ITNIPILL	NGSPVTKD	FRCDRV	RLFVNILGDVVQ-IPRVT
14.Amaranthus hypochondriacus	KQEWPELVG	EYGYKAA	IIERENPN	VRSIVKHE-R	SGFTKDFRC	DRVWVWVVD	STGVVVR-TPRVT
15.Amaranthus caudatus	KQEWPELVG	EYGYKAA	IIERENPN	VRSIVKHE-R	SGFTKDFRC	DRVWVWVVD	STGVVVR-TPRVT
16.Nicotiana glauca	KETWPPELVG	VPAKFARE	IIQKENS	KLTVPS	VNLNGSPVTKD	FRCDRV	RLFVNILGDVVQ-IPRVT
17.Nicotiana glauca	KETWPPELVG	VPAKFARE	IIQKENS	KLTVPS	VNLNGSPVTKD	FRCDRV	RLFVNILGDVVQ-IPRVT
18.Ipomoea batatas	TSSWPPELVG	VDALAK	STIEKENS	DLFVVFLFQ-G	CPESEYIF	DRVRLIVDC	SNVVR-TPSIG
19.Canavalia lineata	KTSPPELVG	VTAEAE	E-KIKEEM	SGVEIQVVP	-GSFVTADY	KPQRVRLV	YVDES
20.Phaseolus angularis	KTSPPELVG	VTAEAE	E-KIKEEM	SGVEIQVVP	-GSFVTADY	KPQRVRLV	YVDES
21.Vicia faba	RTSPPELVG	VSAAE	EAR-KIKEEM	PAEIQVVP	-DSFVTADY	KPQRVRLV	YVDES
22.Zea mays	KTSPPELVG	LSVEDAK	VILKDKP	DADIVVLPV-G	SVVTADYR	PNRVIRFVD	--IVAQ-TPHIG

Figura 3-11: alinhamento entre algumas seqüências de membros da família batata I selecionados. Aqueles marcados (+) possuem estrutura tridimensional determinada. As meias cistinas conservadas estão destacadas em preto.

6.5. FAMÍLIA BATATA II

O primeiro membro da família de inibidores do tipo batata II a ter a sua seqüência de aminoácidos determinada foi o inibidor de tripsina extraído do exocarpo de berinjela *Solanum melongena*. Peptídeos homólogos e imunologicamente relacionados (PCI e PTI), os quais são poderosos inibidores de serinoproteases foram isolados de tubérculos de batata *Solanum tuberosum*, de onde provém o nome da família. Além disso, folhas de tomate e batata mostraram conter genes para inibidores do tipo batata II cujas estruturas primárias foram deduzidas a partir das seqüências de nucleotídeos dos cDNAs. (RICHARDSON, 1991)

	10	20	30	40	50	60	70
1.+SOLANUM TUBEROSUM	-----	-----	-----	-----	-----	PICITNCCAGYKGG	NYNSA
2.+NICOTIANA ALATA	-----	-----	-----	-----	-----	DRICITNCCAGTKG	KYFSD
3.+NICOTIANA TABACUM	-----	-----	-----	-----	-----	DRICITNCCAGTKG	KYFSD
4. SOLANUM MELONGENA	-----	-----	-----	-----	-----	QICITNCCAGRKGG	SYFSE
5. LYCOPERSICON ESCULENTUM	MAVHKEVNFVAYLLI	--VLGMFLYVD	AKACTRECG-NL	GFICPRSEGS	PLNPIC	INCCSGYKGG	NYNSA
6. CAPSICUM ANNUUM	-----	-----	-----	-----	-----	RICITNCCAGRKGG	NYNSA
7. SOLANUM AMERICANUM	MAVHKVSFLACLLVL	GWFLAKHVDA	KACTRECG-HF	SYGICPRSEGS	PQKPICTNCC	SGYKGG	NYNSA
8. ATROPA BELLADONNA	-----	-----	-----	-----	-----	CDPEVAYMVCPS	SGHRIKPVVNCSSARKGGKLF

	80	90	100	110	120	130	140	150
1.	NGAFIC	EGQSDPKKPKAC	PLNCDPHIAYS	KCPR	-----	-----	-----	-----
2.	DGTFVC	EGESDPRNPKAC	PRNCDPRIAYG	ICPLA	-----	-----	-----	-----
3.	DGTFVC	EGESDPRNPKAC	PRNCDPRIAYG	ICPLS	-----	-----	-----	-----
4.	DGTFIC	KGESNPENPKAC	PRNCDGRIAYG	ICPLS	-----	-----	-----	-----
5.	FGKFI	EGESDPRNPKAC	PRNCDGRIAYG	ICPLS	-----	-----	-----	-----
6.	DGTFIC	EGESDPRNPKAC	PRNCDGRIAYG	ICPLS	-----	-----	-----	-----
7.	KGDLIC	EGESDPRNPKAC	PRNCDGRIAYG	ICPLS	-----	-----	-----	-----
8.	NGSVK	TGT	-----	-----	-----	-----	-----	-----

Figura 3-12: alinhamento entre algumas seqüências de membros da família batata II selecionados. Aqueles marcados (+) possuem estrutura tridimensional determinada. As meias cistinas conservadas estão destacadas em preto.

Os membros dessa família apresentam homologia seqüencial interna entre dois domínios, sugerindo duplicação e elongação gênica. O sítio ativo do inibidor de tripsina extraído de berinjela foi identificado como a ligação peptídica Arg(38)–Asn(39). A mesma ligação peptídica também foi encontrada num local homólogo no inibidor de tripsina de batata (PTI-1), enquanto os inibidores de quimotripsina (PCI-1 e 2) contêm Leu–Asn e Phe–Asn, respectivamente, o que é inteiramente consistente com suas especificidades. (RICHARDSON, 1991)

A família batata II de inibidores também tem um grau de homologia muito limitado com as regiões N–terminais dos inibidores de carboxipeptidase de batata e tomate. Mas esse nível de homologia é insuficiente para sugerir que os dois tipos pertencem à mesma família. Da mesma forma que a fraca semelhança de seções curtas do inibidor de tripsina de berinjela com a proteína PAPI de cevada não parece ser biologicamente significativa. (RICHARDSON, 1991)

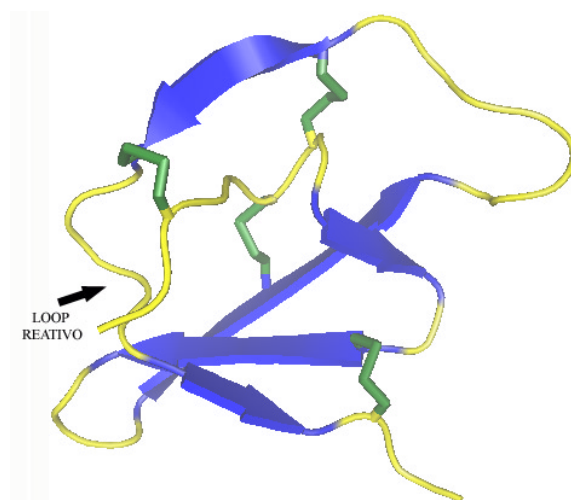


Figura 3-13: representação da estrutura tridimensional do inibidor tipo batata II extraído de *Nicotiana alata* (código PDB 1TIH). As pontes dissulfeto estão em verde e o loop reativo é indicado pela seta.

6.6. FAMÍLIA ABÓBORA (OU FAMÍLIA CUCÚRBITA)

Os menores inibidores conhecidos são os inibidores de tripsina isolados das sementes da família Cucurbitaceae. Todos contêm cerca de 29 aminoácidos e, a despeito de seu pequeno tamanho, são inibidores muito potentes com suas constantes de associação (K_a) na faixa de $5,9 \times 10^{10}$ – $9,5 \times 10^{11} \text{ M}^{-1}$, estando entre as maiores já reportadas para inibidores de tripsina. Esses inibidores altamente homólogos possuem três pontes dissulfeto

e a ligação peptídica do seu sítio ativo é formado por Arg-Ile ou Lys-Ile. (RICHARDSON, 1991)

	10	20	30
1.+Cucúrbita máxima	RVCPRLMBCKKDSDDLAEVCL	LEHG-YCG	
2.+Cucúrbita pepo	RVCPKILMBCKKDSDDLAEVCL	LEHG-YCG	
3.+Momordica charantia	RIQPLIWMBCRKDSDDLACQICVDG--	HCG	
4.+Ecballium elaterium	-GCPRLMRCKQDSDDLACGVCGP	NG-FCG	
5. Momordica repens	GICPRLMBCKKDSDDLACQVCKRQ	YCG	
6. Momordica cochinchinensis	RAQPRILKKRRDSDEPGEICKENG	YCG	
7. Cucumis melo	VGCPRLMKCKTDRDLTGCTCKRNG	YCG	
8. Cucumis sativus	MMCPRLMKCKHDSDDLPGCVCL	EHIEYCG	
9. Luffa cylindrica	RIQPRILMBCKSSDSDDLAEVCL	ENG-FCG	
10.Cyclanthera pedata	RIQPRILMKCKKDSDDLAEVCL	QEHG-FCG	
11.Citrullus lanatus	RRCPRIYMBCKRDADCLADCVCL	QHG-ICG	
12.Lagenaria leucantha	RRCPRIYMBCKHDSDDLADCVCL	EHG-ICG	
13.Bryonia diodica	RGCPRLMRCKKDSDDLACGVCK	QKNG-YCG	
14.Trichosanthes kirilowii	VACPRILMBCKVNDDLRCCKCL	SNG-YCG	

Figura 3-14: alinhamento entre algumas seqüências de membros da família abóbora (squash) selecionados. Aqueles marcados (+) possuem estrutura tridimensional determinada. As meias cistinas conservadas estão destacadas em preto.

Muitas das formas de isoinibidores reportadas parecem ter sido produzidas pela remoção proteolítica de pequenos di ou tripeptídeos a partir do C- e N-terminais de cadeias maiores. Esses pequenos inibidores podem ser facilmente sintetizados *in vitro* usando sistemas de fase sólida e têm as mesmas características imunológicas e inibitórias das proteínas isoladas de sementes. Também, a substituição química da Arg do sítio ativo por Ala converte o inibidor de tripsina num inibidor de elastase. (RICHARDSON, 1991)

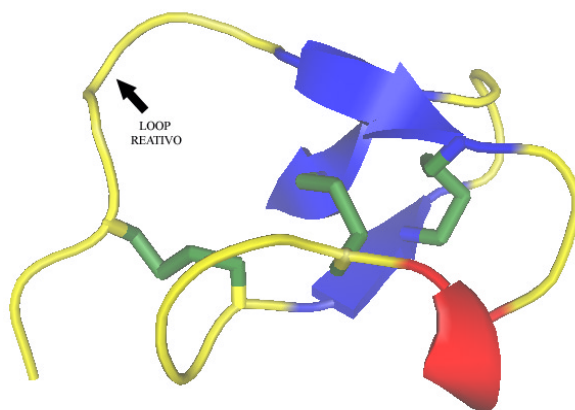


Figura 3-15: representação da estrutura tridimensional do inibidor tipo abóbora extraído de Cucurbita maxima (código PDB 3CTI). As pontes dissulfeto estão em verde e o loop reativo é indicado pela seta.

6.7. SUPERFAMÍLIA CEREAL

Estudando-se as estruturas primárias de numerosos inibidores encontrados em grãos de cereais chegou-se à conclusão de que existe uma superfamília de proteínas homólogas,

as quais incluem inibidores de α -amilase e outras proteinases, inibidores bifuncionais ativos contra duas ou mais classes de proteinases e proteínas de armazenamento com funções inibitórias desconhecidas. A primeira sequência de aminoácidos completa de um inibidor de α -amilase a ser determinada foi da forma monomérica de 13kDa de trigo. Depois, outras formas diméricas de 26kDa de inibidores de α -amilases exógenas de trigo, bem como o inibidor de amilase de centeio, mostraram-se altamente homólogas. Por sua vez as sequências do inibidor bifuncional de amilase/tripsina de *Eleusine coracana*, dos inibidores de tripsina de cevada, milho e centeio e dos inibidores de α -amilase de sorgo exibem altos níveis de homologia (50–76%) entre si. O que pode ser visto no alinhamento feito com alguns membros da superfamília cereal mostrado na figura 3-16. (RICHARDSON, 1991)

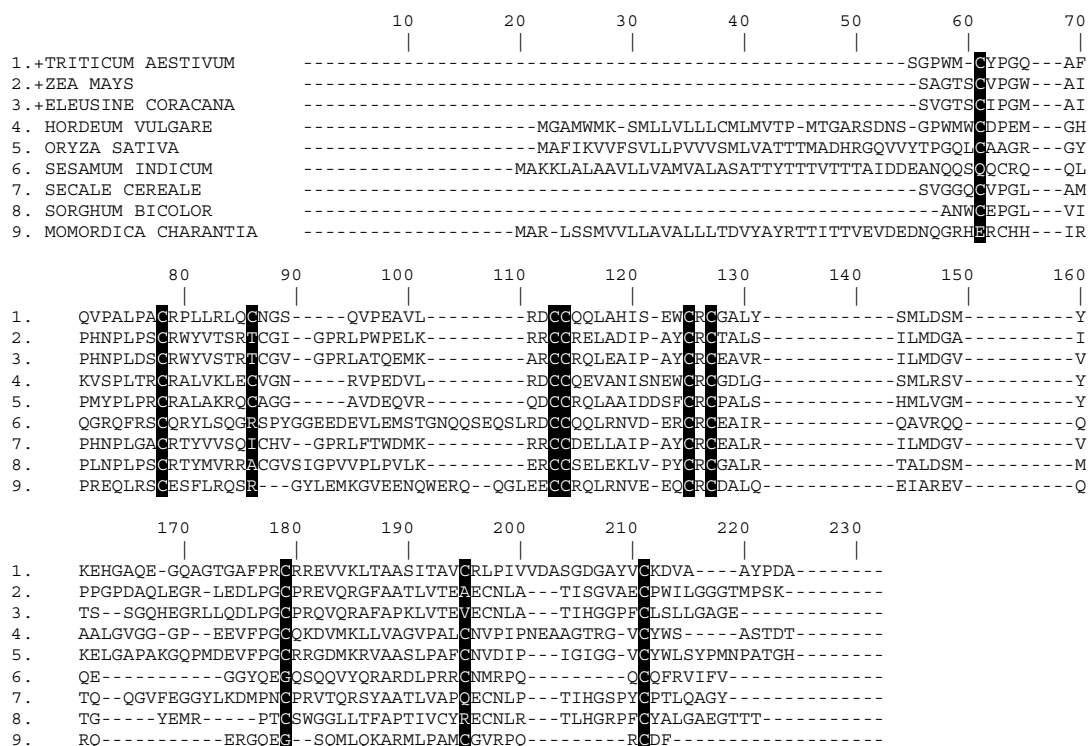


Figura 3-16: alinhamento entre algumas sequências de membros da superfamília cereal selecionados. Aqueles marcados (+) possuem estrutura tridimensional determinada. As meias cistinas conservadas estão destacadas em preto.

Foram incluídas no alinhamento as sequências de algumas das chamadas proteínas solúveis em clorofórmio-metanol (CM) extraídas de cevada e trigo. Inicialmente o papel dessas proteínas era desconhecido, mas sua homologia com os inibidores de cereal deu a primeira pista sobre sua função. O que foi confirmado quando foi demonstrado que CMA de

cevada inibe α -amilase de *Tenebrio molitor*, e tanto CMc quanto CMe inibiram tripsina. CMB e CMD, as quais não têm atividade inibitória por si só, podem combinar com CMA para formar um inibidor tetramérico. Essas proteínas são codificadas em trigo e cevada por famílias multigênicas dispersas em vários locais cromossômicos. (RICHARDSON, 1991)

Todos os inibidores homólogos dessa família contêm uma ligação peptídica Arg-Leu no sítio reativo nas posições que correspondem aos resíduos 33–34 do inibidor de cevada, com exceção dos inibidores de α -amilase.

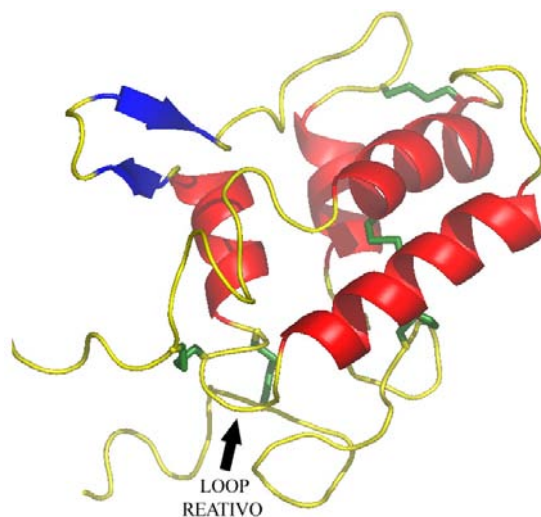


Figura 3-17: representação da estrutura tridimensional do inibidor da superfamília cereal extraído de *Eleusine coracana gaertneri* (código PDB 1BIP). As pontes dissulfeto estão em verde e o loop reativo é indicado pela seta.

6.8. FAMÍLIA RAGI I-2

Um segundo inibidor de α -amilase (ragi I-2) isolado das sementes de *Eleusine coracana* mostrou uma sequência de aminoácidos completamente diferente daquela do inibidor bifuncional da mesma semente (superfamília cereal), a despeito do fato de que ambos têm especificidade de inibição para α -amilase. De fato, relatos posteriores de proteínas similares em outros cereais indicaram que os inibidores do tipo ragi I-2 pertencem a uma família nova e separada. (RICHARDSON, 1991)

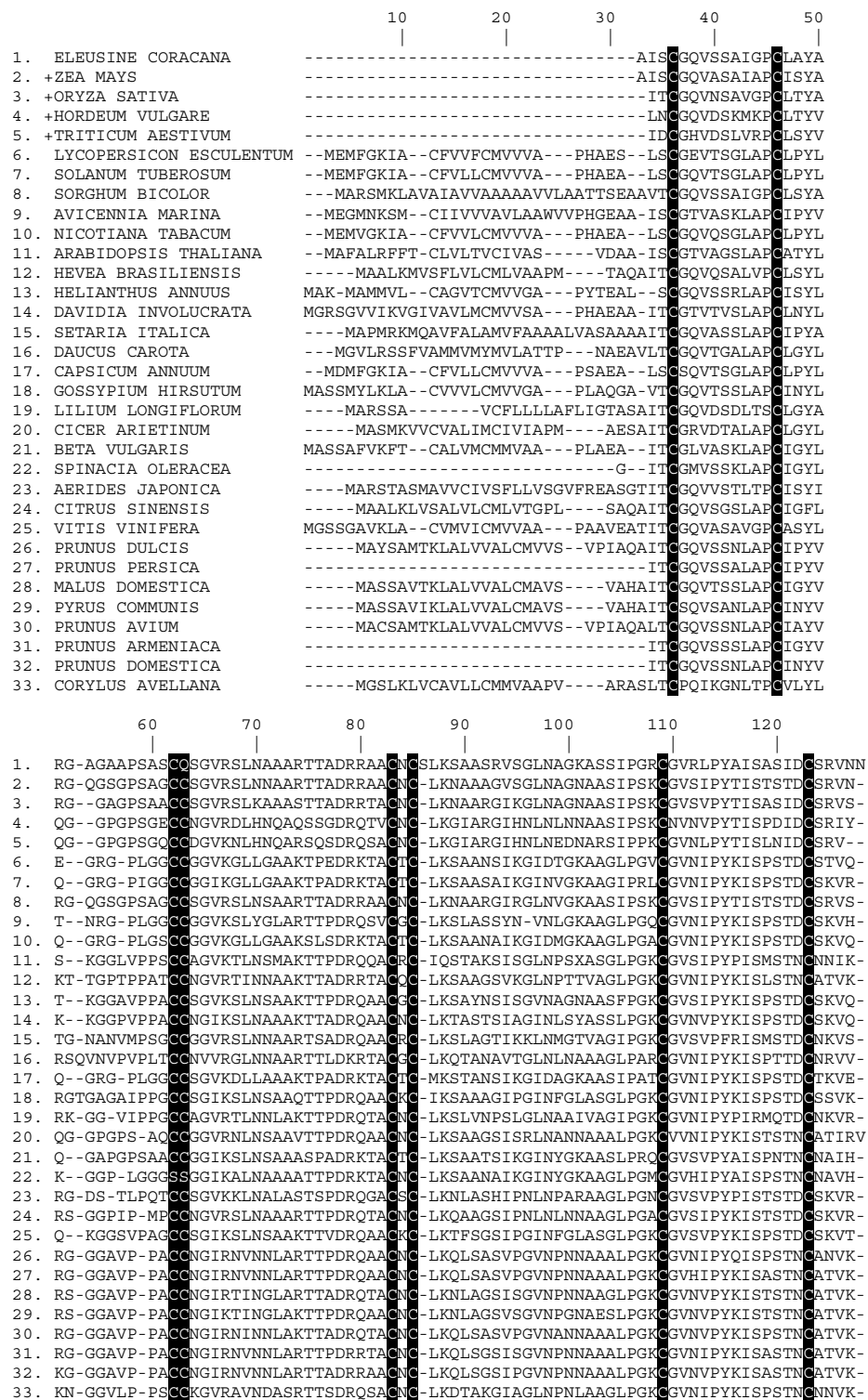


Figura 3-18: alinhamento entre algumas seqüências de membros da família RAGI I-2 selecionados. Aqueles marcados (+) possuem estrutura tridimensional determinada. As meias cistinas conservadas estão destacadas em preto.

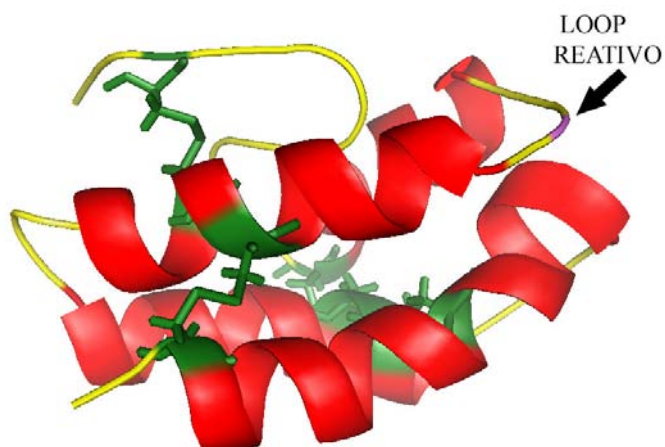


Figura 3-19: representação da estrutura tridimensional do inibidor da família ragi I-2 extraído de *Hordeum vulgare* (código PDB 1LIP). As pontes dissulfeto estão em verde. Note como a estrutura dessa família é bem diferente da superfamília cereal não contendo folhas- β s pregueadas. A conformação do loop reativo também é diferente pois eles são inibidores de α -amilase.

6.9. FAMÍLIA MILHO 22kDa / THAUMATINA / PR-PROTEÍNAS

Talvez um dos mais surpreendentes resultados fornecidos pelos estudos das estruturas de inibidores tenha sido a descoberta de que um dos inibidores de α -amilase isolado de milho é de fato um inibidor bifuncional de tripsina de mamíferos e α -amilases de insetos e tem uma homologia seqüencial grande (52%) com uma proteína intensamente doce chamada taumatina, extraída dos frutos de *Thaumatococcus danielli*, e 57% de homologia com as chamadas ‘proteínas relacionadas com patogênese’ (ou PR-proteínas), induzidas em plantas de tabaco pela infecção com o vírus mosaico do tabaco. Também existe forte homologia seqüencial entre o inibidor bifuncional de milho e proteínas induzidas por estresse osmótico em tomate e tabaco. O inibidor de 22kDa não tem homologia com o inibidor de 13kDa extraído do mesmo tecido de milho. (RICHARDSON, 1991)

Essa inesperada similaridade seqüencial de taumatina e PR-proteínas com o inibidor bifuncional de 22kDa de milho sugere que essas proteínas, cuja função ainda é desconhecida, tenham um efeito inibitório em enzimas hidrolíticas. Entretanto, outros pesquisadores que notaram uma fraca semelhança entre curtas regiões de uma PR-proteína e os inibidores de proteinases da família Bowman-Birk não foram capazes de detectar qualquer inibição de tripsina por essa PR-proteína.

	10	20	30	40	50	60	70	80	
1. +THAUMATOCOCCLUS DANIELLI	-----	-----	-----	ATFEIVNR	CSYTVWAAASKGDAA	-----	LDAGGRQLNS	-GESWTINVEPGT	
2. +NICOTIANA TABACUM	-----	-----	-----	SGVFEVHNN	CPYTVWAAATPV	-----	GGRRRLER	-QGSWWFWAPPGT	
3. +ZEA MAYS	-----	-----	-----	AVFTVVNQ	CPFTVWAAASVPV	-----	GGGRQLNR	-GESWRITAPAGT	
4. GLYCINE MAX	-----	-----	-----	ARFEITNR	CTYTVWAAASVPV	-----	GGGVQLNP	-QGSWSVDVPAGT	
5. LYCOPERSICON ESCULENTUM	-----	-----	FFFLLA	FVITYTYAATFEVRNN	CPYTVWAASTPI	-----	GGRRRLDR	-GQTWVINAPRGT	
6. ORYZA SATIVA	MTKKT	KAMAPSLATSS	TALFL	LVAVS	-IADAATFAITNR	COYTVWPAAVPS	-----	GGGTKLDP	-GQTWVINVPAGT
7. HORDEUM VULGARE	-----	MASLPTSS	VLLPILL	LVVAATADAATFTVINK	COYTVWAAAVPA	-----	GGGQKLDA	-GQTWSINVPAAGT	
8. SOLANUM COMMERSONII	-----	MAYLRSS	SVFLLAF	VITYTYAATIEVRNN	CPYTVWAASTPI	-----	GGRRRLDR	-GQTWVINAPRGT	
9. ARABIDOPSIS THALIANA	-----	MANLLV	STFI	FSALLLISTATAATFEILNQ	CSYTVWAAASP	-----	GGRRRLDA	-QGSWRLDVAAGT	
10. CAPSICUM ANNUUM	-----	MGYLRSS	SVFL	LAFVITYTYAATFEVRNN	CPYTVWAASTPV	-----	GGRRRLDR	-GQTWVINAPPAGT	
11. CICER ARIETINUM	-----	TRSLIT	ITLCSLL	FLLLPSQAANFEIVNN	CPYTVWAAASP	-----	GGRRRLDR	-GQTWNLWVNAAGT	
12. AVENA SATIVA	-----	MASSSVL	QLLLML	LATITATTTVNNK	CSYTVWPGALP	-----	GGVRLLDP	-GKSVALWNAAGT	

	90	100	110	120	130	140	150	160	170	180								
1. KGGKI	WART	CYFDDSGSGI	CTGDC	G-GLLR	C-KRFR	GRPTTLAE	FSNLNQY	-KDYIDIS	NIKGFNV	PMDFSPTT	---RGGRG	-VRCAAD	IVGCP	PAKLK	APG			
2. KMARI	WGR	TNCNFDG	AGRGC	QTGDC	G-GVLE	C-KGWG	KPNTLAE	YALNQF	SNLDFWD	ISLVDG	FNIPMS	FGPTK	PGPK	CHG	-IQCTAN	INGEC	PGSLR	VPG
3. TAARI	WART	CKFDAS	GRGSC	RTGDC	G-GVLC	C-TGYG	RAPNTLAE	YALKQ	NNLDFD	DISLIDG	FNVPMS	FLPD	-GGSGC	SRG	-PRCAVD	VNAR	CPAELR	QDG
4. KGARV	WART	CCNFDG	SGRGC	QTGDC	G-GVLD	C-KAYG	APNTLAE	YGLNG	FNNLDFD	ISLVDG	FNVPMS	FLPD	-GGSGC	SRG	-PRCAVD	VNAR	CPAELR	QDG
5. KMARI	WGR	TNCNFDG	AGRGC	QTGDC	G-GVLC	C-TGWG	KPNTLAE	YALNQF	SNLDFWD	ISLVDG	FNIPMT	FAPTNP	SGGK	CHA	-IHCTAN	INGEC	PGSLR	VPG
6. TGGRV	WAGR	-CGFDG	SNGGC	ETGDC	G-GKLR	C-TAYG	APNTLAE	FALNQ	NNLDFD	ISLIDG	FNVPMA	FLPAGS	GAGC	PKGGP	RCATAI	TPOC	PSLRA	PG
7. TSGRV	WART	CCSFDG	AGNGR	QTGDC	G-GKLR	C-TQYG	QAPNTLAE	FGLNK	YMGQDFD	ISLIDG	YNVPMS	FLPAGS	TGCP	KGGP	RCPKVI	TPOC	PSLRA	PG
8. KMARI	WGR	TNCNFDG	AGRGC	QTGDC	G-GVLC	C-TGWG	KPNTLAE	YALNQF	SNLDFWD	ISLVDG	FNIPMT	FAPTNP	SGGK	CHA	-IHCTAN	INGEC	PGSLR	VPG
9. KMARI	WGR	TNCNFDG	SGRGC	QTGDC	G-GVLC	C-TGWG	HPPNTLAE	YALNQF	NNLDFD	ISLVDG	FNIPMT	FAPTNP	SGGK	CHA	-IHCTAN	INGEC	PGSLR	VPG
10. AMARI	WGR	TNCNFDG	SGRGC	QTGDC	G-GVLC	C-TGWG	KPNTLAE	YALNQF	NNLDFWD	ISLVDG	FNIPMT	FAPTNP	SGGK	CHA	-IQCTAN	INGEC	PGSLR	VPG
11. SMARI	WGR	TNCNFDG	SGRGC	ETGDC	G-GVLC	C-TGWG	VPPNTLAE	FALNQ	NNLDFD	ISLVDG	FNIPMT	FAPTNP	SGGK	CHA	-IQCTAN	INGEC	PGSLR	VPG
12. RGARV	WPR	TGCTFD	ASGRGH	VTGDC	G-GALC	C-RVSG	QAPNTLAE	YTLG	KGAKDFD	ISLVDG	FNVPMS	FQPVG	-GAPRG	-ATCAAD	ITKCP	QELK	VAG	

	190	200	210	220	230	240	250	260	270	280									
1. GG-----	CNDAC	CTVFQ	TSEYCC	CT-----	TGRC	GPTEYS	RFFKRL	CPD	AFSVY	LDKPTTV	---TCPGS	-SNYRV	TF	PTA	-----	-----	-----	-----	
2. G-----	CNNP	CTTFG	QQYCC	CT-----	QGFC	GPTELS	RWFKQR	CPD	AYSYP	QDDPTST	--FTCT	SWTTD	YKVM	FCP	---YG	-----	-----	-----	
3. V-----	CNNA	CPVFK	DEYCC	GSA--	ANDCH	PTNYS	RYFKQR	CPD	AYSYP	KDDATST	--FTCP	PAG	-TNYK	VVFCP	-----	-----	-----	-----	
4. G-----	CNNP	CTVFK	TDQYCC	CT-----	SGSC	GPTD	YSRFFKQR	CPD	AYSYP	KDDPPST	--FTCNGG	-TDYR	VVFCP	-----	-----	-----	-----	-----	
5. G-----	CNNP	CTTFG	QQYCC	CT-----	QGFC	GPTD	LSRFFKQR	CPD	AYSYP	QDDPTST	--FTCP	SGSTN	YRVVFCP	-----	NGVTS	-PNFPLE	MPSSD	EDEAK	
6. G-----	CNNA	CTVFR	QDRYCC	GSA--	ANS	CGPTN	YSEFFKRL	CPD	AYSYP	KDDASST	--YTCP	PAG	-TNYQ	VVFCP	-----	-----	-----	-----	
7. G-----	CNNA	CTVFK	EDRYCC	GSA--	ANS	CGPTD	YSRFFKQR	CPD	AYSYP	KDDATSI	--FTCP	PGG	-TNYQ	VVFCP	-----	-----	-----	-----	
8. G-----	CNNP	CTTFG	QQYCC	CT-----	QGFC	GPTD	LSRFFKQR	CPD	AYSYP	QDDPTST	--FTCP	SGSTN	YRVVFCP	-----	NGVTS	-PNFPLE	MPSSD	EDEAK	
9. G-----	CNNP	CTVFK	TNQYCC	CTN-G	-QGS	CDTD	YSRFFKQR	CPD	AYSYP	QDDPTST	--FTCT	CTN	-TNYR	VVFCP	-----	PRSR	LATG	SHQLPI	KMVTEEN
10. G-----	CNNP	CTTFG	QQYCC	CT-----	QGFC	GPTELS	KFFKQR	CPD	AYSYP	QDDATST	--FTCP	SGSTN	YRVVFCP	-----	NGVTG	-PNFPLE	MPSSD	GVAK	
11. G-----	CNNP	CTVYK	TNEYCC	CTN-G	-QGS	CGPTN	FSTFFKDR	CHD	AYSYP	QDDPTST	--FTCP	PAG	-SNYK	VVFCP	-----	LGAPH	-IEMPL	IQTNVY	
12. G-----	ASA	GRFG	GDV	RGGR	-TDK	PPNTN	YSKFFKQR	CPD	AYSYP	KDDATST	--FTCP	VVG	-TNYQ	VVFCP	-----	-----	-----	-----	

Figura 3-20: alinhamento entre membros da família milho22kDa/taumatina/PR-proteínas selecionados. Aqueles marcados (+) possuem estrutura tridimensional determinada. As meias cistinas conservadas estão destacadas em preto.

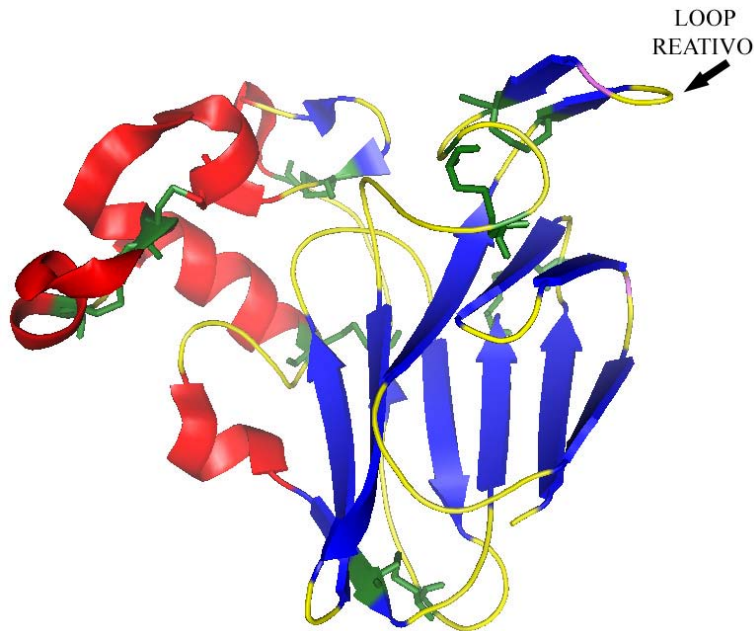


Figura 3-21: representação da estrutura 3D do inibidor extraído de milho Zea mays (código PDB 1DU5). Apesar de ser bifuncional, inibindo α -amilase de insetos e tripsina de mamíferos, apenas o loop reativo da tripsina é mostrado.

6.10. FAMÍLIA KAZAL

Tabela 2-7: Família Kazal de inibidores.

INIBIDOR	ATIVIDADE INIBITÓRIA PRINCIPAL	OUTRAS FUNÇÕES CONHECIDAS
Ovoinibidor	Inibe uma grande variedade de serinoproteinases tendo vários domínios de inibição	Inibe o canal ativado por $\text{Ca}^{++}/\text{K}^{+}$.
Folistatina	Atividade inibitória desconhecida	Modula a ação de vários membros da família do fator β de crescimento
Agrina Dentre outros, contém nove domínios folistatina	Alguns dos domínios Kazal/folistatina da agrina inibem CARL e HLE	Está associada com membranas basais de vários tecidos como a lâmina basal sináptica da junção neuromuscular.
BM-40 (SPARC ou osteonectina) Dentre outros, contém um domínio folistatina	Atividade inibitória desconhecida	É uma glicoproteína antiadesiva secretada a qual acredita-se estar envolvida no remodelamento dos tecidos.
PEC-60	Inibe tripsina	É estruturalmente similar ao fator de liberação da colecistokinina. Também é conhecido por suprimir a secreção de insulina induzida por glicose.
Trombospondina 1 (uma glicoproteína com multidomínios)	Contém dois domínios tipo Kazal e inibe HLE e catepsina G	Acredita-se estar envolvida no desenvolvimento e remodelamento de tecidos.
PSTI II	Inibe tripsina.	Mostra atividade de liberação de colecistokinina.
Rodnina	O primeiro domínio da rodniina inibe fortemente trombina	O segundo domínio da rodniina se liga ao sítio da trombina que reconhece e liga fibrinogênio mas não possui atividade inibitória.

Essa família foi nomeada depois que L. Kazal descobriu o primeiro dos inibidores de tripsina da secreção pancreática (PSTI – *Pancreatic Secretory Trypsin Inhibitor*). Os inibidores da família Kazal estão presentes em todos os vertebrados, inclusive no homem, e a sua função fisiológica é evitar a ativação prematura de pró-enzimas, inibindo traços de tripsina ativa que possam estar livres no pâncreas. Os inibidores do tipo Kazal inibem uma série de serinoproteinases, como tripsina e elastase. Mas eles não estão limitados aos vertebrados. Extratos da sanguessuga *Hirudo medicinalis* contém numerosos inibidores de tripsina e um deles mostra pelo menos um domínio do tipo Kazal. Na tabela 2-7 são apresentados os principais membros da família Kazal. A família tem muitos membros e,

além dos inibidores propriamente ditos também inclui outras proteínas que não apresentam propriedades inibitórias como as folistatinas e as osteonectinas as quais apresentam vários domínios do tipo Kazal em sua estrutura. Outras proteínas que se destacam são os ovoinibidores, proteínas abundantes em ovos de aves, e as agrinas pois ambas possuem de um a sete domínios do tipo Kazal em sua estrutura. (MOSS *et al.*, 1996; MATZUK *et al.*, 1995; DENZER, *et al.*, 1998; HOHENESTER *et al.*, 1997; AGERBERTH *et al.*, 1989; HOGG *et al.*, 1994; TSUZUKI *et al.*, 1991; VAN DE LOCHT *et al.*, 1995)

Cada domínio do tipo Kazal é composto por cerca de 56 resíduos de aminoácidos e três pontes dissulfeto. A estrutura do domínio Kazal lembra a estrutura do BPTI, tendo também a forma de pinça com uma haste sendo a α -hélice e a outra haste sendo a folha- β formada pelas três fitas- β s.

PSTI	
HOMO SAPIENS	NCSQYRLP-----GCPRH-----FNPVCGSDMSTYANEC ~~T~~ LCMKIRE-----GGHNIKIIRNGFC
BOS TAURUS	DCKVYTE-----ACTRE-----YNPICD ~~S~~ AAKTYSN ~~E~~ CF ~~C~~ NEKMN-----NDADIHFNFHGEC
SUS SCROFA	DCDVYRSHLF-----FCTRE-----MDPICGTNGKSYANPC ~~I~~ FCSEKLG-----RNEKFD ~~F~~ GHWGHC
CANIS FAMILIARIS	BCTEYSD-----MCTMD-----YRPLCGSDGKNYSNK ~~S~~ FCNAVKK-----SRGTIFLAKHGEC
RATTUS NORVEGICUS	NCPKQIM-----GCPRI-----YDPVCGTNGITYPSE ~~C~~ SLCFENRK-----FGT ~~S~~ IHIQRRGTC
PANTHERA LEO	NCSQYNR-----KSGGIACSKQ-----LKPICGIDHKTYSN ~~E~~ CMFCLLNQN-----KQFQIRKLYDDKC
MELES MELES	NCSKYNA-----KGSQFACSRH-----LDPVCGTDHRTYSN ~~E~~ CMFCMLTQN-----KRFSVRILQDNNC
ANEMONIA SULCATA	DCPL-----ICTMQ-----YDPVCGSDGITYGNAC ~~M~~ LLGASCR-----SDTPIELVHKGR ~~C~~
OVOINIBIDOR	
GALLUS GALLUS	
DOMÍNIO 1	NCSLYASGIGKDGTSWVAC ~~P~~ RN-----LKPVC ~~G~~ TDGSTYSN ~~E~~ C ~~G~~ ICLYNRE-----HGANVEKEYDGE ~~C~~
DOMÍNIO 2	DCSRFPNATDK ~~E~~ GKDVLCNKD-----LRPIC ~~G~~ TDGVITYND ~~C~~ LLCAYSIE-----FGTNISKEHDGE ~~C~~
DOMÍNIO 3	DCSKYKTSTLKDGRQV ~~A~~ CTMI-----YDPVCATNGVTYASE ~~C~~ TLCAHNLE-----QRTNLGKRKNGR ~~C~~
DOMÍNIO 5	HCR ~~E~~ FQKVSP-----ICTME-----YVPHCGSDGV ~~T~~ YSNR ~~C~~ FFCNAYVQ-----SNRTNLNLVSM ~~A~~ AC
DOMÍNIO 6	DCTQYLSNTQN-GEAITAC ~~P~~ FI-----LQEVCGTDGV ~~T~~ YSND ~~C~~ SLCAHNIE-----LGT ~~S~~ VAKKHDGR ~~C~~
FOLISTATINA	
MUS MUSCULUS	
DOMÍNIO 1	ICPE-----PSSS-----EQYLCGNDGV ~~T~~ YSSA ~~C~~ HLRKATCL-----LGRSIGLAYEGK ~~C~~
DOMÍNIO 2	RCSLCDE-----LCPDSKS-----DEPV ~~C~~ ASDNATYASE ~~C~~ AMKEAACS-----SGV ~~L~~ LEVKHSGS ~~C~~
AGRINA	
GALLUS GALLUS	
DOMÍNIO 1	ECQQ-----VCQGR-----YDPVCGSDNRTYGNP ~~C~~ ELNAMACV-----LKREIRVKHKGFC
DOMÍNIO 2	QC-----VCPRCEQQP-----LAQVCGTDGLTYDNRC ~~E~~ LRAASCQ-----QQKSIEVAKMGFC
DOMÍNIO 3	VCPT-----ECVPS-----SQPVCGTDGNTYGSE ~~C~~ ELHVR ~~A~~ CT-----QQKNILVAAQGD ~~C~~
DOMÍNIO 5	EE-----NCPSK-----REPVC ~~G~~ DDGV ~~T~~ YASE ~~C~~ VMGRTGAI-----RGLEIQKVRSGQC
DOMÍNIO 6	HC-----ECPSPLCSEAN---MTKVC ~~G~~ SDGV ~~T~~ YGDQC ~~O~~ LKT ~~I~~ ACR-----QGQLITVKHVGQC
DOMÍNIO 7	SCSGVA-----ESIVCGSDGKDYRSE ~~C~~ DLNKHACD-----KQENVFKKFDGAC
DOMÍNIO 8	SCDR-----ITCDGT-----YRPVCARDSRTYSND ~~C~~ ERQKAECH-----QKAAIPVKHSGFC
OSTEONECTIN	
GALLUS GALLUS	VCQDPS-----SCPAHSGV-----FEKVCGTDNKTYDSS ~~C~~ HFFATKCTLEGTK---KGHKLHLDYIGFC
RODNINA	
RHODNIUS PROLIXUS	CQE-----CD-----GDEYKPVCGSDDITYDNN ~~C~~ RL---ECA-----SISSSPGV ~~E~~ LKHEGFC

Figura 3-22.: alinhamento entre algumas seqüências de membros da família Kazal selecionados. As meias cistinas conservadas estão destacadas em preto. Note que a grande maioria das proteínas da família é composta de vários domínios do tipo Kazal.

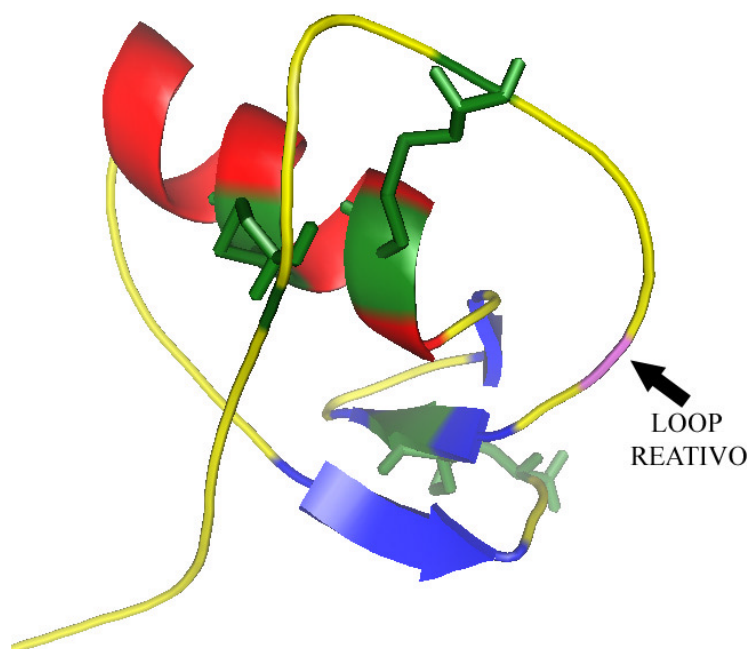


Figura 3-23: representação da estrutura tridimensional do inibidor de tripsina humano do tipo Kazal (código PDB 1HPT). As pontes dissulfeto são mostradas em verde.

6.11. FAMÍLIA DOS INIBIDORES DE SUBTILISINA DE STREPTOMYCES (SSI)

A família *Streptomyces* de bactérias produz uma série de inibidores, os quais são caracterizados por sua forte atividade contra subtilisina. Eles são coletivamente conhecidos como SSI ou “*Streptomyces* Subtilisin Inhibitors”. Alguns SSI também inibem tripsina e quimotripsina. Além disso, o inibidor extraído de *Streptomyces griseus* inibe também a metaloendopeptidase dessa bactéria e verificou-se que os sítios de inibição para serinoproteinases e para a metaloproteinase coincidem. (KUMAZAKI *et al.*, 1993)

SSI é um homodímero com cada monômero contendo duas folhas- β s antiparalelas e duas α -hélices curtas. A associação com a proteinase induz o alargamento de uma estrutura que forma uma espécie de canal no qual as cadeias laterais hidrofóbicas estão sanduichadas entre dois lóbulos (HIRONO *et al.*, 1984). A perda do tetrapeptídeo C-terminal reduz drasticamente o efeito inibitório das proteínas quando menos de uma molécula de inibidor está presente por molécula de enzima, implicando que o tetrapeptídeo é necessário para manter o arranjo 3D correto (SAKAY *et al.*, 1980).

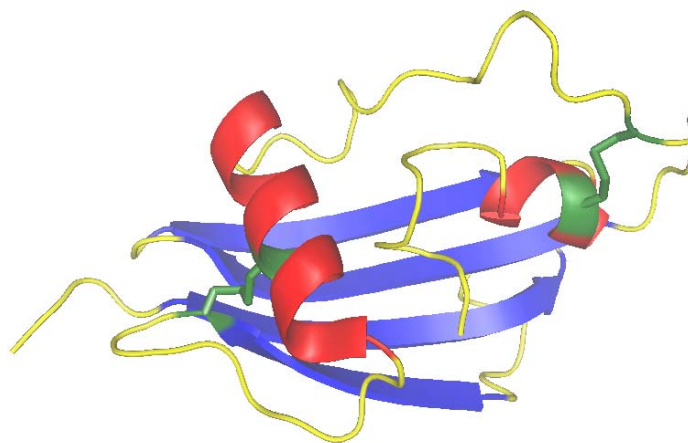


Figura 3-24: representação da estrutura 3D do inibidor de subtilisin de *Streptomyces albogriseolus* da família SSI (código PDB 3SSI.). As pontes dissulfeto estão em verde .

As similaridades estruturais entre SSI e os inibidores ovomucóides e PSTI sugerem a evolução das duas subfamílias a partir de um ancestral comum (HIRONO *et al.*, 1984).

	10	20	30
1. Streptovercillium orinoci	LYAPSALVLTIGQGD-SASAGIQRAVTLSC		
2. Streptovercillium luteovercicillatus	-YAPSALVLTIGQGDSAAATAGVQRAVTLTC		
3. Streptovercillium eurocidicus	-YAPSALVLTIGQGATAAESGVQRAVTLTC		
4. Streptovercillium netropsis	LFAPSALVLTVGEGESAADSGVQRAVTLTC		
5.+Streptomyces albogriseolus	LYAPSALVLTVGKGVSAATTAAPERAVTLTC		
6. Streptomyces antifibrinolyticus	LYAPSALVLTMGHGNSAATVNPERAVTLNC		
7. Streptomyces lavendulae	LYAPSALVLTIGHGGAATATPERAVTLTC		
8. Streptomyces longisporus	LYAPSALVLTVGHGTSAAAASPLRAVTLNC		
9. Streptomyces rochei	LYAPSALVLTIGQGESAAATSPRAVTLTC		
10. Streptomyces cacaoui	LYAPSAVVISKTQGA-SADAPAQRAVTLRC		
11. Streptomyces virginiae	LYAPSAMVFSVAQGDDVAAPTIVRATTVSC		

	40	50	60	70	80	90
1. MPTPSGTHPDARDACAQLRQADGKFDLTATKA--GT	CTKEWNPVTVTATGVWEGQRVNYSHTFGNPC					
2. TPKAAGSHPTSGACAQLRLSNGDFDKLVKIKD--GT	MCTREWNPTSTVTAEGVWEGRRVSFEHTFANPC					
3. TPKSSGTHPDAGACTQLRAAGGDFDKVTRIKS--DT	VTCTKEWNPVTVTAEVWDGRRISYEHTFANPC					
4. TPKASGTHPAARAACDQLRAVDGDFKALVTTKS--DR	VTCTKEYRPIVITAEGVWDGHRVSYEHKFANPC					
5. APGPSGTHPAAGSACADLAAGVGDNLALTRGE--D	VMCPMVYDVPVLLTVDGVMQGRVSYERVFSNEC					
6. APTASGTHPAALQACAEELRGAGGDFDALTVRG--D	VACTKQFDPVVVTVDGVMQGRVSYERTFANEC					
7. APTSSGTHPAASAACAELRGVGGDFALKARD---D	VWCKNLYDPVVVTAQGVWQGRVSYERTFGNSC					
8. APTASGTHPAPALACADLRGVGGDIDALKARD--G	VICNLYDPVVVTVDGVMQGRVSYERTFGNEC					
9. APKATGTHPAADAACAELRRAGGDFDALSAAD--G	VMCTREYAPVVVTVDGVMQGRRLSYERTFANEC					
10. LPVG-GDHPAPEKACAALREAGGDPALPRYVEDTGR	VCTREYRPVTVSVQGVWDGRRIDHAQTFNSC					
11. APGARGTHPDKAACAALKSTGGAFDRLLEPNP-DR	ACPMHYAPVTVSAVGWEGRRVAWDHTFANSC					

Figura 3-25: alinhamento entre algumas seqüências de membros da família SSI selecionados. As meias cistinas conservadas estão destacadas em preto.

6.12. FAMÍLIA ANTISTASINAS

Apesar de extremamente pequenos, os inibidores de tripsina dessa família possuem um conteúdo de meias cistinas bastante alto o que lhes confere uma estrutura rígida e resistente. O padrão de resíduos de cisteínas é bastante conservado sendo considerado como uma assinatura dessa família de proteínas. (LAPATTO *et al.*, 1997)

		10	20	30
Haementeria officinalis	CSGV-R	CRMH---	CPHGFQSR	SGCE-FCKC
HUMANO	CGELSN	CTLTGKDC	INGFKRDH	NGCR-TCQC
Gallus gallus	CELLVN	CTLTEKDC	CIYSFKLD	QNGCR-ICQC
Mus musculus	CSLK---	EEKD---	CVYGFKLD	HNGCR-TCQC
Hirudo medicinalis	CSDL-H	CKVK---	CEHGFKKD	DNGCEYACIC
Theromyzon tessulatum	CNDA-Q	CRIY---	CPLGFTTD	ANGCESFCIC
Hirudo nipponia	CVIG-Q	CRKY---	CPNGFKKD	ENGCTFPIC

Figura 3-26: alinhamento entre algumas seqüências do domínio inibitório de alguns membros da família das antistasinas selecionados. As meias cistinas conservadas estão destacadas em preto.

Esse domínio também é encontrado em múltiplas cópias em algumas proteínas extracelulares grandes. Muitas das proteínas que pertencem à essa família são anticoagulantes. Como mostrado na figura a seguir, os inibidores são encontrados como monômeros sendo ativos como dímeros.

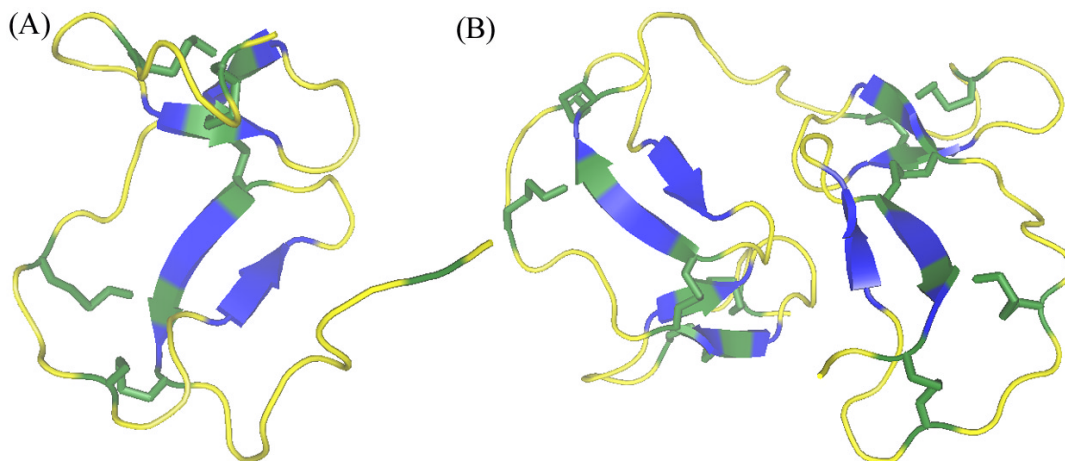


Figura 3-27: representação da estrutura tridimensional do inibidor da família das antistasinas extraído da saliva de sanguessuga mexicana *Haementeria officinalis* (código PDB 1SKZ). As pontes dissulfeto estão em verde. Em (A) está o monômero e em (B) o dímero ativo.

6.13. FAMÍLIA GRASSHOPPER

Também é conhecida como família da pacifastina, que é uma proteína de 155kDa encontrada no plasma do peixe *Pacifastacus levisculus*, a qual inibe quimotripsina, tripsina e uma proteinase hemolinfática desse peixe (LIANG *et al.*, 1997).

Os membros dessa família são inibidores compostos de duas subunidades covalentemente ligadas. Uma cadeia leve com 44kDa rica em cisteínas que contém a subunidade inibitória. E uma cadeia pesada com 105kDa que é estruturalmente e

funcionalmente relacionada à transferrina, contendo três lóbulos transferrina dois dos quais são sítios de ligação de ferro (LIANG *et al.*, 1997)

PACIFASTACUS LENIUSCULUS	
DOMÍNIO 1	ACTPGSRKY-DG ENW C IC SSGG-A-W--I CT LKY CP SSGGGLTF
DOMÍNIO 2	VCSEGSRWKADD ENW C IC IDG--S-P--S CT KRL CR TKLAKGMFA
DOMÍNIO 3	QCVPGSRWK-KD ENW C IC TETA-I-G--M CT LIG CL NYEPKPGEA
DOMÍNIO 4	VCTDGSKWK-DD ENW C IC NNG--S-A--S CT EKL CQ YKPDGSLPD
DOMÍNIO 5	MCVPFSRWK-DE ENW C IC EANG-A-A--P CT RM GC SEDYKPPGE
DOMÍNIO 6	VCIDGSRWK-VD ENW C IC NNG--S-S--A CT EKL CL KPGGQCTEG
SCHISTOCERCA GREGÁRIA	
DOMÍNIO 1	ECTPGDTKK-ED NT C IC CTPTG-V-W--V CT RKG CV TKREVNCTP
DOMÍNIO 2	HCTPNTTFK-KD NT C IC SCNRDGT-A-A--V CT LK ACL SRSKREVSC
LOCUSTA MIGRATORIA	
DOMÍNIO 1	ECTPGQTKK-QD NT C IC CTPTG-I-W--G CT RK AC RTTREAEEPA
DOMÍNIO 2	QCTPNKSFK-KD NT C IC CKNDGT-A-A--I CT Q IA CLNRGRQVNC
PIMPLA HYPOCHONDRIACA	
DOMÍNIO 1	PCTPGENFK-YY ND C IC CLDGLRAHA--M CT RM CR DRNVFNEDGT
DOMÍNIO 2	SCAPGASFK-YY NS C IC CGAEGKV-AEAQ CT SQ EC DRYKWKDGS

Figura 3-28: alinhamento entre algumas seqüências do domínio inibitório de membros da família da pacifastina selecionados. As meias cistinas conservadas estão destacadas em preto. Apenas a cadeia leve com a subunidade inibitória é mostrada.

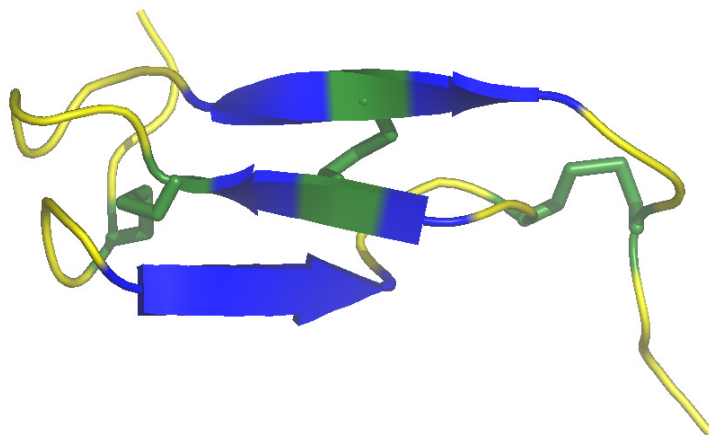


Figura 3-29: representação da estrutura 3D da cadeia leve da pacifastina de Locusta migratoria (código PDB 1PMC). As pontes dissulfeto estão em verde.

6.14. FAMÍLIA CHELONIANINA

Essa família é composta de um grupo de proteínas contendo oito resíduos de cisteínas caracteristicamente espaçados, os quais estão envolvidos em pontes dissulfeto. Embora o padrão das cisteínas seja conservado, o grau de similaridade global das seqüências é baixo com poucas prolínas e glicinas sendo razoavelmente bem conservadas (HENNIGHAUSEN & SIPPEL, 1982).

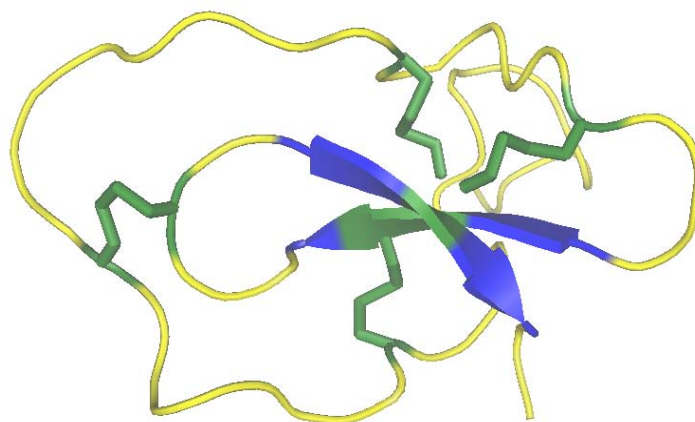


Figura 3-30: representação da estrutura 3D da elafina humana (código PDB 2REL) que possui o arranjo característico da família chelonianina. As pontes dissulfeto estão em verde.

WAP		
MUS MUSCULUS		
DOMÍNIO 1	VQSMFQKAS....PIEGTECII...CQTNEECAQNAMCCPGSCGRTRKTP	
DOMÍNIO 1	VQSMFQKAS.....PIEGTECII---CQTNEECAQNAMCCPGSCGRTRKTP	
DOMÍNIO 2	KAGRCF..WNLLQTI..SSTGPCPMKIE..CSSDRECSGNMRCCNVDCVMTCTTP	
RATTUS NORVEGICUS		
DOMÍNIO 1	VQSMCSDDS.....FSEDETCIN...CQTNEECAQNDMCCPSSCGRSCKTP	
DOMÍNIO 2	KAGRCF..WNPIQMI..A.AGPCPKDNP..CSIDSDCSGTMRCCKNGCIMSMDP	
ORYCTOLAGUS CUNICULUS		
DOMÍNIO 1	VQVMCP.....EPSSSEETL..CLSDNDCLGSTVCCPSAAGGSCRTTP	
DOMÍNIO 2	KAGRCF..WVQAPML..S..QLCEELSD..CANDIECRGDKKCCFSRCAMRILEP	
SUS SCROFA		
DOMÍNIO 1	KAGRCF..WVPAPLA..P..ELCLEKNE..CSRDDQCRGNKKCCFSSCAMRCLDP	
DOMÍNIO 2	KPGVCP.....KII..F.CPLVNPPIK..CWRDSHCPGVKKCCPSLCKKGCVTP	
ELAFINA		
HUMANA	KPGSCF....I.II.L.IRCAMLNPPNRCLKDTDCPGIKKCCCGSGCMACFVP	
SUS SCROFA	KPGSCPRIL.....IRCLMVNPPNR..CLSDAQCPGLKKCCCGSGCMACMDP	
WDNM		
RATTUS NORVEGICUS	KPGKCPKNP.....PRSIGTCVEL--..CSGDQSCPNIQKCCSNGCGHVCKSP	
MUS MUSCULUS	KPGACPKPP.....PRSFGTCDER--..CTGDGSCSGNMKCCSNGCGHACKPP	

Figura 3-31: alinhamento entre algumas seqüências selecionadas das diversas sub-famílias em que a família chelonianina é dividida. As meias cistinas conservadas estão destacadas em preto. Note que algumas proteínas são compostas de vários domínios do tipo chelonianina.

O grupo de seqüências que dividem esse padrão de estrutura inclui as proteínas ácidas que foram encontradas na parte líquida do leite coalhado, chamadas WAP (de *whey acidic proteins*) (HENNIGHAUSEN & SIPPEL, 1982), as elafinas que são homólogas a um inibidor elastase-específico encontrado na pele humana (WIEDOW *et al.*, 1990), as chamadas proteínas WDNM1 as quais estão envolvidas no potencial de metástase de adenocarcinomas em ratos (DEAR *et al.*, 1988), proteínas envolvidas com a síndrome de

Kallmann (LEGOUIS *et al.*, 1991), e as proteínas homólogas a caltrina II extraída de porco a qual inibe o transporte de cálcio em spermatozoa (CORONEL *et al.*, 1990).

6.15. FAMÍLIA ASCARIS

Os inibidores pertencentes à essa família possuem uma estrutura característica conhecida como domínio rico em cisteínas que é encontrado em muitas proteínas extracelulares. O domínio contém tipicamente dez meias-cistinas que formam cinco pontes dissulfeto conservadas numa estrutura composta por 50–60 aminoácidos. Esse alto conteúdo de pontes dissulfeto confere robustez ao inibidor que é extremamente resistente frente a altas temperaturas e variações extremas de pH bem como a proteólise por várias enzimas. (GRASBERGER *et al.*, 1994)

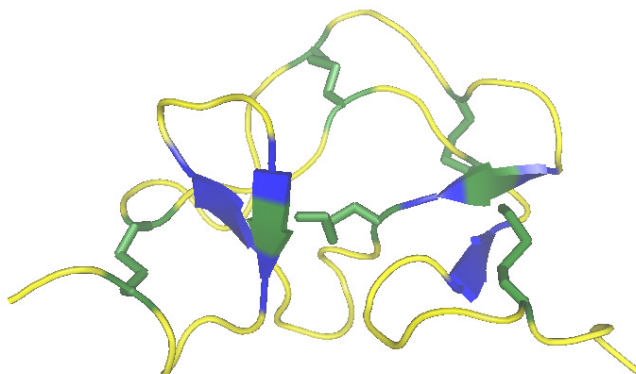


Figura 3-32: representação da estrutura tridimensional do inibidor da família ascaris extraído de *Ascaris lumbricoides* (código PDB 1ATA). As pontes dissulfeto estão em verde.

	10	20	30	40	50
<i>Apis mellifera</i>	CGP-NEVFNT	CGSACAPT	CAQPKTRIC	TMQCRIG-CCCQ--EG-FLRNGEGA	VLPENC
<i>Ascaris suum</i>	CTKPNEQWTK	CGG-CEGTCAQK-IVPCTRECKPPRC	CECIASAG-FVRDAQGN	CIKFEDC	
<i>Anisakis simplex</i>	CGP-NEQWSG	CPK-CELOS	SESD-KPCATICGEPKCYCS--	PDKYRRI	PDGRIRKIQ

Figura 3-33: alinhamento de três inibidores da família ascaris. As dez meias cistinas conservadas estão destacadas em preto.

6.16. FAMÍLIA ECOTINA

A ecotina é um inibidor de serinoproteínases encontrado no periplasma de *Escherichia coli*. Um dos seus aspectos característicos é que ela é um potente inibidor de

uma variedade de serinoproteínases com especificidades bem diferentes (CHUNG *et al.*, 1983; MCGRATH *et al.*, 1995).

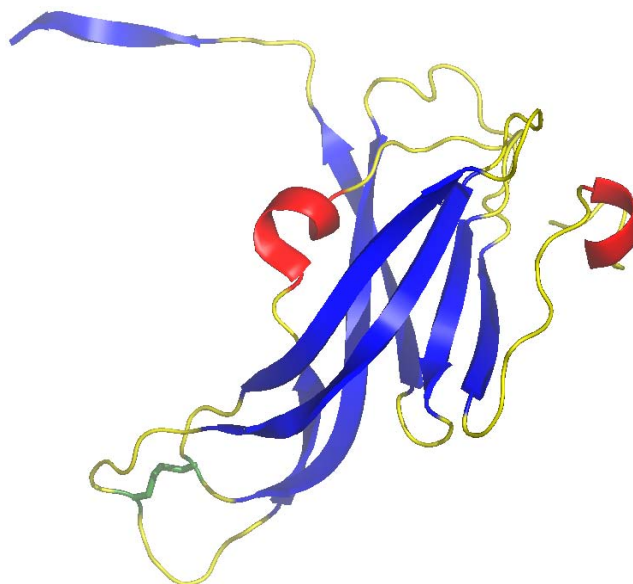


Figura 3-34: representação da estrutura 3D da ecotina de *Escherichia coli* (código PDB 1ECY). A única ponte dissulfeto é mostrada em verde.

Os membros da família são proteínas homodiméricas. As duas subunidades estão unidas por suas longas fitas- β s que são arranjadas em duas folhas- β s antiparalelas. Um dímero de ecotina pode ligar duas moléculas de proteinase, cada uma delas ligada a uma subunidade de ecotina simultaneamente através de dois diferentes sítios. Entretanto o sítio secundário é muito pequeno e ainda não há evidências suficientes para afirmar se se trata de um sítio de reconhecimento real ou uma superfície de contato que foi confundida (YANG *et al.*, 1998).

	10	20	30	40	50	60	70
1. <i>Pseudomonas aeruginosa</i>	MKALLIAAGVAALSSTAMAAK	-----	LDEKVPYPKADAGFTRQVIHL	PKQDAEDAFKVEII	IAGKTLEA		
2. <i>Pseudomonas putida</i>	MTAIL-ALSLAA-AAPAMAAS	-----	LKDIAPYPEAEKGFTRQVIHL	PAQADESAYKLEIL	LAGKTLQV		
3. <i>Rickettsia conorii</i>	-----						
4. <i>Salmonella typhimurium</i>	MKMFVPAVVFAALASASAWANN	GDTAQPLEKIAPYPQAEKGMKRQVITL	TPQQDESTLKV	ELLIGQTLNV			
5. <i>Salmonella typhi</i>	MKMFVPAVVFAALASASAWANN	GDTAQPLEKIAPYPQAEKGMKRQVITL	TPQQDESTLKV	ELLIGQTLNV			
6. <i>Escherichia coli</i>	MKTILPAVLFAAFATTSAWAA	--ESVQPLEKIAPYPQAEKGMKRQVI	QLTPQDESTLKV	ELLIGQTLNV			
7. <i>Yersinia pestis</i>	LASVLLATSINAIADTPTPLN	---QQQPLEKIAPYPQAEKGMSRQVIF	LEPQKDES	RFKVELLIGQTLNV			
8. <i>Shewanella oneidensis</i>	-----PTGLDAPMISVSSMNA	---NNYAPVETVKMFPA	PKKGMVQHILTL	PKLENETDYMVEIQIGQTLV			

	80	90	100	110	120	130	140	150	160
1. <i>DCNQRLGGLEEHTLEGWGY</i>	SYRRLDKVSG-PMSTM	ACPGQKKEQ-RFIPV-V	EGGFLLCYNSKLP	IVVYAPKDVEVRYR	ISASEKVEKAV				
2. <i>DCNQRLGGNLEAHTLEGWGY</i>	NYRDLNVSG-PASTL	MACPDGKKTE-AFVPV-V	GDGFLLRYSKLP	IVVYVPKDV	EVRYRVSASQDVQKAK				
3. <i>DCNRVWFGGKLETKTLEGWGY</i>	NYIIDQVSDHPASTM	MACPNVKATTI-QTVS	VFLGD-----						
4. <i>DCNQHRLGGTLETKTLEGWGY</i>	DYVFDNVTSPVSTM	MACPEGKKEQ-KFVT	AWLGEDGMLRYSKLP	IVVYTPANVDVKYRI	WKADANVQNAV				
5. <i>DCNQHRLGGTLETKTLEGWGY</i>	DYVFDNVTSPVSTM	MACPEGKKEQ-KFVT	AWLGEDGMLRYSKLP	IVVYTPANVDVKYRI	WKADANVQNAI				
6. <i>DCNLHRLGGKLENKTLEGWGY</i>	DYVFDKVSSPVSTM	MACPDGKKEK-KFVT	AYLGDAFMLRYSKLP	IVVYTPDNVDVKYRV	WKAEKIDNAV				
7. <i>DCNRHMLGGNLETRTL</i>	SGWGFYLVMDKISQ-P	ASTM	MACPEDSKPQVKFVT	ANLGDAAMQRYNSR	LPIVVYVPGQVEVKYRI	WEAGEDIRSAQ			
8. <i>DCNKHGLNGQLKELT</i>	VEGWGYNYQVDEISE-G	PSTM	MACFELAKKE-AFVQI--	PDELTLRYDSRLPKV	FYLP	PEGAELRFRTWKADSTYQYSK			

Figura 3-35: alinhamento entre algumas seqüências de membros da família ecotina selecionados. A única ponte dissulfeto está destacada em preto.

CAPÍTULO 4

FUNÇÕES E APLICAÇÕES DOS INIBIDORES DE PROTEINASES

1. FUNÇÃO DOS INIBIDORES DE PROTEINASES NOS VEGETAIS

Muitas pesquisas vêm sendo realizadas, buscando determinar o papel biológico dos inibidores de proteinases em plantas e hoje se sabe que os inibidores exercem uma multiplicidade de papéis. Além do seu envolvimento nos processos de regulação das atividades proteolíticas das enzimas durante a germinação (RICHARDSON, 1991), existem outras funções biológicas atribuídas aos inibidores, como a sua atuação como proteínas de reserva (RYAN, 1981), participação nos mecanismos de defesa das plantas ao ataque de insetos (GATEHOUSE, 1984; MACEDO *et al.*, 1994), e o seu papel antifúngico e antimicrobiano (CHRISTELLER *et al.*, 1992; HUYNH *et al.*, 1992; BLANCOLABRA *et al.*, 1995).

1.1. PROTEÍNAS DE RESERVA

Muitos aspectos dos inibidores de proteinases encontrados em sementes sugerem fortemente que eles podem ter um papel como proteínas de reserva as quais são imunes à digestão até serem requeridas durante a germinação. As proteínas que desempenham esse papel são, na sua maioria, particularmente resistentes a desnaturação por calor, extremos de

pH e ao ataque de enzimas proteolíticas, ou seja, são proteínas robustas. As características estruturais, mencionadas tantas vezes aqui, que conferem aos inibidores uma forma compacta, rígida e estável depõem a favor do seu papel como proteínas de reserva. Além disso, o alto conteúdo de pontes dissulfeto encontradas particularmente nos inibidores extraídos de plantas apontam para o seu valor como reservas de enxofre para as plantas. (RYAN, 1981)

Existem estudos mostrando que a biossíntese dos inibidores durante o desenvolvimento acontece em paralelo com a de algumas proteínas de reserva já conhecidas. Além da sua presença em sementes, esses inibidores também estão presentes nas paredes celulares, espaços intercelulares e no citosol. (RYAN, 1981)

Qualquer inibidor que está presente em sementes dormentes em altas concentrações e é subsequente degradado durante a germinação é de fato uma proteína de reserva. Muitos inibidores apresentam esse comportamento. Entretanto, em alguns casos a extensão da degradação é limitada e em outros existe mesmo um aumento na contração do inibidor durante a germinação. Isso acontece porque alguns inibidores atuam controlando a atividade proteolítica das enzimas durante a germinação. (RICHARDSON, 1991)

1.2. CONTROLE DE ENZIMAS ENDÓGENAS NAS SEMENTES

Durante as últimas décadas, muitos estudos têm dado respaldo ao conceito que os inibidores de proteinases presentes em sementes atuam controlando a atividade de enzimas endógenas como, por exemplo, as α -amilases, prevenindo a hidrólise prematura dos materiais de reserva. Existem, como já vimos, numerosos inibidores de α -amilases endógenas, as quais são responsáveis por iniciar a mobilização das proteínas de reserva durante a germinação. A superfamília cereal, por exemplo, possui inibidores de α -amilase extraídos de trigo, milho, cevada, arroz, sorgo, gergelim, etc. (RICHARDSON, 1991)

Além da ação das serinoproteinases, a mobilização de proteínas de reserva durante a germinação é também devida à ação de proteinases cisteínicas e inibidores para essa classe de enzimas e outras proteinases endógenas têm sido descobertos. (RICHARDSON, 1991)

1.3. DEFESA CONTRA PESTES E PREDADORES

As plantas possuem diferentes maneiras de se defender do ataque de pestes e predadores. A grande variedade dos diferentes inibidores encontrados nas plantas parece se adequar admiravelmente a esse papel. Por exemplo, alguns inibidores podem fazer com que as sementes sejam menos agradáveis ao paladar, até mesmo letais, para os insetos predadores os quais podem atacá-las durante períodos de dormência ou quando armazenadas. Também durante esses períodos, eles podem inibir enzimas extracelulares, retardando ou prevenindo a invasão de tecidos por microorganismo saprófitos e parasitas como fungos e bactérias. Além disso, não se pode esquecer que muitas plantas dependem de animais para dispersar as suas sementes e essa dispersão freqüentemente envolve a passagem da semente através do aparelho digestivo do animal. A presença de inibidores pode ser, em alguns casos, a razão principal porque tantas sementes permanecem intactas durante esse trajeto. (LAWRENCE & KOUNDAL, 2002)

Muitas observações reforçam esse papel dos inibidores no mecanismo de defesa contra pestes. Como vimos na família taumatococina, o inibidor bifuncional de α -amilase/tripsina de milho apresenta forte homologia seqüencial com as proteínas induzidas por vírus relacionadas com patogêneses – PR proteínas – e com outras proteínas induzidas por estresse. (LAWRENCE & KOUNDAL, 2002)

A ação dos inibidores nesses casos está intimamente ligada às suas propriedades antinutricionais. Em um dos mecanismos propostos, eles atuam reduzindo a digestão das proteínas da dieta dos insetos pela inibição direta das enzimas digestivas desses predadores, que morreriam de inanição. Estudos mostram que se adicionando inibidores puros na ração artificial reduz-se significativamente o crescimento e desenvolvimento de vários insetos e pragas. Mas outros estudos posteriores mostraram que a ingestão de inibidores causa uma significativa elevação dos níveis de tripsina no estômago dos insetos e, por isso, foi proposto que o verdadeiro modo de ação dos inibidores seria causar a hiperprodução perniciosa de tripsina, o que levaria o inseto à morte por danos causados ao aparelho digestivo numa espécie de autodigestão. (LAWRENCE & KOUNDAL, 2002)

Um aspecto interessante do desenvolvimento dessa área de pesquisa foi o estudo demonstrando que, quando um gene codificando um inibidor do tipo Bowman-Birk foi tomado de sementes de ervilhas e geneticamente engeheirado em tabaco, ele conferiu ao

tabaco resistência a *Heliothis virescens*, um dos insetos herbívoros que é praga da ervilha (HILDER et al, 1987). Desde então muitas plantas geneticamente modificadas foram construídas a fim de se obter plantas mais resistentes às pragas, o que tem contribuído sensivelmente para avanços e melhorias para a agricultura.

Além de participar na defesa das plantas contra insetos e outros predadores, os inibidores também participam na defesa contra microorganismos como fungos e vírus. Em 1976, PENG & BLACK mostraram que os níveis de inibidores de proteinases em plantas de tomate aumentam logo após a infecção pelo fungo patogênico *Phytophthora infestans*, mas essa resposta só é observada nas variedades de tomate resistentes ao fungo e não em variedades susceptíveis. Similarmente, BRUENING & SANDERSON em 1984 observaram que um inibidor de proteinases cisteínicas em algumas variedades de ervilha, as quais são resistentes ao vírus mosaico da ervilha, previne a atividade da enzima codificada pelo vírus que desempenha um papel no processo normal de infecção.

Outro possível modo de defesa contra microorganismos seria a secreção de alguns inibidores de proteinases que as sementes de leguminosas fazem no meio externo durante a germinação. Isso seria a fim de inibir a ação microbiana na rizosfera das plantas pela inibição de microorganismo patogênicos e, possivelmente, pela estimulação de simbiontes benéficos (TAN-WILSON et al., 1982).

2. USOS MEDICINAIS DOS INIBIDORES DE PROTEINASES

Há um interesse muito grande no possível uso de inibidores de origem vegetal que atuam sobre proteinases no tratamento de uma grande variedade de distúrbios metabólicos associados às enzimas proteolíticas, como, por exemplo, pancreatites, enfisemas, alergias, inflamações e hipertensão (RICHARDSON, 1991).

Experimentos *in vivo* mostraram que inibidores de proteinases de alguns tipos, especialmente os da família Bowman-Birk e Kunitz, são anticarcinogênicos (KENNEDY, 1998). As propriedades anticarcinogênicas incluem a habilidade para reduzir a formação de radicais de oxigênio (FRENKEL et al., 1987), suprimir o crescimento de tumores anais e de cólon induzidos quimicamente (BILLINGS et al., 1990), tumor de mama em ratos e em humanos (TAMIR et al., 1990), e tumor de pulmão em camundongos (WITSCHI &

KENNEDY, 1989). Incluem também a habilidade para suprimir as transformações celulares induzidas quimicamente ou por radiação (BILLINGS *et al.*, 1989).

Além disso, amplos estudos epidemiológicos verificaram que a ocorrência de câncer em humanos está ligada à dieta, sendo que, em populações que ingerem alimentos com alta concentração de inibidores, especialmente da família Bowman-Birk, a proporção de casos de câncer de cólon, mama, próstata e pele é muito menor em relação à de outras populações (KENNEDY, 1998).

Pesquisas envolvendo inibidores de tripsina tipo Kunitz e tipo Bowman-Birk da soja e inibidores de papaína indicaram que os mesmos apresentam um grande potencial para reduzir a frequência de mudanças cromossômicas de células de pacientes com síndrome de floração (KENNEDY, 1994).

3. USOS LABORATORIAIS DOS INIBIDORES DE PROTEINASES

Hoje se tem uma grande quantidade de inibidores que estão disponíveis para o uso laboratorial. É o caso do inibidor de α -amilase de trigo que é usado atualmente em kits comerciais para exames laboratoriais para diagnosticar hiperamilasemia (RICHARSON, 1991).

Os inibidores de proteinases, na sua maioria convenientemente disponíveis em grandes quantidades a custos relativamente baixos, têm providenciado uma ferramenta muito útil para a purificação por cromatografia de afinidade de uma forma simples, freqüentemente em um passo, das suas enzimas inibidas. Diferentes inibidores imobilizados têm sido utilizados para a purificação por afinidade de proteinases de microorganismos e animais, bem como para a purificação de α -amilases de uma variedade de animais e microorganismos.

PARTE

II

CRISTALOGRAFIA DE PROTEÍNAS

CAPÍTULO 5

FUNDAMENTOS DA CRISTALOGRAFIA DE PROTEÍNAS

1. INTRODUÇÃO

Quando queremos determinar a forma de objetos a nossa volta, basta simplesmente olharmos para eles. Se eles são muito pequenos basta usar um microscópio.

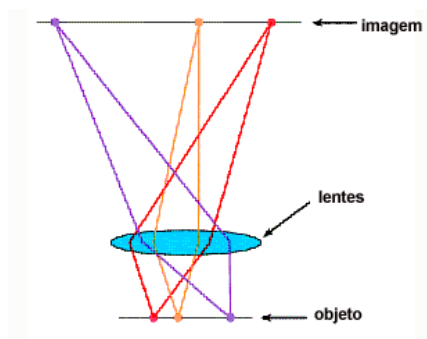


Figura 5-1: A formação da imagem de um objeto num microscópio óptico.

Entretanto existe um limite para o quão pequeno um objeto pode ser para que possa ser visto num microscópio ótico. Esse limite, o *limite de difração*, é dado pelo comprimento de onda da luz usada pois não é possível formar a imagem de um objeto muito menor do que esse comprimento de onda. O comprimento de onda para a luz visível é da ordem de centenas de nanômetros enquanto átomos são separados por distâncias da ordem de 0,1 nanômetro ou 1 Ângstron. Consultando-se o espectro eletromagnético verifica-se que os

raios-x se encontram na faixa adequada de comprimento de onda. No entanto não existem microscópios de raios-x para se observar moléculas, simplesmente porque não existem lentes para raios-x – talvez porque elas teriam de ser construídas com tolerâncias significativamente menores que a distância entre dois átomos! Esse obstáculo pode ser superado simulando-se uma lente para raios-x em computador: os raios atingem o objeto, são difratados em várias direções e então a lente coleta esses raios difratados e os rearranja para formar a imagem.

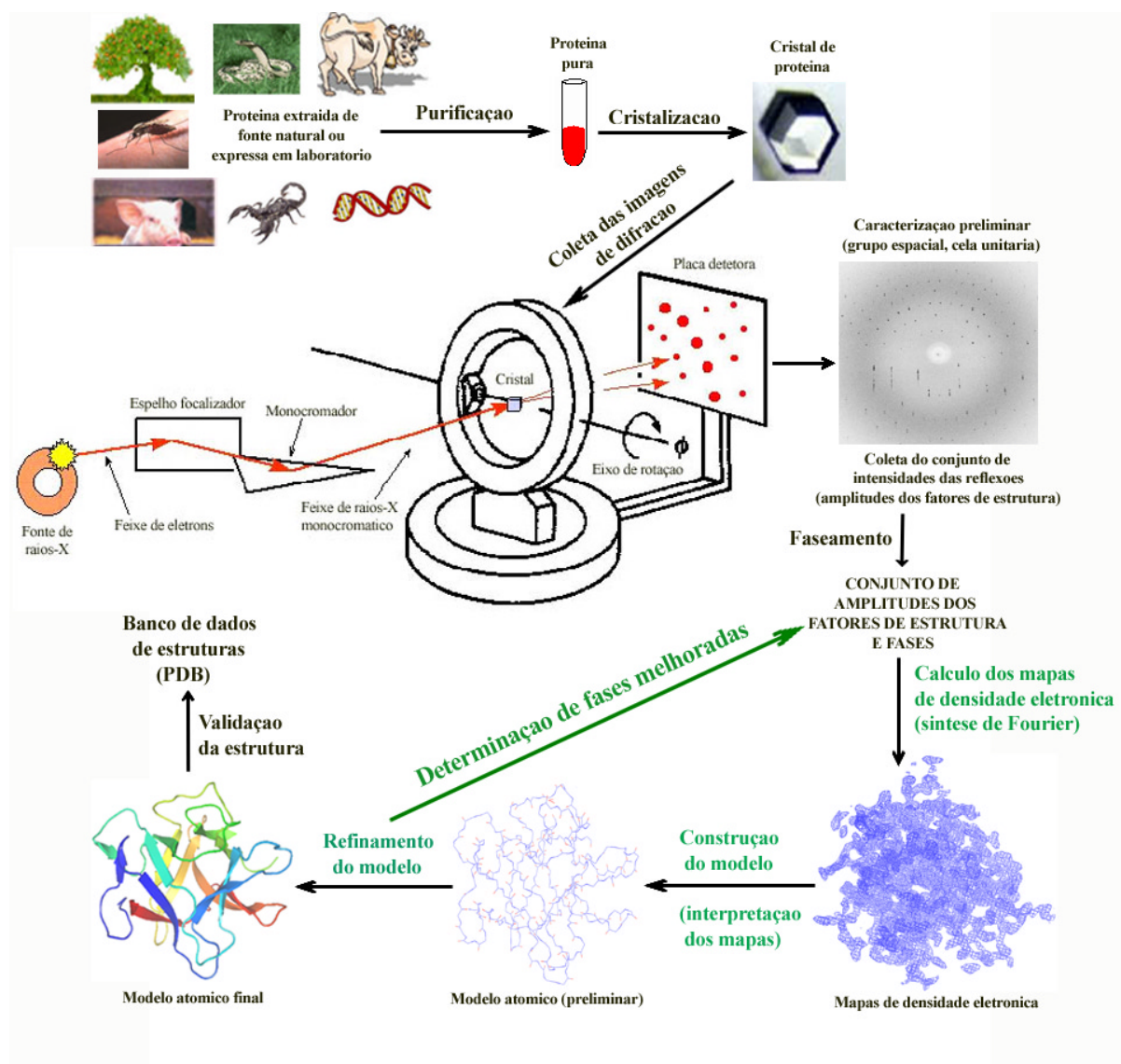


Figura 5-2: Fluxograma com as principais etapas envolvidas na determinação da estrutura cristalográfica de uma macromolécula biológica de interesse pela técnica de cristalografia de proteínas – uma visão panorâmica da técnica.

Com efeito, foi verificado que, a exemplo dos cristais de compostos inorgânicos, cristais de proteína também difratam raios-x. Esses raios-x difratados são coletados e formam um padrão de difração. Com o uso de computadores, esses dados são processados e usados para se gerar a *imagem* da proteína. Embora esse processo não seja tão simples assim essa é a essência do método.

Na figura 5-2 é dada uma visão panorâmica dos principais passos envolvidos na determinação da estrutura cristalográfica de uma macromolécula biológica.

2. DIFRAÇÃO: ONDAS ELETROMAGNÉTICAS

O fenômeno de difração acontece a partir da interação da matéria com as ondas eletromagnéticas da radiação envolvida, por isso é preciso se conhecer um pouco sobre o comportamento das ondas. A difração resulta da adição de ondas espalhadas por diferentes objetos, então é preciso saber como as ondas podem ser somadas e como isso está relacionado com as posições relativas dos objetos que as espalham. Para que se possa compreender o fenômeno de difração não se faz necessária uma abordagem quântica, por isso na abordagem feita aqui os raios-x serão tratados como ondas clássicas sem preocupação quanto às suas propriedades de partículas.

2.1. DESCREVENDO ONDAS

Ondas eletromagnéticas, como os raios-x, variam no tempo e no espaço. Embora elas possuam um componente magnético que varia perpendicularmente a um componente de campo elétrico, é a variação no campo elétrico que nos interessa. Isso porque os raios-x interagem com partículas carregadas da matéria, particularmente com os elétrons.

Quando se considera uma onda eletromagnética em função da posição, ao longo da direção de propagação da onda, em um determinado tempo ou em função do tempo em determinada posição, o campo elétrico apresenta uma forma cossenoidal. Considerando primeiro a onda como uma função do tempo: para simplificar, pode-se representar uma onda numa posição cujo pico ocorra no tempo zero, como na figura 5-3 a seguir.

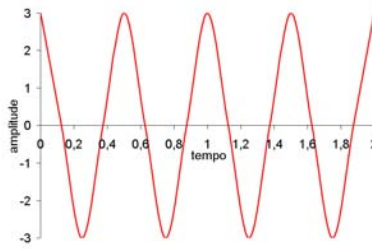


Figura 5-3: uma onda representada como uma função do tempo

Essa onda pode ser descrita pela sua amplitude – dada pela altura dos picos, que nesse caso é 3 – e pela sua frequência – ou quantas vezes ela se repete por unidade de tempo, que nesse caso é 2. Ao invés de usar a frequência, também se pode usar o período da onda, ou quanto tempo ela leva para se repetir, nesse caso metade da unidade de tempo, que é o inverso da frequência. Se a amplitude for chamada A , o tempo t e a frequência ν , então essa onda pode ser descrita pela equação 5.1:

$$A \cos(2\pi \nu t) \quad [5.1]$$

Observe que sempre que o tempo é um múltiplo do período, ou seja, quando o produto do tempo pela frequência é um inteiro, o argumento do cosseno é um múltiplo de 2π e a onda se repete.

Considerando a onda como uma função da posição ao longo do seu eixo de propagação: toma-se um corte da onda num determinado tempo. Para simplificar pode-se tomar o tempo zero e então a onda estará num pico na posição inicial, como na figura 5-4 a seguir.

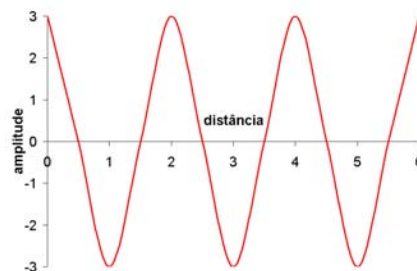


Figura 5-4: uma onda representada como uma função da distância.

Novamente esta onda tem uma amplitude 3. Ela também pode ser caracterizada pelo seu comprimento de onda, que é a distância entre dois picos, nesse caso 2. Se o

comprimento de onda for chamado λ e a distância ao longo da direção de propagação de x então a expressão que descreve a onda é dada por:

$$A \cos\left(\frac{2\pi x}{\lambda}\right) \quad [5.2]$$

Observe que sempre que x é um múltiplo do comprimento de onda, o argumento do cosseno é um múltiplo de 2π e a onda se repete. Também, e isso será importante, essa expressão pode ser representada por $A \cos(-2\pi x / \lambda)$ porque o cosseno é o mesmo para argumentos positivos ou negativos.

Como o x foi escolhido ao longo da direção de propagação da onda, daqui a um determinado tempo ela terá se movido. Por exemplo, se esperar-se por um quarto do período, ele estará na posição representada a seguir:

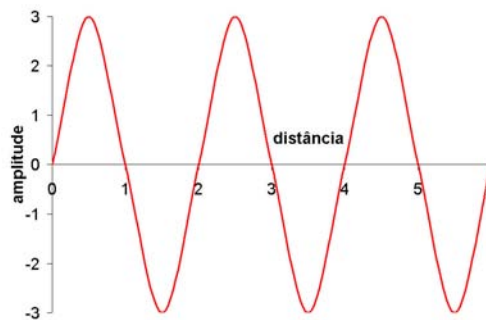


Figura 5-5: a mesma onda representada depois de transcorrido um quarto do período.

Observa-se que a amplitude caiu para zero na posição zero, como acontece após um quarto do período na figura 5-3 da onda como uma função do tempo. Os efeitos do tempo e da posição podem ser combinados numa mesma expressão para a onda. Como se vê, o efeito do tempo pode ser cancelado quando se desloca um quarto do comprimento de onda para se tomar o pico que começava na origem. Dessa forma, posição e tempo atuam em sentidos opostos na onda e seus efeitos podem ser combinados subtraindo-se o efeito da posição do efeito do tempo. Assim a expressão fica:

$$A \cos\left[2\pi\left(\frac{vt - x}{\lambda}\right)\right] \quad [5.3]$$

2.2. ONDAS COMO VETORES

O cosseno de um ângulo é o componente x de um vetor unitário após rotacioná-lo por esse ângulo ao redor de um círculo unitário. Se o vetor é rotacionado numa velocidade constante, então seu componente x traçará uma onda cossenoidal como uma função do tempo.

Quando uma onda é descrita como um vetor em um plano o comprimento do vetor representa a amplitude da onda e o ângulo que o vetor faz com o eixo horizontal representa a fase da onda. A velocidade de rotação do vetor determina o comprimento de onda da onda.

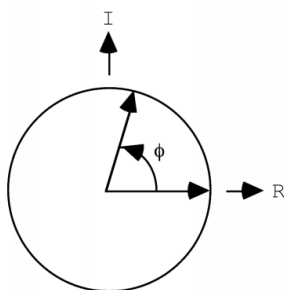


Figura 5-6: Representação vetorial de uma onda cossenoidal.

Existem grandes vantagens matemáticas em se descrever ondas como vetores. Por exemplo, quando se deseja somar duas ondas pode-se adicionar as expressões trigonométricas que as representam:

$$A \cos(\alpha + \phi_1) + B \cos(\alpha + \phi_2) \quad [5.4]$$

Ou pode-se representar as duas ondas individuais como dois vetores, um com um comprimento **A** rotado por um ângulo ϕ_1 a partir do eixo horizontal e outro com um comprimento **B** rotado por um ângulo ϕ_2 . Dessa forma basta somar os dois vetores de forma geométrica trivial.

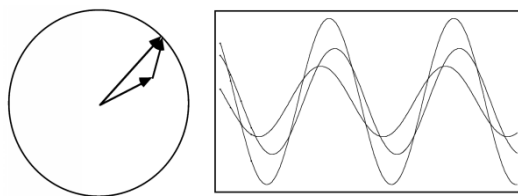


Figura 5-7: somando ondas vetorialmente.

O resultado é uma onda cossenoidal com o mesmo comprimento de onda mas com diferentes amplitude e fase as quais são determinadas pelo vetor resultante da soma.

2.3. ONDAS COMO NÚMEROS COMPLEXOS

Pode-se pensar em números complexos como vetores no plano complexo, com um componente real e um imaginário ao invés dos componentes x e y de um vetor 2D real. Os números complexos têm propriedades adicionais que os tornam mais úteis que vetores 2D reais para a representação de ondas, por isso elas são quase sempre expressas assim.

Uma das vantagens dos números complexos é que a rotação do vetor pode ser manipulada como uma multiplicação ou, de forma equivalente, adição de expoentes. Na figura 5-8 a seguir são representados alguns números complexos. O eixo horizontal representa a parte real e o eixo vertical a parte imaginária.

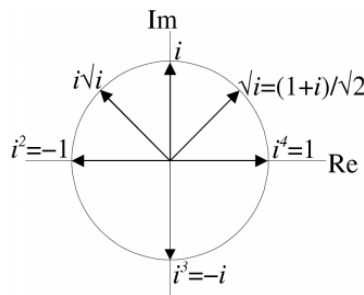


Figura 5-8: Representação gráfica de alguns números complexos.

A raiz quadrada de i é um vetor no plano complexo orientado 45 graus (ou $\pi/4$ radianos) a partir do eixo real. Se i é multiplicado pela sua raiz quadrada, o vetor é rotado de 45 graus. Assim, a multiplicação de quaisquer dois vetores unitários no plano complexo resultará num novo vetor unitário cuja orientação é dada pela soma dos ângulos de orientação desses dois vetores.

É necessário generalizar para considerar números complexos com magnitudes arbitrárias. Primeiro tem-se de considerar duas formas para representar números

complexos: como a soma de uma parte real e uma parte imaginária ou com coordenadas polares.

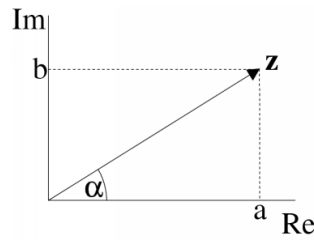


Figura 5-9: Representação gráfica do número complexo $\mathbf{Z}$ com a parte real e a imaginária.

Na figura 5-9, $\mathbf{z}$ representa o número complexo, $a=\text{Re}(\mathbf{z})$ ou a parte real de $\mathbf{z}$, $b=\text{Im}(\mathbf{z})$ ou a parte imaginária de $\mathbf{z}$, α é o ângulo entre $\mathbf{z}$ e o eixo real e $\mathbf{z}$ pode ser expresso por:

$$\mathbf{z} = a + ib = |\mathbf{z}| (\cos \alpha + i \text{sen} \alpha) \quad [5.5]$$

$$|\mathbf{z}| = (a^2 + b^2)^{1/2} = [(a + ib)(a - ib)]^{1/2} = (\mathbf{z} \times \mathbf{z}^*)^{1/2} \quad [5.6]$$

$\mathbf{z}^*$ é chamado de complexo conjugado de $\mathbf{z}$, com a parte imaginária negativa.

Agora multiplicando-se dois números complexos quaisquer:

$$\mathbf{z}_1 = a_1 + ib_1 = |\mathbf{z}_1| (\cos \alpha_1 + i \text{sen} \alpha_1) \text{ e } \mathbf{z}_2 = a_2 + ib_2 = |\mathbf{z}_2| (\cos \alpha_2 + i \text{sen} \alpha_2)$$

$$\mathbf{z}_1 \mathbf{z}_2 = |\mathbf{z}_1| |\mathbf{z}_2| (\cos \alpha_1 + i \text{sen} \alpha_1)(\cos \alpha_2 + i \text{sen} \alpha_2)$$

$$\mathbf{z}_1 \mathbf{z}_2 = |\mathbf{z}_1| |\mathbf{z}_2| (\cos \alpha_1 \cos \alpha_2 + i^2 \text{sen} \alpha_1 \text{sen} \alpha_2) + i(\cos \alpha_1 \text{sen} \alpha_2 + \cos \alpha_2 \text{sen} \alpha_1)$$

$$\mathbf{z}_1 \mathbf{z}_2 = |\mathbf{z}_1| |\mathbf{z}_2| (\cos \alpha_1 \cos \alpha_2 - \text{sen} \alpha_1 \text{sen} \alpha_2) + i(\cos \alpha_1 \text{sen} \alpha_2 + \cos \alpha_2 \text{sen} \alpha_1)$$

$$\mathbf{z}_1 \mathbf{z}_2 = |\mathbf{z}_1| |\mathbf{z}_2| \cos(\alpha_1 + \alpha_2) + i \text{sen}(\alpha_1 + \alpha_2) \quad [5.7]$$

De fato, como já se havia mencionado, o produto de dois números complexos tem uma magnitude proporcional ao produto de suas magnitudes individuais e a sua direção é determinada pela soma dos ângulos dos vetores multiplicados.

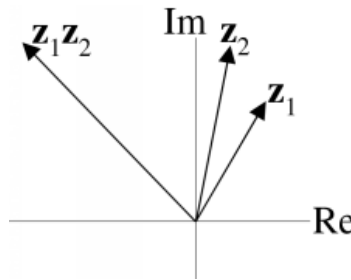


Figura 5-10: Representação gráfica dos números complexos Z_1 , Z_2 e do produto entre eles.

Resumindo, os números complexos podem ser representados com uma magnitude e um ângulo de fase no plano complexo e o produto de dois números complexos tem uma magnitude igual ao produto das magnitudes e uma fase igual à soma das fases.

2.4. A EQUAÇÃO DE EULER

De forma a simplificar a expressão dos números complexos, pode se utilizar a equação derivada por Euler:

$$e^{i\alpha} = \exp(i\alpha) = \cos\alpha + i\sin\alpha \quad [5.8]$$

E assim, a adição dos ângulos de fase que acontece na multiplicação se torna óbvia:

$$z_1 z_2 = |z_1| \exp(i\alpha_1) |z_2| \exp(i\alpha_2) = |z_1| |z_2| \exp[i(\alpha_1 + \alpha_2)]$$

Devido à simplificação, essa notação é a mais útil para se trabalhar com ondas.

3. DIFRAÇÃO: O FENÔMENO DE INTERFERÊNCIA

Os raios-x interagem com a matéria sendo espalhados – ou reemitidos – em todas as direções a partir dos elétrons que eles encontram. Raios-x espalhados de diferentes elétrons percorrerão distâncias diferentes, ou seja, eles diferirão em suas fases relativas. As ondas espalhadas interferem umas nas outras sendo somadas. Elas podem ser adicionadas em fase, nesse caso a amplitude resultante será a soma das amplitudes individuais, o que é chamada interferência construtiva. Ou elas podem ser somadas fora de fase, e nesse caso a amplitude resultante é a diferença das amplitudes individuais, o que é chamada interferência destrutiva.

3.1. QUANDO AS ONDAS SÃO ESPALHADAS EM FASE

Os pontos que formam o padrão de difração de um cristal são frequentemente chamados de reflexões porque se pode imaginar um cristal como sendo composto de milhares de espelhos que refletem os raios-x. Esses espelhos são chamados de planos de Bragg.

Quando a luz é refletida por um espelho, o ângulo de incidência – o ângulo no qual os raios de luz atingem o plano do espelho – é igual ao ângulo de reflexão. O mesmo é verdade para os planos de Bragg e a razão é que quando o ângulo de incidência é igual ao ângulo de reflexão, os raios de luz atingem o plano em fase e saem em fase. A figura 5-11 mostra porque:

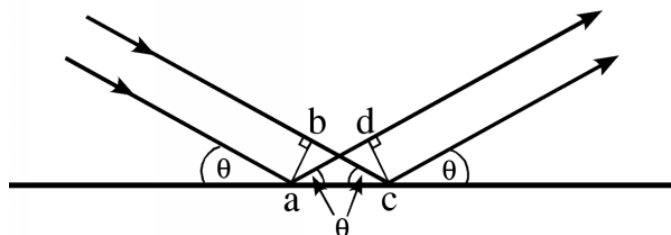


Figura 5-11: Dois raios atingem o plano de Bragg com ângulo θ e são refletidos pelo mesmo ângulo

Pela figura 5-11, os dois raios incidentes sairão em fase porque percorrem o mesmo caminho: se as linhas ad e bc forem diferentes no comprimento, os raios sairão fora de fase nos pontos c e d . Mas se essas linha têm o mesmo comprimento, eles estarão em fase em c e d . Pode-se ver que essas linhas terão o mesmo comprimento se e somente se o ângulo de incidência for igual ao ângulo de reflexão. Considere os dois triângulos abc e cda na figura 5-11: eles são ambos retângulos e compartilham o lado ac . Se eles tiverem outro ângulo igual então serão congruentes. Agora, o ângulo acb é o ângulo de incidência e o ângulo cad é o ângulo de reflexão. Quando esses ângulos são iguais, então os triângulos são congruentes e, conseqüentemente, os lados ad e bc devem ser iguais. Dessa forma os raios refletidos estarão em fase em cd .

Note que os raios incidentes e refletidos diferem na direção por um ângulo total de 2θ . Por isso, quando se observa um padrão de difração deve-se sempre lembrar de dividir o ângulo do feixe direto por dois para se saber o ângulo de incidência θ .

Assim, raios incidentes num mesmo plano percorrem caminhos idênticos sendo refletidos em fase mas, se eles forem refletidos de planos diferentes percorrerão caminhos diferentes. É justamente essa diferença de caminho que determina se os raios serão refletidos em fase. Essa diferença de caminho por sua vez depende do ângulo de incidência (e reflexão) e também do comprimento de onda da radiação incidente. A figura 5-12 ajuda a trabalhar essa relação a qual é conhecida como *Lei de Bragg*.

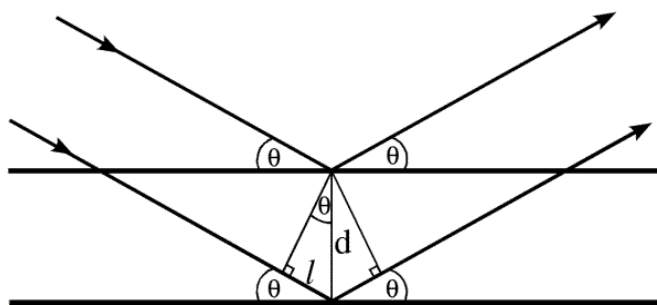


Figura 5-12: Aqui os dois raios incidentes atingem diferentes planos de Bragg.

Como se pode ver na figura 5-12, a diferença de caminho entre os raios refletidos de dois diferentes planos separados por d é duas vezes a distancia l . Como a soma dos ângulos internos de um triângulo deve ser 180 graus e, como se pode ver, um desses ângulos é 90

graus e o outro é $(90-\theta)$ graus então o último ângulo deve ser θ . Por trigonometria simples chega-se ao valor da distância l sendo igual a $d \sin \theta$. Para que os dois raios sejam difratados em fase, a diferença de caminho entre os raios refletidos deve ser igual ao comprimento de onda da radiação incidente, o que leva a seguinte relação:

$$\lambda = 2d \sin \theta \quad [5.9]$$

De fato, as duas ondas estarão em fase se os caminhos diferirem por qualquer múltiplo inteiro do comprimento de onda, por isso a Lei de Bragg é usualmente expressa como $n\lambda = 2d \sin \theta$. Entretanto, do ponto de vista da informação no padrão de difração, faz mais sentido escolher d para o qual n é igual a um.

Pela lei de Bragg, enquanto objetos sobre os planos difratam em fase, objetos entre os planos difratarão fora de fase. O deslocamento da fase será proporcional ao quanto o objeto estiver longe do plano, como uma fração da distância até o próximo plano. Como se pode ver, um simples evento de difração pode fornecer informações sobre as posições relativas dos objetos ao conjunto de planos.

Pela Lei de Bragg, com um aumento no ângulo θ , deve haver uma diminuição em d para que o caminho permaneça igual a um comprimento de onda. Isso pode ser mais bem visualizado por um simples rearranjo na equação 5.9:

$$d = \frac{\lambda}{2 \sin \theta} \quad [5.10]$$

Essa é uma das formas de se entender o conceito de espaço recíproco: quanto maior o ângulo de difração, menor o espaço no qual o padrão de difração será detectado.

Na figura 5-13 a seguir são mostrados dois raios com ângulos de incidência (e reflexão) diferentes e como a distância entre os planos deve ser afetada para que o caminho percorrido pelos dois permaneça igual ao comprimento de onda. Ela nos ajuda a pensar sobre dois limites para o ângulo de espalhamento: quando $\theta=0^\circ$ e $\theta=90^\circ$. Quando $\theta=0^\circ$ os raios incidentes são espalhados sem mudanças na direção independente da posição dos objetos. A distância d correspondente é infinita, o que significa que não existe distância

entre os planos que provoque mudança na fase, não existe difração e não existe informação espacial. Quando $\theta=90^\circ$, as ondas são refletidas voltando para a fonte. A diferença de caminho será duas vezes a distância entre os planos d , porque as ondas vão até o plano e voltam. Assim, a informação sempre estará limitada a espaçamentos iguais a metade do comprimento de onda da radiação que se está usando. Para se obter informação a maiores resoluções, é necessário escolher comprimentos de onda menores – isso explica mais uma vez porque se utilizam raios-x ao invés de luz visível.

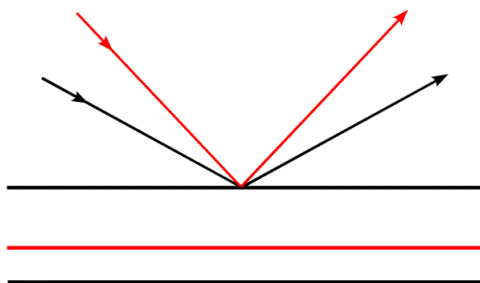


Figura 5-13: Se os ângulos de incidência dos raios são diferentes, a distância entre os planos de Bragg deve ser afetada para que eles continuem saindo em fase.

Até agora, tem-se mantido o objeto fixo e mudado o ângulo da radiação incidente. Mas para se observar os mesmos eventos de difração, pode-se também fixar a radiação incidente e rotar o objeto – o que é feito experimentalmente. Também é interessante considerar o que acontece se o objeto e a radiação incidente forem fixados. Diferentes raios serão difratados de diferentes conjuntos de planos que não são paralelos entre si, como mostrado na figura 5-14 a seguir:

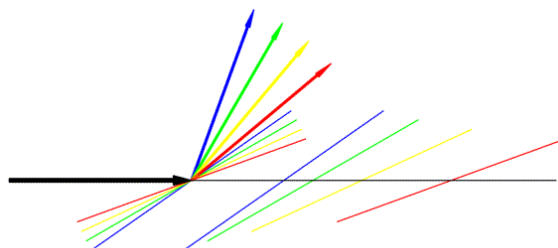


Figura 5-14: A seta preta mais forte representa a radiação incidente. As setas coloridas representam diferentes raios difratados de diferentes planos e um par de planos de Bragg é mostrado na mesma cor.

Pode-se pensar sobre o tipo de informação que um simples evento de difração pode fornecer. Como visto, objetos localizados nos planos de Bragg espalharão em fase. Se todos os objetos estiverem sobre planos de Bragg seria observada uma difração muito forte para a correspondente direção do espalhamento. Entretanto, se metade dos objetos estiverem sobre os planos e a outra metade na metade da distância entre eles, os dois conjuntos espalhariam fora de fase e não seria observada difração. Dessa forma, um ponto no padrão de difração contém informação sobre a concentração (distribuição) de objetos nos planos correspondentes. Ele informa sobre a posição média dos objetos nas direções perpendiculares aos planos mas nada sobre sua posição nas direções paralelas aos planos.

A figura 5-15 ilustra isso. Quando se considera o evento de difração representado pela seta preta, os objetos azul e magenta espalham fora de fase e suas contribuições para a onda total difratada serão canceladas. Entretanto, considerando o evento representado pela seta vermelha, eles serão adicionados quase em fase resultando numa onda difratada mais significativa.

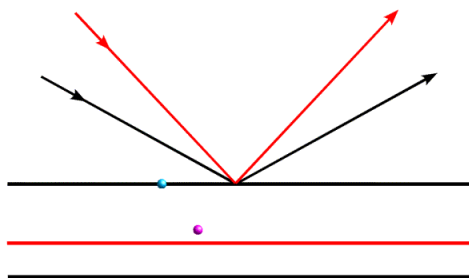


Figura 5-15: Agora foram adicionados dois diferentes objetos – em azul e magenta – em diferentes posições em relação aos planos de Bragg.

4. DESCRIÇÃO MATEMÁTICA DE DIFRAÇÃO

O objetivo de se descrever a difração matematicamente, é relacionar a amplitude e a fase das ondas difratadas em termos das posições dos elétrons (átomos) num cristal e as direções das ondas de raios-x incidentes e difratados.

Primeiro é necessário descrever as ondas de raios-x incidentes. Assumindo-se que um raio-X difratado por um elétron na origem do cristal tem uma fase relativa de zero,

então todas as outras fases serão expressas em relação à origem. A direção da onda pode ser expressa com um vetor na direção de propagação da onda. O vetor de onda é tipicamente denotado como $\mathbf{k}$.

O raio-x incidente é considerado como uma onda plana, com picos no campo elétrico ocorrendo em planos perpendiculares ao vetor de onda e separados por um comprimento de onda. Como discutido, a interferência é determinada pelo caminho percorrido por diferentes raios-X. Considerando-se a difração de um determinado elétron numa posição $\mathbf{r}$ com relação à origem do cristal, a figura 5-16 mostra o caminho que o raio-x percorre antes de atingir o elétron.

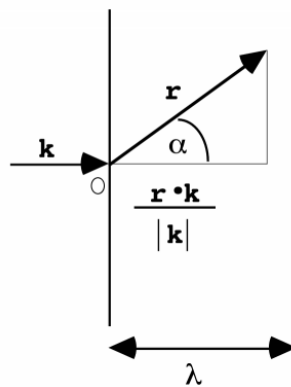


Figura 5-16: O caminho que o raio-x descrito pelo vetor de onda $\mathbf{k}$ percorre até atingir o elétron na posição $\mathbf{r}$ é dado pelo componente de $\mathbf{r}$ na direção de $\mathbf{k}$.

O que interessa é o componente de $\mathbf{r}$ na direção de $\mathbf{k}$, o qual é dado pelo produto escalar entre os dois vetores. Lembrando que $\mathbf{r} \bullet \mathbf{k} = |\mathbf{r}| |\mathbf{k}| \cos \alpha$ então o componente de $\mathbf{r}$ na direção de $\mathbf{k}$ é dado por $|\mathbf{r}| \cos \alpha = \frac{\mathbf{r} \bullet \mathbf{k}}{|\mathbf{k}|}$. Divide-se por λ para encontrar a

fração de comprimento de onda percorrido, o que resulta $|\mathbf{r}| \cos \alpha = \frac{\mathbf{r} \bullet \mathbf{k}}{\lambda |\mathbf{k}|}$. Por conveniência define-se a magnitude do vetor de onda $\mathbf{k}$ igual ao inverso do comprimento de onda pois assim o denominador desaparece. Lembrando que a posição aparece com sinal negativo na equação para a fase da onda, a fase na posição $\mathbf{r}$ é dada por $2\pi \mathbf{r} \bullet \mathbf{k}$ com relação à origem.

4.1. A FASE DA DIFRAÇÃO DE UM ÚNICO ELÉTRON

Vamos definir o feixe incidente pelo vetor de onda $\mathbf{k}_0$ e o feixe difratado pelo vetor de onda $\mathbf{k}$. Como vimos ambos têm uma magnitude igual ao recíproco do comprimento de onda. Como a fase da onda difratada a partir da origem é definida como zero, a fase de uma onda difratada a partir de um determinado local é obtida pela diferença nos caminhos percorridos.

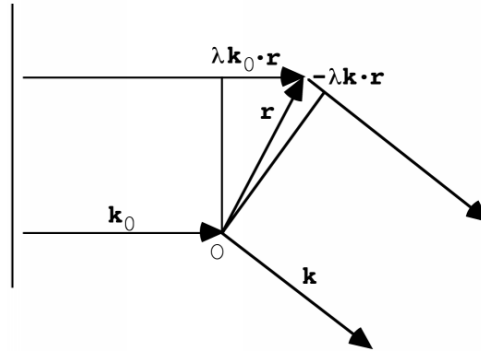


Figura 5-17: diferença de caminho entre o feixe incidente e refletido em termos dos vetores de onda que os representam.

Como ilustrado na figura 5-17, a diferença no caminho, expressa em múltiplos do comprimento de onda, é $\mathbf{k}_0 \bullet \mathbf{r}$ para o feixe incidente e $\mathbf{k} \bullet \mathbf{r}$ para o feixe difratado e, dessa forma, a fase total de um evento de difração a partir de um elétron localizado na posição $\mathbf{r}$ é $2\pi(\mathbf{k}_0 \bullet \mathbf{r} - \mathbf{k} \bullet \mathbf{r}) = 2\pi(\mathbf{k} - \mathbf{k}_0) \bullet \mathbf{r}$.

Agora pode-se definir o vetor de difração como $\mathbf{s} = \mathbf{k} - \mathbf{k}_0$ e a fase como $2\pi\mathbf{s} \bullet \mathbf{r}$.

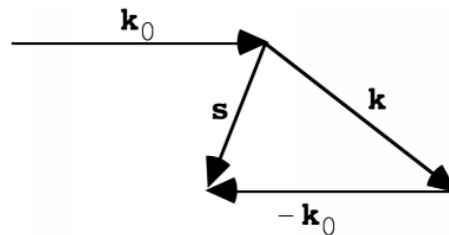


Figura 5-18: representação esquemática dos vetores de onda dos feixes incidente ($\mathbf{k}_0$) e difratado ($\mathbf{k}$) e do vetor de difração correspondente.

Assim, o importante para a fase da difração é o componente de $\mathbf{r}$ na direção do vetor da difração $\mathbf{s}$. O conjunto de pontos que satisfaz a equação $\mathbf{s} \cdot \mathbf{r} = c$, ou seja o conjunto de pontos para o qual o produto escalar do vetor $\mathbf{s}$ por $\mathbf{r}$ é uma constante, pertence ao plano perpendicular a $\mathbf{s}$ e difrata com a mesma fase. A equação para $c=0$ define o plano de Bragg que passa pela origem do cristal. Veja na figura 5-19:

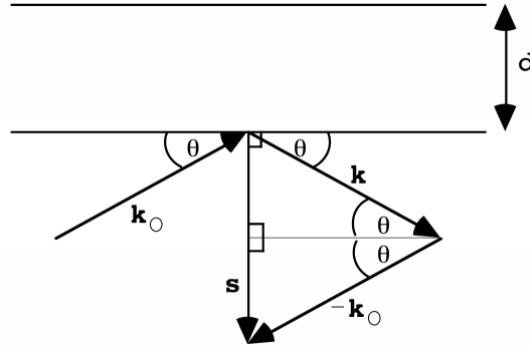


Figura 5-19: Representação do vetor de difração em relação ao plano de Bragg com separação d em relação ao próximo sobre o qual incide o feixe $\mathbf{k}_0$ que é refletido ($\mathbf{k}$).

A partir da figura 5-19 pode-se chegar novamente à lei de Bragg. A fase relativa para difração a partir da origem depende do valor de $\mathbf{s} \cdot \mathbf{r}$ ou o componente de $\mathbf{r}$ paralelo a $\mathbf{s}$. Na figura pode-se ver que:

$$|s| = 2 \sin \theta \quad |k| = \frac{2 \sin \theta}{\lambda} \quad [5.11]$$

Quando $\mathbf{s} \cdot \mathbf{r} = 1$, a fase de difração é deslocada por 2π e o componente de $\mathbf{r}$ na direção de $\mathbf{s}$ é igual ao espaçamento d entre os planos para esse ângulo de difração. Isso significa que:

$$|s| d = 1 \quad \text{ou} \quad |s| = \frac{1}{d} \quad [5.12]$$

Se as equações 5.11 e 5.12 forem igualadas e rearranjadas o resultado é a lei de Bragg (equação 5.9 e 5.10):

$$\lambda = 2d \sin \theta$$

4.2. INTERPRETAÇÃO FÍSICA DO PRODUTO $\mathbf{s} \bullet \mathbf{r}$

A fim de se expressar de forma mais intuitiva o significado da equação do fator de estrutura, a qual será desenvolvida a seguir, é útil considerar a interpretação física do produto $\mathbf{s} \bullet \mathbf{r}$. Lembrando que o produto escalar de dois vetores pode ser interpretado como a projeção de um dos vetores sobre o outro, ou o componente desse vetor que é paralelo ao outro, multiplicado pelo comprimento do outro vetor.

$$\mathbf{s} \bullet \mathbf{r} = |\mathbf{s}| |\mathbf{r}| \cos \phi \quad [5.13]$$

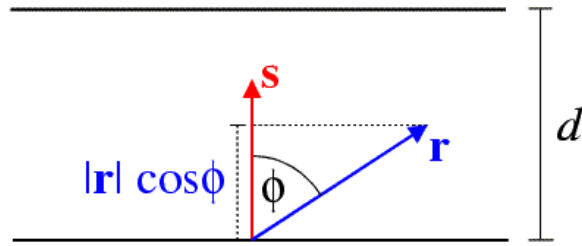


Figura 5-20: Representação gráfica do produto $\mathbf{s} \bullet \mathbf{r}$.

Pela figura 5-20, $|\mathbf{r}| \cos \phi$ é o componente do vetor posição $\mathbf{r}$ na direção do vetor de difração $\mathbf{s}$ que é perpendicular aos planos de Bragg. Como o comprimento do vetor de difração, $|\mathbf{s}|$, é igual a $1/d$, o produto $\mathbf{s} \bullet \mathbf{r}$ é igual à fração da distância que o vetor posição $\mathbf{r}$ percorre a partir da origem de um plano de Bragg até o próximo. É claro que $\mathbf{s} \bullet \mathbf{r}$ pode ser qualquer número real, inclusive ser maior que um. Como foi definido uma onda difratada a partir do plano de Bragg passando pela origem tem uma fase zero. As ondas difratadas a partir do próximo plano de Bragg têm uma fase de 2π (que é equivalente a fase de zero) e, generalizando, a difração a partir de qualquer ponto $\mathbf{r}$ terá uma fase de $2\pi \mathbf{s} \bullet \mathbf{r}$.

5. FATORES DE ESTRUTURA

5.1. FATOR DE ESTRUTURA DE UM ÚNICO ELÉTRON

O fator de estrutura representa a onda que resulta da difração, o que significa que ele é um número complexo com amplitude e fase. Para colocar o fator de estrutura numa escala absoluta, a difração a partir de um elétron num determinado ponto é definida como tendo uma amplitude de $1e$. O fator de estrutura freqüentemente é escrito como um número adimensional, mas deve-se sempre lembrar que ele tem unidade de elétrons.

$$F(s) = 1e \exp(2\pi i s \cdot r) \quad [5.14]$$

Se fosse possível medir a fase de um evento de difração a partir de um único elétron, isso apenas diria que o elétron está localizado em uma das séries de planos de Bragg separados pela distância d correspondente ao ângulo de difração. Quanto maior o ângulo, mais próximos esses planos ou menos espaçados.

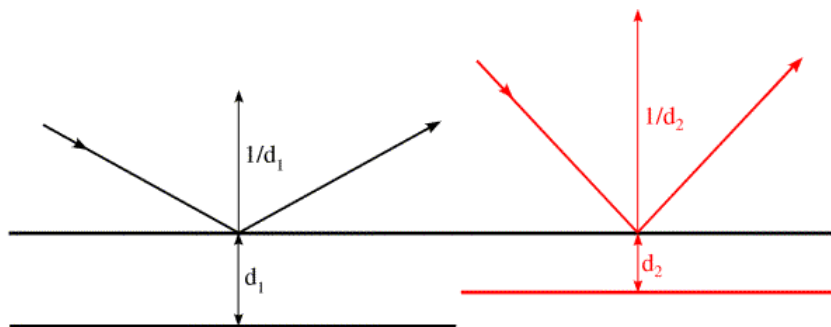


Figura 5-21: Dois diferentes eventos de difração ocorrendo em diferentes planos de Bragg.

5.2. FATOR DE ESTRUTURA DE UM NÚMERO DE ELÉTRONS

Se for medida a difração de um objeto contendo muitos elétrons localizados em diferentes pontos, a difração em cada ponto será a soma das ondas espalhadas por cada

elétron. Sabemos agora que a soma de ondas pode ser representada mais facilmente pela adição de números complexos expressos usando a equação de Euler, o que nos dá:

$$F(s) = \sum_j \exp(2\pi i s \bullet r_j) \quad [5.15]$$

O padrão de difração agora fornece informação sobre as posições relativas dos elétrons, devido aos efeitos de interferência. Quanto maior as diferenças de fase, maior os efeitos de interferência, o que significa que pequenas diferenças na posição têm efeitos mensuráveis quando o ângulo de difração aumenta. (Esta é a base do conceito de resolução.)

5.3. FATOR DE ESTRUTURA DE UMA DISTRIBUIÇÃO CONTÍNUA DE ELÉTRONS

Para se ir de uma distribuição discreta para uma contínua, basta substituir um somatório por uma integral. Ao invés de um elétron num ponto, tem-se agora uma função contínua que é a densidade eletrônica $\rho(\mathbf{r})$.

$$F(s) = \int_{\text{espaço}} \rho(r) \exp(2\pi i s \bullet r) dr \quad [5.16]$$

Esta é uma equação muito interessante. A operação que transforma a densidade eletrônica no fator de estrutura é o que os matemáticos conhecem como transformada de Fourier. Em outras palavras, o padrão de difração é a transformada de Fourier da densidade eletrônica.

6. DIFRAÇÃO A PARTIR DE UM CRISTAL

Um cristal amplifica o padrão de difração em certas direções, nas quais as várias células unitárias difratam em fase, e o elimina em outras direções. A fim de que as células

unitárias difratam em fase, os planos de Bragg devem passar através dos mesmos pontos em todas as células unitárias no cristal. O caso mais fácil de se imaginar é quando os planos estão separados por uma unidade ao longo de um dos eixos da célula unitária ou uma fração integral dessa unidade. Na figura 5-22 são mostrados os conjuntos de planos separados por uma unidade ao longo do eixo **a** da célula unitária e por um terço de unidade que passam através de átomos equivalentes nas diferentes células unitárias. Por outro lado, se os planos dividem a célula unitária por um número não integral, as diferentes células unitárias irão difratar fora de fase e as ondas serão canceladas.

Definindo a célula unitária: A célula unitária é delimitada por três eixos, **a**, **b**, e **c** ao longo das direções x, y e z respectivamente. O uso de negrito indica que os eixos da célula são vetores com comprimento e direção. A célula unitária é frequentemente descrita pelo comprimento dos eixos, $a=|\mathbf{a}|$, $b=|\mathbf{b}|$ e $c=|\mathbf{c}|$ chamados parâmetros da célula, e pelos ângulos entre eles, α =ângulo entre **b** e **c**, β =ângulo entre **a** e **c** e γ =ângulo entre **a** e **b**.

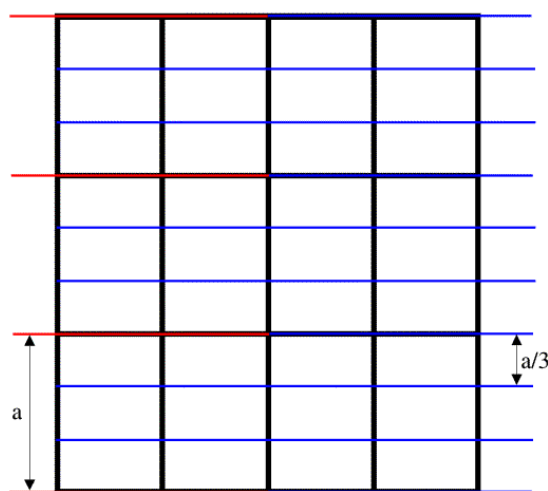


Figura 5-22: Diferentes células unitárias de um cristal.

Na figura 5-22, as linhas pretas delimitam as células unitárias, as vermelhas indicam os planos separados por uma unidade ao longo do eixo **a** da célula unitária e as azuis indicam os planos separados por um terço de unidade ao longo do eixo **a**. Os planos em vermelho são chamados planos (100) porque dividem o parâmetro **a** da célula uma vez e os

planos azuis são chamados planos (300) porque dividem o parâmetro **a** da célula três vezes. Isso remete à idéia de espaço recíproco: os planos – e os correspondentes pontos de difração – com índices maiores correspondem aos espaçamentos menores. Esses três índices que identificam os planos são chamados *h*, *k* e *l*, conhecidos como *índices de Miller*. Na figura são mostrados dois conjuntos de planos (*h*00).

Generalizando, para se observar difração a partir de um cristal, os planos de Bragg têm que dividir os lados da célula em um número inteiro de vezes. Nos planos (300) os lados **b** e **c** da célula são divididos zero vezes. É mais fácil ilustrar isso em duas dimensões, por exemplo, com os planos (*h k* 0). Três conjuntos desses planos são mostrados a seguir:

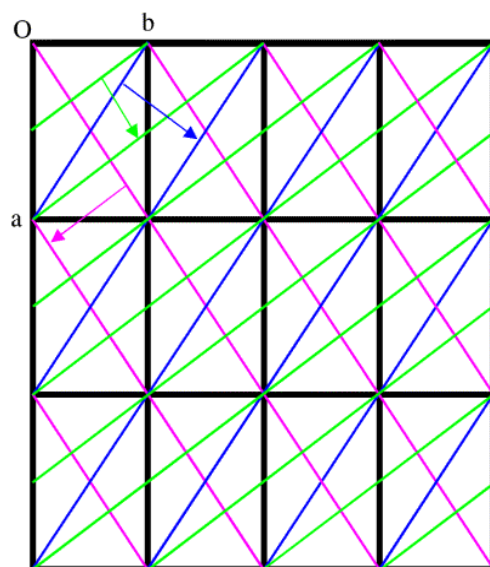


Figura 5-23: Célula unitária com três diferentes planos de Bragg representados: em azul os planos (110), em rosa os planos (1-10) e em verde os planos (210).

Fica um pouco mais complicado. Primeiro considere os planos azuis, basta seguir a seta azul de um plano para o próximo para perceber que eles estão separados por uma unidade tanto na direção do eixo **a** quanto na direção de **b**, por isso esses são os planos (110). Os planos em rosa são similares, seguindo a seta rosa percorre-se uma unidade na direção **a** até o próximo plano mas *volta-se* uma unidade na direção **b** até o próximo, por isso esses são os planos (1-10). Com os planos em verde, segue-se a seta verde meia unidade na direção **a** até o próximo plano e uma unidade inteira na direção **b** e assim, esses são os planos (210).

7. DIFRAÇÃO: O ESPAÇO RECÍPROCO

Por várias vezes tem se feito alusão ao espaço recíproco: maiores ângulos de espalhamento correspondem aos menores espaçamento entre os planos; o aumento nos índices dos planos corresponde a dividir mais finamente a célula unitária. Entretanto, matematicamente o espaço recíproco é definido pelo vetor de difração $\mathbf{s}$. Esses vetores de difração, como já descritos, são perpendiculares aos seus correspondentes planos de Bragg e seu comprimento é inversamente proporcional ao espaçamento d entre esses planos. O espaço dos vetores de difração $\mathbf{s}$ é também o inverso do espaço criado pela aplicação da transformada de Fourier à densidade eletrônica definida sobre o espaço real $\mathbf{r}$.

7.1. REDE RECÍPROCA

Até agora foi vista apenas uma maneira de pensar sobre o espaço recíproco: os índices de Miller (hkl) os quais definem os planos de Bragg que permitem a difração (interferência construtiva) a partir de um cristal. Os índices de Miller são úteis para definir os pontos da rede recíproca no espaço recíproco definido por $\mathbf{s}$. A rede recíproca é definida por uma célula unitária recíproca da mesma maneira que o retículo cristalino é definido pela célula unitária do cristal no espaço real.

Define-se a célula unitária recíproca por três vetores de difração correspondentes a planos separados pelos três eixos da célula unitária no espaço real. Considerando-se primeiro o eixo $\mathbf{a}$. Tomando-se os planos (100), separados por múltiplos inteiros de $\mathbf{a}$, se poderia pensar que o correspondente vetor de difração $\mathbf{s}$ deveria ser paralelo a $\mathbf{a}$ mas isso não é necessariamente verdadeiro. Os planos de Bragg são definidos pelos eixos $\mathbf{a}$ e $\mathbf{b}$ e o vetor de difração deve ser perpendicular ao plano $\mathbf{bc}$. Assim o primeiro eixo da célula recíproca, chamado $\mathbf{a}^*$, é definido como o vetor de difração para os planos (100), sendo perpendicular ao plano $\mathbf{bc}$ e o seu comprimento é o recíproco da distância entre um plano $\mathbf{bc}$ e o próximo. A figura 5-24 a seguir mostra uma seção de uma célula unitária monoclinica, na qual apenas o ângulo β não é ortogonal.

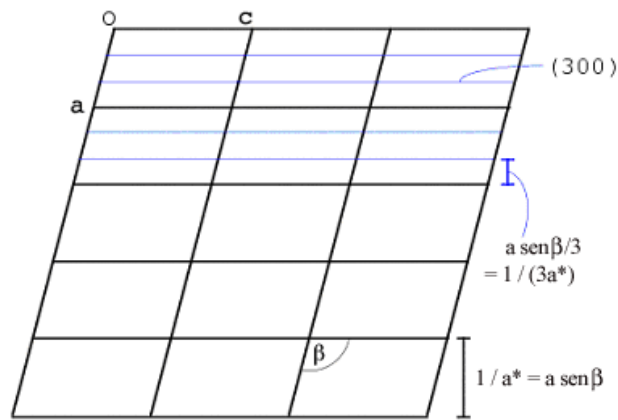


Figura 5-24: Uma seção de um cristal com célula unitária monoclinica mostrando os planos (300).

Os planos (300) são separados por um terço do espaçamento entre os planos **bc**, então o vetor de difração é dado por $3\mathbf{a}^*$. Generalizando, o vetor de difração para qualquer conjunto de planos ($h00$) é dado por $h\mathbf{a}^*$. Os outros dois eixos recíprocos são definidos do mesmo jeito. Então qualquer ponto na rede recíproca pode ser definido como a soma dos múltiplos inteiros dos eixos recíprocos:

$$\mathbf{s} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \quad [5.17]$$

A célula unitária recíproca

Os eixos recíprocos são definidos como tendo direção perpendicular aos outros dois eixos no espaço real e um comprimento igual ao recíproco do espaçamento entre os planos definidos pelos outros dois eixos no espaço real. Matematicamente:

$$\mathbf{a}^* = \mathbf{b} \times \mathbf{c} / [\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})]$$

Explicação:

$\mathbf{b} \times \mathbf{c}$ = vetor perpendicular a **b** e **c**, com comprimento igual à área do quadrilátero definido por eles, isto é

$$|\mathbf{b} \times \mathbf{c}| = b c \sin \alpha$$

$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ = componente de **a** na direção do vetor $\mathbf{b} \times \mathbf{c}$ multiplicado pelo comprimento desse vetor $|\mathbf{b} \times \mathbf{c}|$.

O espaçamento interplanar para o conjunto de planos (100) é então $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) / |\mathbf{b} \times \mathbf{c}|$, e assim a definição para $\mathbf{a}^*$ tem direção e comprimento certos.

O produto triplo $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ é também igual ao volume V da célula unitária e então se pode escrever:

$$\mathbf{a}^* = \mathbf{b} \times \mathbf{c} / V$$

$$\mathbf{b}^* = \mathbf{c} \times \mathbf{a} / V$$

$$\mathbf{c}^* = \mathbf{a} \times \mathbf{b} / V$$

Note que $\mathbf{a} \cdot \mathbf{a}^* = \mathbf{a} \cdot \{\mathbf{b} \times \mathbf{c} / [\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})]\} = 1$. Similarmente, $\mathbf{b} \cdot \mathbf{b}^* = \mathbf{c} \cdot \mathbf{c}^* = 1$.

Por outro lado, $\mathbf{a} \cdot \mathbf{b}^* = 0$ porque $\mathbf{b}^*$ é, por definição, perpendicular a $\mathbf{a}$. Similarmente,

$$\mathbf{b} \cdot \mathbf{a}^* = \mathbf{b} \cdot \mathbf{c}^* = \mathbf{c} \cdot \mathbf{b}^* = \mathbf{a} \cdot \mathbf{c}^* = \mathbf{c} \cdot \mathbf{a}^* = 0$$

Como visto, os índices de Miller (hkl) são muito úteis para descrever pontos na rede recíproca mas eles também são bem apropriados para descrever posições no espaço real com coordenadas fracionárias (xyz). As coordenadas fracionárias são definidas como frações dos eixos da célula unitária de tal forma que uma posição $\mathbf{r}$ é expressa como:

$$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c} \quad [5.18]$$

Que é ilustrada na figura 5-25 junto com a direção de $\mathbf{a}^*$:

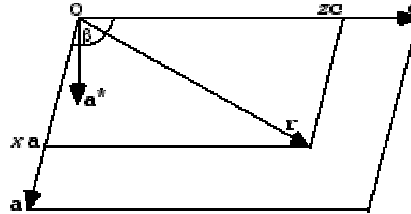


Figura 5-25: Aqui a seção representada se restringe à célula unitária monoclinica; as coordenadas fracionárias que indicam a posição $\mathbf{r}$ são mostradas; o eixo recíproco $\mathbf{a}^*$ também é ilustrado.

Os índices de Miller são apropriados porque implicitamente incorporam as dimensões da célula de tal maneira que o produto $\mathbf{s} \cdot \mathbf{r}$ pode ser substituído por $h \cdot x$:

$$\mathbf{s} \cdot \mathbf{r} = (h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*) \cdot (x\mathbf{a} + y\mathbf{b} + z\mathbf{c})$$

$$\mathbf{s} \cdot \mathbf{r} = hx + ky + lz$$

$$\mathbf{s} \cdot \mathbf{r} = h \cdot x$$

As simplificações ocorrem devido às relações entre os eixos reais e recíprocos nas quais o produto é igual a um para os pares como $\mathbf{a}$ e $\mathbf{a}^*$ e igual a zero para os pares como $\mathbf{a}$ e $\mathbf{b}^*$.

Agora se pode expressar a equação do fator de estrutura, a partir da equação 5.16, em termos dos índices de Miller e coordenadas fracionárias:

$$F(s) = \int_{\text{espaço}} \rho(r) \exp(2\pi i s \cdot r) dr \quad [5.16]$$

$$F(s) = \int_{\text{célula}} \rho(x) \exp(2\pi i h \cdot x) dx \quad [5.19]$$

Note que a integral é agora feita apenas sobre uma célula unitária. Por convenção o fator de estrutura é definido como a transformada de Fourier de uma única célula unitária. Isso porque para se fazer a integral sobre um grande número de células unitárias, como num cristal, basta simplesmente multiplicar o fator de estrutura pelo número de células unitárias.

7.2. FATORES DE ESTRUTURA EM TERMOS DE ÁTOMOS: O FATOR DE ESPALHAMENTO ATÔMICO

Até agora a equação do fator de estrutura foi expressa em termos dos índices de Miller, que estão relacionados com os pontos individuais observados no padrão de difração. Entretanto, quando se está construindo um modelo trabalha-se com átomos e não diretamente com as suas distribuições de densidade eletrônica. Assim o próximo passo útil é expressar a equação do fator de estrutura em termos dos tipos e posições dos átomos individuais.

Lembrando que a equação do fator de estrutura foi primeiro derivada como a soma das contribuições de elétrons individuais, pode-se quebrá-la na soma das contribuições de elétrons dos átomos individuais. Uma outra forma de dizer isso é que a transformada de Fourier da soma é a soma das transformadas de Fourier individuais. Lembrando-se ainda que a contribuição de um elétron na origem para o fator de estrutura é $1e$ com uma fase

zero para todos os vetores de difração. Movendo-se o átomo para um ponto diferente na célula unitária apenas muda a fase da sua contribuição de uma maneira que depende tanto da sua posição ($\mathbf{r}$) quanto do vetor de difração ($\mathbf{s}$): a nova fase é dada por $2\pi\mathbf{s} \bullet \mathbf{r}$. Da mesma forma, pode-se calcular a contribuição do fator de estrutura de um átomo na origem e então mudar a sua fase de acordo com a sua posição na célula unitária.

$$F(\mathbf{s}) = \int_{\text{espaço}} \rho(\mathbf{r}) \exp(2\pi i \mathbf{s} \bullet \mathbf{r}) d\mathbf{r} \quad [5.16]$$

$$F(\mathbf{s}) = \sum_j \int_{\text{espaço}} \rho_j(\mathbf{r}) \exp(2\pi i \mathbf{s} \bullet \mathbf{r}) d\mathbf{r} \quad [5.20]$$

onde j denota os átomos individuais e

$$\int_{\text{espaço}} \rho_j(\mathbf{r}) \exp(2\pi i \mathbf{s} \bullet \mathbf{r}) d\mathbf{r} = \int_{\text{espaço}} \rho_j(\mathbf{r} - \mathbf{r}_j) \exp(2\pi i \mathbf{s} \bullet \mathbf{r}) d\mathbf{r}$$

onde $\mathbf{r}_j$ é o centro do átomo j de tal maneira que $\rho_j(\mathbf{r} - \mathbf{r}_j)$ é a densidade desse átomo centrado na origem e então

$$\begin{aligned} &= \int_{\text{espaço}} \rho_j(\mathbf{r} - \mathbf{r}_j) \exp(2\pi i \mathbf{s} \bullet \mathbf{r}_j) \exp(2\pi i \mathbf{s} \bullet (\mathbf{r} - \mathbf{r}_j)) d\mathbf{r} \\ &= \exp(2\pi i \mathbf{s} \bullet \mathbf{r}_j) \int_{\text{espaço}} \rho_j(\mathbf{r}') \exp(2\pi i \mathbf{s} \bullet \mathbf{r}') d\mathbf{r}' \end{aligned}$$

substituindo $\mathbf{r}'$ por $\mathbf{r}' = \mathbf{r} - \mathbf{r}_j$

$$= f_j(|\mathbf{s}|) \exp(2\pi i \mathbf{s} \bullet \mathbf{r}_j) \text{ assim}$$

$$F(\mathbf{s}) = \sum_j f_j(|\mathbf{s}|) \exp(2\pi i \mathbf{s} \bullet \mathbf{r}_j) \quad [5.21]$$

$$F(\mathbf{h}) = \sum_j f_j(|\mathbf{s}|) \exp(2\pi i \mathbf{h} \bullet \mathbf{x}_j) \quad [5.22]$$

O padrão de difração de um átomo centrado na origem depende apenas do tipo de átomo. Para uma aproximação muito boa, pode-se considerar a densidade eletrônica de um átomo centrado na origem como sendo esfericamente simétrica, o que tem duas consequências:

1. A fase do espalhamento total será zero, porque para qualquer densidade eletrônica em $+\mathbf{r}$ haverá uma densidade eletrônica igual em $-\mathbf{r}$ sendo ambas canceladas
2. A difração dependerá apenas da resolução ($|\mathbf{s}|$) e não da direção do vetor de difração

Para ângulos de espalhamento muito baixos – que correspondem a espaçamentos d muito maiores que as dimensões do átomo – os elétrons do átomo espalharam em fase e a contribuição do espalhamento total será aproximadamente igual ao número de elétrons que ele contém. Para ângulos de espalhamentos maiores, diferentes partes da distribuição eletrônica do átomo começaram a espalhar parcialmente fora de fase de tal forma que a contribuição do espalhamento total diminuirá com o aumento da resolução.

7.3. REDE RECÍPROCA E A ESFERA DE EWALD

Ewald elaborou uma construção geométrica que ajuda a visualizar quais planos de Bragg estão na orientação correta para difratar. É a chamada esfera de Ewald e está ilustrada na figura 5-26.

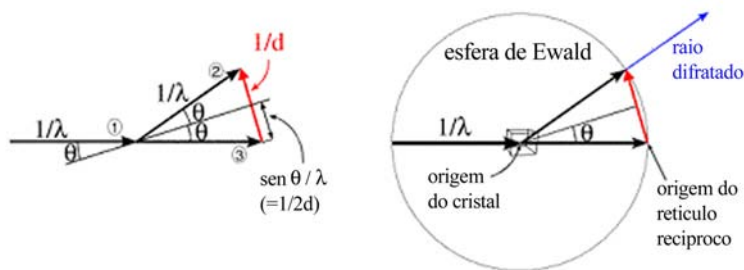


Figura 5-26: A esfera de Ewald mostra a correlação entre o espaço real e o recíproco.

Os raios incidente e difratado estão representados como vetores e, convenientemente, foi dado a eles o comprimento de $1/\lambda$. O raio incidente (1) e o raio difratado (2) formam ambos um ângulo θ com um conjunto de planos de Bragg no cristal.

Considere a diferença, mostrada em vermelho, entre o feixe direto que atravessa o cristal sem qualquer desvio (3) e o raio difratado. Por geometria simples, este vetor vermelho é perpendicular aos planos de Bragg, então ele está na direção do vetor do espaço recíproco. Pelos triângulos pode-se notar que os lados correspondentes às duas metades do vetor vermelho têm, cada um, um comprimento de $\sin\theta/\lambda$, que, pela lei de Bragg, deve ser igual a $1/2d$. Assim o vetor vermelho é, de fato, o vetor do espaço recíproco com um comprimento de $1/d$. Essa construção pode ser feita para qualquer conjunto de planos em condição de difração e o correspondente vetor do espaço recíproco estará sempre entre as pontas dos vetores que representam o feixe direto e o raio difratado. Por todos esses vetores do espaço recíproco começarem sempre no mesmo ponto, a base do vetor vermelho na figura 5-26, esse ponto foi definido como a origem do espaço recíproco.

O ângulo 2θ pode ser qualquer entre 0 e 180 graus, ou seja, θ pode ser qualquer ângulo entre 0 e 90 graus. Assim o raio difratado pode ir em qualquer direção em três dimensões e o vetor que o representa pode ter a ponta em qualquer lugar na superfície de uma esfera com um raio de $1/\lambda$. Se um conjunto de planos está em condição de difração, o correspondente vetor do espaço recíproco tem que terminar na superfície da esfera de Ewald e, se o feixe direto não incide nos planos com o ângulo θ correto, o vetor do espaço recíproco não estará na superfície da esfera de Ewald. A origem do cristal está no centro da esfera de Ewald e os raios incidentes são difratados a partir do cristal. A origem do espaço recíproco está no ponto onde o feixe de raios-x incidente deixa a esfera de Ewald. Se o cristal for rotado, os planos de Bragg e o espaço recíproco serão rotados na mesma direção. Como a difração de um cristal está limitada a pontos na rede recíproca – que correspondem aos planos que podem ser especificados por índices inteiros – pode-se dizer que há a rotação da rede recíproca quando o cristal é rotado. A figura 5-27 a seguir mostra isto esquematicamente, ilustrando planos de pontos na rede recíproca. Os planos de pontos na rede recíproca interseccionam a esfera de Ewald para dar um círculo de pontos em condição de difração. Quando os planos estão alinhados perpendicularmente ao feixe de raios-x, esses círculos na esfera de Ewald projetarão círculos de pontos ao redor da posição do feixe direto mas, como o cristal é rotado (e a rede recíproca) os círculos na esfera de Ewald serão distorcidos e projetarão o que são chamadas *luas* de pontos.

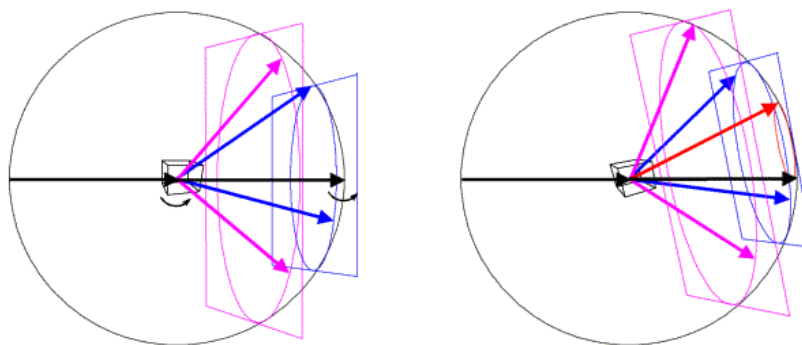


Figura 5-27: À esquerda é mostrado um cristal orientado de tal forma que os planos na rede recíproca estão perpendiculares ao feixe de raios-x. Os diferentes planos, círculos e feixes difratados que interseccionam a esfera de Ewald são mostrados em diferentes cores para cada plano de Bragg. Quando o cristal é rotado, a rede recíproca também é rotada, como mostrado à direita. Os pontos para um plano no nível zero, em vermelho, estarão atrás do beam stop.

8. DIFRAÇÃO DE UM CRISTAL REAL

Quando um cristal é posto num feixe de raios-x, alguns dos planos de Bragg estarão na orientação correta para difração e serão vistos pontos para eles no padrão de difração. Se o cristal for rotado, outros conjuntos de pontos irão atingir a orientação correta e entrar em condição de difração e novos pontos de difração serão vistos para eles.

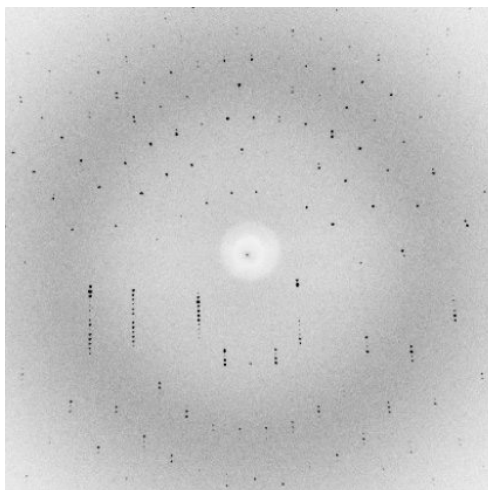


Figura 5-28: Padrão de difração formado pela coleta de pontos no espaço recíproco para um cristal real.

A imagem 5-28 mostra o resultado quando se coloca um cristal de proteína num feixe de raios-x. O cristal é rotado por ângulos pequenos, em torno de 1 grau, e a cada rotação é coletada uma nova imagem. O detector é posicionado de tal maneira a interseccionar a esfera de Ewald e assim se forma o padrão de difração. Observe que as *luas de pontos* formam um conjunto de círculos concêntricos ao redor da posição onde o feixe direto atinge o detector.

Um cristal real não é perfeito e tem um tamanho finito, particularmente os cristais de proteína estão longe de serem perfeitos e são freqüentemente bem pequenos devido ao tamanho e forma das moléculas (proteínas) que são empacotadas para formar o cristal. Por isso os cristais são considerados como sendo compostos de pequenos blocos mosaicos, os quais podem ser considerados como fragmentos óticamente independentes, ou seja, um conglomerado de cristais perfeitos com diferentes orientações ou diferentes ângulos de espalhamento.

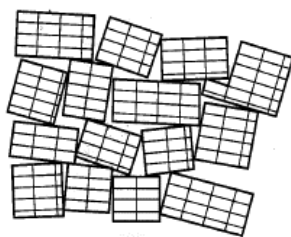


Figura 5-29: Mosaicidade comumente presente em cristais de proteína.

Essas imperfeições expressam a *mosaicidade* do cristal, que é medida em graus e reflete a desorientação relativa média de um bloco de cristais perfeitos, que também são chamados domínios mosaicos ou blocos mosaicos, dentro do cristal. Um alto valor de mosaicidade é vista por uma amplificação dos pontos no padrão de difração, os pontos aparecem borrados. Os pontos aparentam ser mais que o usual, porque para um cristal mosaico qualquer ângulo de rotação para o cristal inteiro realmente representa uma série de ângulos de rotação, um para cada bloco mosaico, e cada um desses blocos difrata independentemente dos outros. Conseqüentemente, pontos de cristais muito mosaicos permanecerão em condição de difração por mais tempo do que os pouco mosaicos, assim cada ponto conterà o espalhamento sobre um amplo ângulo de rotação e, será sobreposto

com outros pontos. Nos casos mais severos a difração é muito fraca e/ou os pontos coletados muito borrados e sobrepostos. Nesses casos o cristal deve ser desprezado. Nos casos em que a mosaicidade está dentro da faixa considerada normal, 0,25–0,75° para uma proteína típica, é aplicado um fator de correção à equação da intensidade do ponto medido.

$$I(hkl) = \frac{\lambda^3}{\omega} \left(\frac{e^2}{mc^2} \right)^2 \frac{V_{\text{cristal}}}{V^2} I_o LPA |F(hkl)|^2 \quad [7.2]$$

Na equação (que será cuidadosamente discutida no próximo capítulo) λ é o comprimento de onda, ω é a velocidade angular de rotação do cristal, e é a carga e m é a massa de um elétron, c é a velocidade da luz no vácuo e I_o é a intensidade do feixe incidente. Os demais termos correspondem a outras correções que são aplicadas. O termo V_{cristal}/V^2 corrige a mosaicidade do cristal. O fator de Lorentz L depende da técnica de aquisição dos dados e leva em conta a diferença de tempo que cada plano de Bragg entra em condição de difração. P é o fator de polarização que corrige a redução das intensidades devido à polarização parcial do feixe de raios-x. E A é o fator de absorção que corrige a absorção anisotrópica de raios-x pelo cristal.

Todos os fatores de correção são importantes durante a coleta dos dados de difração para um cristal de proteína e serão cuidadosamente abordados no capítulo 7 sobre a coleta dos dados de difração. Como se viu, toda a teoria da cristalografia se baseia na difração por cristais, considerando implicitamente que eles já foram obtidos e com qualidade suficiente para o experimento de difração. Entretanto, o problema inicial de um cristalógrafo de proteínas é a obtenção de bons cristais da proteína e esse muitas vezes é o maior problema. É justamente sobre a obtenção de cristais adequados que trata o próximo capítulo.

PRINCÍPIOS DA CRISTALIZAÇÃO DE MACROMOLÉCULAS

1. PORQUE CRISTAIS?

O espalhamento de raios-x por uma única molécula de proteína seria inimaginavelmente fraco e nunca poderia ser detectado acima do nível do ruído, o qual inclui o espalhamento causado pelo ar e pela água. Um cristal é o arranjo de um grande número de moléculas numa mesma orientação, o que permite que as ondas espalhadas em fase sejam somadas e elevem o sinal para um nível mensurável. Nesse sentido um cristal atua como um amplificador.

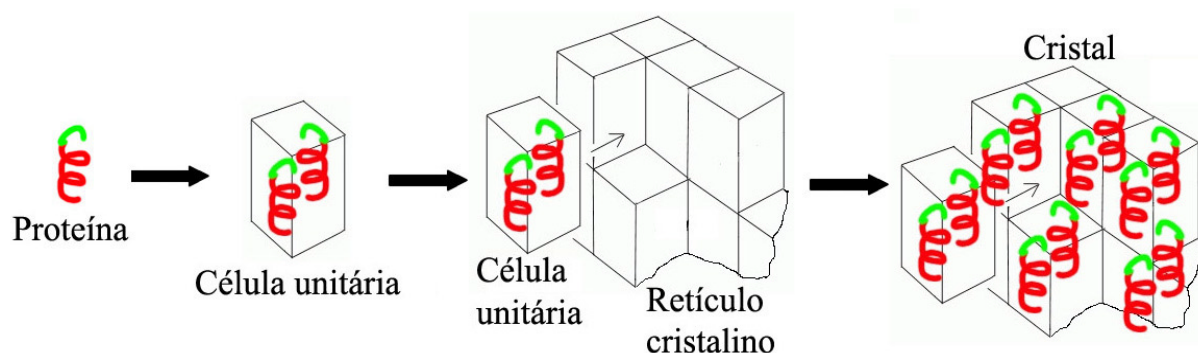


Figura 6-1: Representação esquemática da formação de um cristal de proteína.

Existem vários entraves na determinação da estrutura cristalográfica de uma macromolécula, mas o crescimento de um cristal adequado talvez seja o mais sério deles.

Enquanto algumas proteínas podem ser cristalizadas muito facilmente, existe um número bem maior que necessita de muitas tentativas até que se obtenha condições para a formação de bons cristais.

2. PRINCÍPIOS FÍSICO-QUÍMICOS DO PROCESSO DE CRISTALIZAÇÃO

A partir do diagrama de fases da figura 6-2 é possível se ter uma idéia de como ocorre o processo de cristalização. Para obter cristais de um composto, suas moléculas devem ser levadas à supersaturação, estado termodinamicamente instável, que pode resultar numa fase cristalina ou amorfa, quando retorna ao equilíbrio. Partindo-se da supersaturação, o processo de formação de cristais passa por um estado metaestável, e a conversão desse estado metaestável ao estado mais estável é cineticamente controlada. A cristalização de uma macromolécula acontece pela diminuição lenta de sua solubilidade, se o processo acontece muito rapidamente ocorre precipitação. (MCPHERSON, 1994)

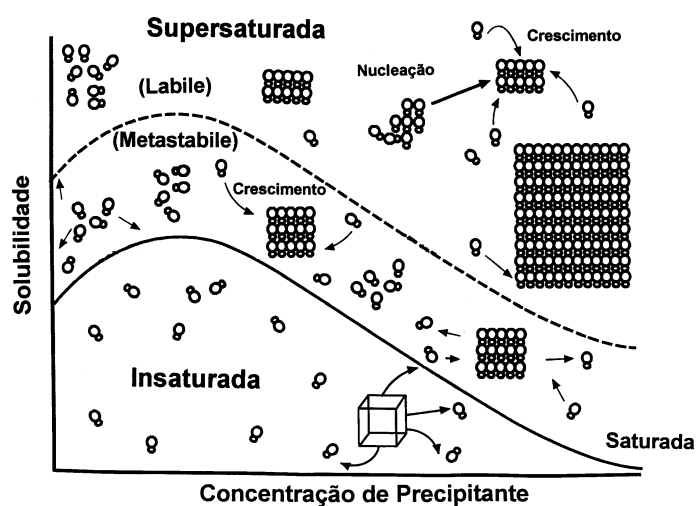


Figura 6-2: Diagrama de fases que representa a solubilidade de uma macromolécula em função da concentração de precipitante adicionado à solução.

O processo de cristalização pode ser dividido em três estágios: nucleação, crescimento e cessação do crescimento.

Para que o processo de nucleação, a mais importante das etapas do processo de formação de cristais, possa ocorrer, partindo-se de uma região de supersaturação (termodinamicamente instável) aglomerados são simultaneamente formados e solubilizados até que consigam transpor uma barreira de energia, a energia de ativação para a nucleação,

e comecem a crescer, “roubando” moléculas de proteína da solução e passando para uma região de insaturação (mais estável) formando cristais que crescem até que a incorporação e a saída de moléculas de proteína ao cristal sejam igualadas, então o crescimento cessa. (MCPHERSON, 1994)

A formação de cristais a partir de uma solução pode ser representada similarmente ao diagrama de fases de reações químicas convencionais. As moléculas livres em solução na região de supersaturação, representadas à esquerda na figura 6-3, se encontram num estado de maior energia em relação ao estado cristalino, representado à direita, indicando que a formação de cristais é um processo energeticamente favorável. Entretanto, para passar do estado livre para o estado cristalino, um núcleo crítico correspondente a um intermediário altamente energético deve ser formado. Essa barreira de energia que deve ser transposta corresponde ao estado de transição e determina a cinética do processo, isto é, a razão com que a nucleação ocorre. (MCPHERSON, 1994)

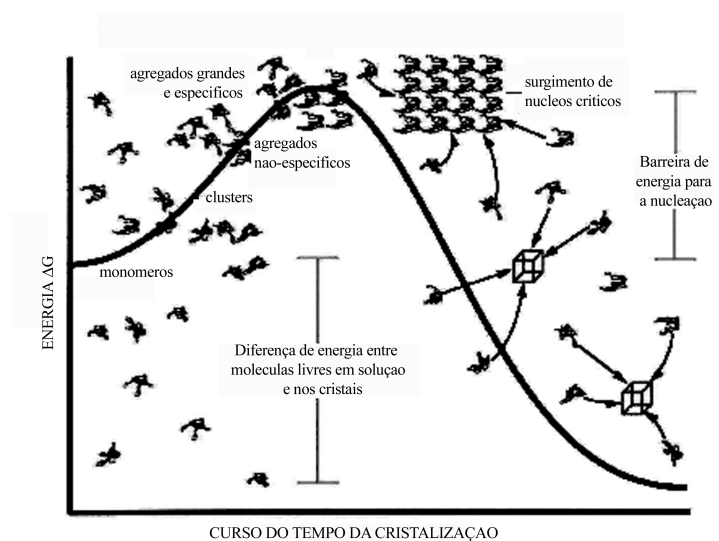


Figura 6-3: Diagrama de fases ilustrando a barreira energética que deve ser vencida para que o processo de nucleação ocorra.

Numa solução suficientemente supersaturada, núcleos podem formar-se a partir de dois mecanismos principais: se o núcleo se forma dentro da solução a nucleação é chamada homogênea e, se o núcleo se forma sobre um substrato sólido, é chamada heterogênea. A nucleação homogênea ocorre em soluções extremamente puras e requer uma altíssima supersaturação. Se as moléculas do soluto têm alguma afinidade por um substrato sólido, como qualquer impureza que esteja presente ou mesmo pelas paredes do recipiente de

cristalização, pode ocorrer nucleação heterogênea sobre essas superfícies e com uma supersaturação relativamente mais baixa. (MCPHERSON, 1994)

Quando o núcleo formado atinge um tamanho crítico, ele entra em fase de crescimento. Um cristal pequeno pode crescer incorporando moléculas nas faces planas em duas dimensões ou em espiral, ou sobre defeitos na superfície do cristal, incorporando moléculas por interações em três dimensões. O mecanismo dependerá da energia das interações entre os constituintes do cristal nas direções consideradas e a taxa de crescimento de cada face do cristal é uma função dessas energias. O hábito do cristal, seu aspecto externo ou morfologia, depende diretamente da taxa de crescimento de cada face: faces crescidas lentamente serão maiores e bem desenvolvidas mas, se o crescimento for rápido, essas faces serão menores e mal formadas. (MCPHERSON, 1994)

Cristais nunca crescem indefinidamente e isso é particularmente verdadeiro para cristais de proteínas. De fato, um dos problemas mais comuns é a obtenção de cristais muito pequenos para análise por difração com raios-x. Isso pode ser devido ao acúmulo de defeitos dentro da rede cristalina, impedindo o crescimento, ou à presença de impurezas, pois foi observado que quanto maior a pureza do material de partida maiores cristais podem crescer. (MCPHERSON, 1994)

3. PARÂMETROS QUE INFLUENCIAM NA CRISTALIZAÇÃO DE MACROMOLÉCULAS

Na prática, freqüentemente um número muito grande de condições tem que ser testadas até que seja obtido sucesso, pois existem muitos parâmetros que influenciam na formação de cristais. Dentre esses parâmetros destacam-se: (DUCRUIX & GIEGÉ, 1999)

3.1. PARÂMETROS FÍSICO-QUÍMICOS INTRÍNSECOS

- Supersaturação (concentração da macromolécula e precipitante)
- Temperatura e pH
- Tempo
- Força iônica e pureza dos reagentes (natureza dos precipitantes, tampões, aditivos, etc)
- Difusão e convecção (géis, microgravidade)

- Efeito de partículas sólidas (nucleação homogênea e heterogênea)
- Efeito da densidade e viscosidade
- Efeito de campos gravitacionais, elétricos e magnéticos.
- Ondas acústicas vibracionais e sonoras
- Uso de robôs para cristalização (experimentalista versus robô: menor chance de erros)

3.2. PARÂMETROS BIOQUÍMICOS E BIOFÍSICOS

- Sensibilidade conformacional da macromolécula aos parâmetros físicos (temperatura, força iônica, pH, etc)
- Associação e ligantes (substratos, cofatores, íons metálicos, outros íons, etc)
- Aditivos específicos (agentes redutores, detergentes não iônicos, poliaminas, etc)
- Propriedades da macromolécula (oxidação, hidrofiliabilidade, hidrofobicidade, etc)

3.3. PARÂMETROS BIOLÓGICOS

- Raridade de diversas macromoléculas
- Estado em que são encontradas nos organismos e células
- Contaminantes bacteriológicos
- Pureza da macromolécula coletada ou extraída

4. ESTRATÉGIAS DA CRISTALIZAÇÃO DE MACROMOLÉCULAS

A estratégia mais abordada na prática é solubilizar a proteína num solvente, comumente água, contendo agente tamponante para controlar-se o pH, aditivos se necessários para manter a estabilidade da proteína, e precipitante e conduzir essa solução protéica à supersaturação. Então induzi-la ao retorno para o equilíbrio, promovendo um aumento lento e gradativo da concentração da proteína, na expectativa de que haja a formação de alguns pontos de nucleação para que alguns desses núcleos cresçam e se tornem cristais.

Normalmente o precipitante é um sal adicionado à solução. Seus íons competem com as moléculas de proteína pelas moléculas de água, pois ambos necessitam de hidratação para manter a solubilidade. Quando a competição se torna intensa, as moléculas protéicas se associam para satisfazer seus requerimentos eletrostáticos, eliminando assim as moléculas de água e, conseqüentemente, começam a perder solubilidade. Este fenômeno é conhecido por “*salting-out*”, que é diminuição da solubilidade da proteína com o aumento da concentração de sal, e a cristalização ocorre normalmente nessas condições. Entretanto, há proteínas cuja solubilidade aumenta à medida que a concentração de sal aumenta. Este fenômeno é conhecido por “*salting-in*” e a cristalização pode ocorrer nessas condições para algumas macromoléculas.

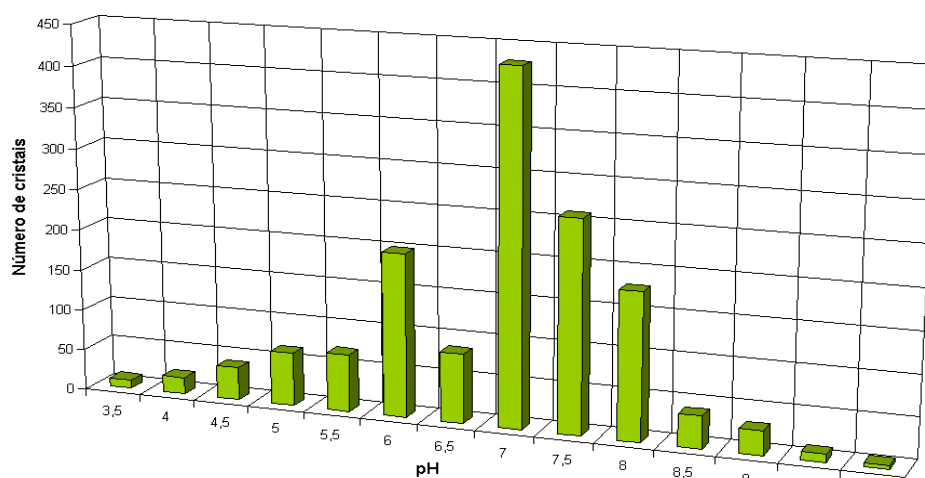


Figura 6-4: Efeito do pH no crescimento de cristais protéicos.

Uma outra abordagem estratégica é alterar-se o pH da solução. Alterando a concentração do precipitante modificam-se as forças iônicas mas, mantendo-se constante a concentração do precipitante e assim a força iônica, a variação do pH da solução acarreta mudanças na distribuição de cargas pela superfície da molécula que podem viabilizar a cristalização. Como se pode ver pela figura 6-4, o pH no qual foi cristalizado o maior número de macromoléculas está em torno de 7. Esse é o pH fisiológico no qual, pelo menos as macromoléculas extraídas de organismos vivos, particularmente as proteínas humanas, são mais estáveis.

Um outro parâmetro muito importante na cristalização de macromoléculas é a temperatura. Ela afeta a solubilidade de todas as substâncias na solução por isso é

importante preparar cristais à temperatura constante. Pela figura 6-5 pode-se ver que a temperatura na qual mais macromoléculas foram cristalizadas está entre 0-4°C e entre 20-24°C o que reflete apenas as temperaturas nas quais é mais fácil de se trabalhar. As câmaras frias mais comumente usadas para cristalização disponíveis comercialmente operam na faixa de 0-4°C, que é a temperatura das geladeiras domésticas também. E a temperatura entre 20-24°C é a temperatura ambiente.

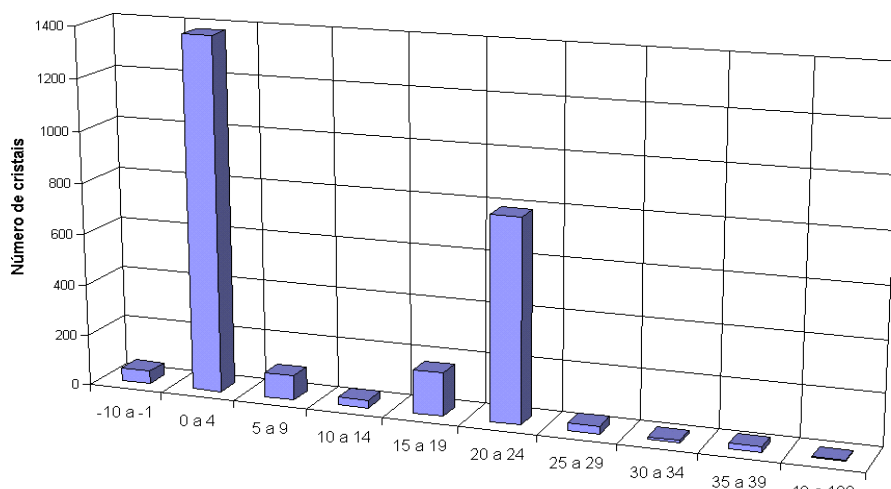


Figura 6-5: Efeito da temperatura no crescimento de cristais protéicos.

5. TÉCNICAS DE CRISTALIZAÇÃO

Existe uma série de técnicas que podem ser utilizadas nas tentativas de cristalização e, embora todas as técnicas possuam o mesmo princípio, a maneira de conduzir o processo é particular a cada uma e favorece a cristalização de um ou outro tipo de proteína.

As técnicas nas quais os resultados obtidos são melhores e por isso as mais utilizadas são a difusão de vapor, diálise e banho. A seguir está uma breve descrição de cada uma delas.

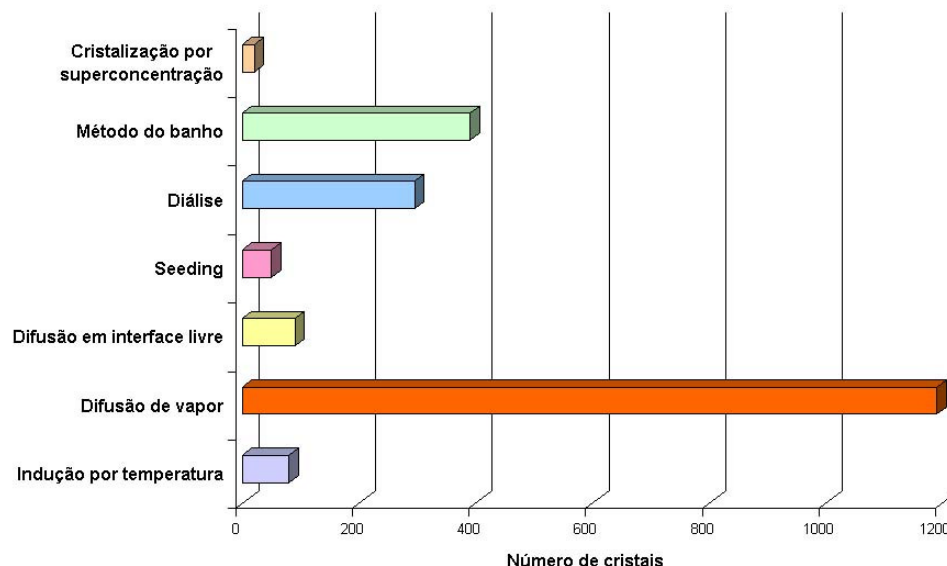


Figura 6-6: Histograma mostrando o uso relativo dos vários métodos para a cristalização de macromoléculas. Atualmente, os métodos que usam a difusão de vapor são os mais populares.

5.1. DIFUSÃO DE VAPOR

O princípio dessa técnica é o processo de equilíbrio entre duas soluções através da fase de vapor num meio fechado. Uma gota de solução da proteína a ser cristalizada contendo uma baixa concentração do agente precipitante é equilibrada contra um volume muito maior de solução com uma concentração bem mais alta do agente precipitante. Assim, devido à diferença entre os potenciais químicos, a solução menos concentrada perde água numa tentativa de que a concentração de precipitante das duas soluções se iguale. À medida que a gota perde água, a concentração de proteína também aumenta lentamente favorecendo a sua cristalização.

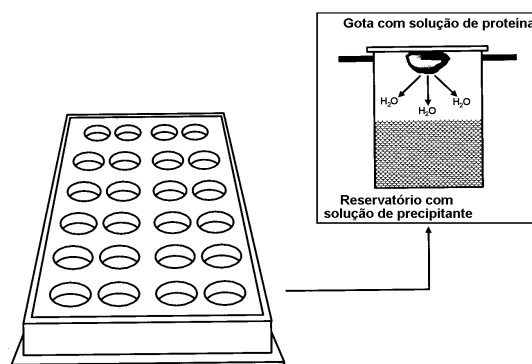


Figura 6-7: Placa comumente utilizada nos experimentos de cristalização. Um dos poços foi amplificado para mostrar como uma laminula de vidro é posicionada sobre o poço. Sobre essa laminula é colocada a gota com a solução de proteína.

Três variações do método são usadas na prática: com a gota pendurada, sentada e em sanduíche.

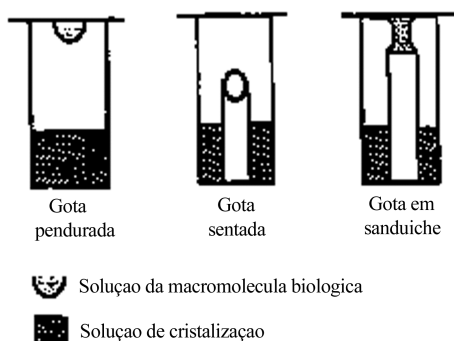


Figura 6-8: Variações na posição da gota no método de difusão de vapor: pendurada, sentada e em sanduíche.

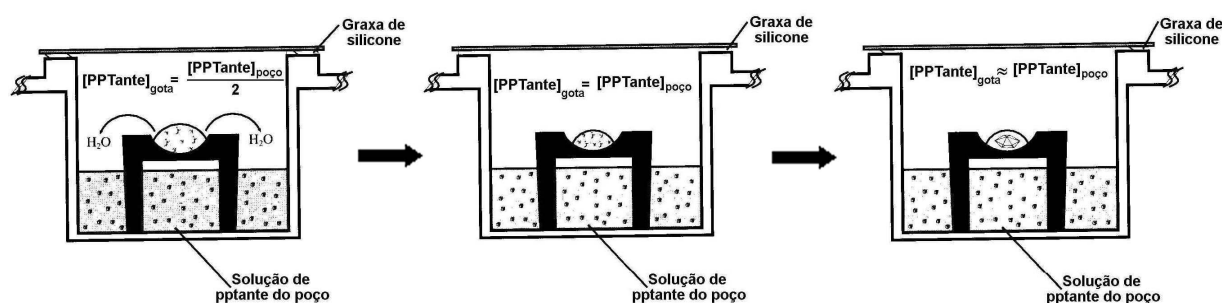


Figura 6-9: Como o processo de difusão de vapor induz a formação de cristais, exemplificação com o método da gota sentada. Como a gota é formada misturando-se o mesmo volume de solução de proteína e de solução do poço, a concentração de precipitante na gota será a metade daquela presente no poço, o que induz a perda de água da gota a fim de igualar ambas as concentrações de precipitante. Conforme a gota perde água a concentração de proteína aumenta, levando a uma condição ótima à formação de cristais de proteína.

5.2. DIÁLISE

Nessa técnica uma solução de proteína é separada dos agentes cristalizantes por uma membrana semipermeável. A concentração da solução de proteína permanece constante, porém o estado de supersaturação é atingido através da difusão de componentes de baixo peso molecular pela membrana. Essa técnica é indicada para cristalizações em condições de baixa força iônica. A vantagem sobre a difusão de vapor é sua reversibilidade, se a proteína precipita ou forma cristais de forma indesejável, esses podem ser redissolvidos por diálise contra uma solução sem precipitante ou com uma concentração menor deste. Por outro lado, um inconveniente da técnica é que precipitantes de peso molecular grande não podem ser usados. (MCPHERSON, 1994)

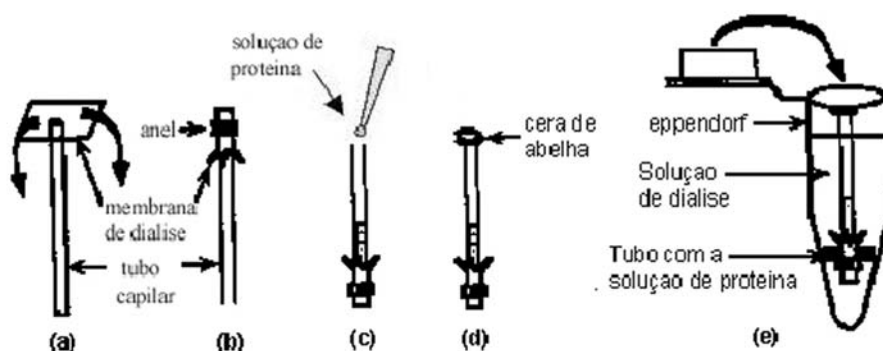


Figura 6-10: Arranjo montado para se fazer diálise num tubo eppendorf: (a) e (b) prende-se uma membrana adequada num capilar com um anel de borracha; (c) introduz-se a solução de proteína nesse capilar; (d) veda-se com cera a extremidade do capilar que não contém a membrana; (e) coloca-se esse arranjo num tubo eppendorf contendo a solução de cristalização contra a qual a diálise será feita. Esse método é usado para facilitar a montagem do cristal num capilar pois o cristal já cresce dentro dele. É um método mais antigo.

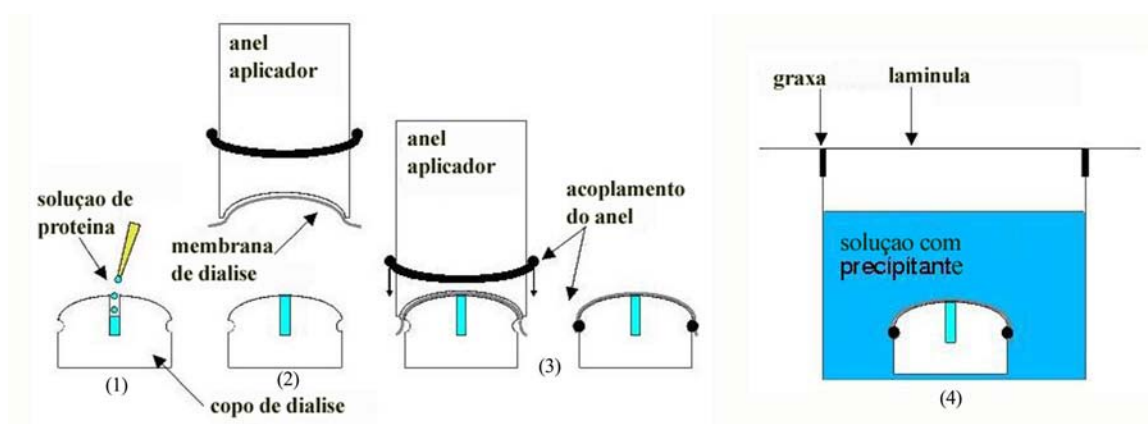


Figura 6-11: Arranjo para diálise usando uma placa comercial própria. Esse método é mais moderno.

5.3. TÉCNICA DE BANHO

Nessa técnica, todos os componentes para a cristalização são colocados em uma única solução juntamente com a solução de proteína preparada em uma concentração elevada, de modo que a condição inicial parta do estado de supersaturação. O sistema então é isolado e deixado em repouso e a proteína cristalizará espontaneamente com o tempo. Pode-se aplicar um gradiente de temperatura, por exemplo, se a proteína for mais solúvel a temperaturas mais altas. Nesse caso, aquecendo-se a solução e permitindo seu resfriamento lento pode-se induzir a formação de cristais. (MCPHERSON, 1994)

O atrativo dessa técnica é a sua simplicidade e, geralmente, é utilizada para a produção de cristais quando já são conhecidas as condições de cristalização, pois se necessita de uma quantidade elevada de proteína – utiliza-se cerca de 10-50µL de uma solução da proteína numa concentração de 100mg/mL.

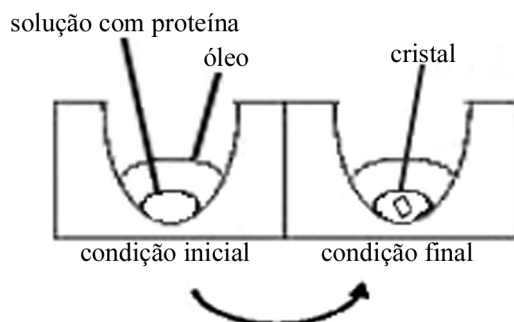


Figura 6-12: representação esquemática para a técnica de banho.

6. EXPERIMENTOS DE CRISTALIZAÇÃO

Tendo decidido sobre quais parâmetros podem influenciar na solubilidade da proteína, existem duas maneiras de iniciar os experimentos de cristalização. Uma delas consiste em selecionar um pequeno número de valores para cada parâmetro envolvido e então “varrer” todas as combinações possíveis para esses valores, mas, nesse caso, o número de experimentos aumenta drasticamente com cada novo parâmetro. Uma outra maneira seria usar uma abordagem estatística e neste caso faz-se uma hipótese de que o efeito de cada parâmetro envolvido é linear e então se podem extrapolar os resultados obtidos. (BERGFORDS, 1999)

Como o número de variáveis que afetam a cristalização é grande e combinatório, o número de condições para soluções a serem testadas é muito grande. Um meio de evitar este problema é usar o método de fatorial incompleto no qual uma matriz de condições de cristalização é explorada e os resultados são analisados para construir matrizes esparsas para refinar essas condições a fim de se obter resultados melhores. (BERGFORDS, 1999)

Existem várias empresas que comercializam kits para cristalização. Esses kits são compostos de uma série de soluções preparadas de acordo com testes estatisticamente comprovados sendo eficazes para uma série de proteínas. Essas soluções cobrem uma ampla quantidade de tamponantes, sais e precipitantes além disso, elas podem ser testadas a diferentes temperaturas. (BERGFORDS, 1999)

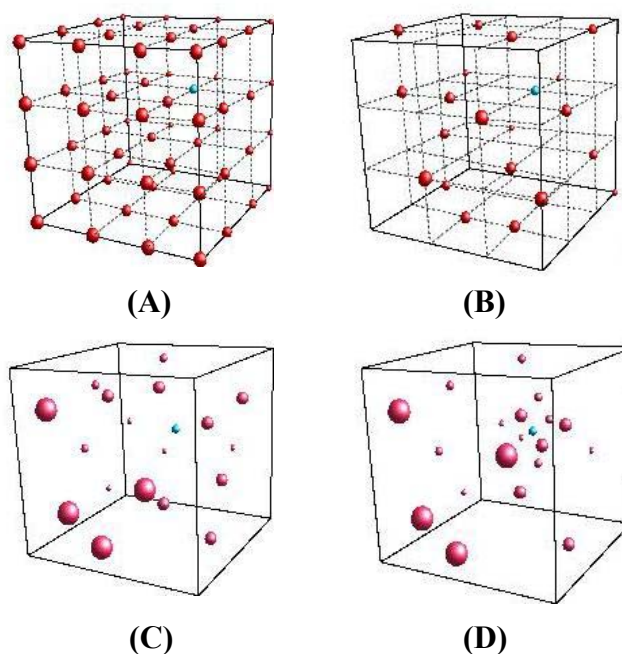


Figura 6-13: Matrizes que representam os métodos usados para varrer as condições de cristalização num experimento: (A) fatorial completo, no qual todas as condições seriam tentadas; (B) fatorial incompleto, no qual algumas condições estrategicamente escolhidas são testadas; (C) matriz aleatória, igual ao fatorial incompleto, mas com condições aleatoriamente escolhidas e (D) matriz esparsa, usado para o refinamento de condições nas quais foram conseguidos resultados prévios.

Após uma série inicial de experimentos com a finalidade de determinar as condições de cristalização, é bastante usual obter-se microcristais da macromolécula sob ao menos uma dessas condições. As condições nas quais foram obtidos os primeiros indícios da formação de cristais são refinadas para melhorar-se o tamanho e a qualidade dos cristais. Para isso parte-se da solução inicial e varia-se adequadamente o pH ou a concentração de precipitante. Em muitos casos, entretanto, a melhora da qualidade do cristal é difícil e várias estratégias diferentes devem ser adotadas. Em alguns casos, a adição de cofatores (para enzimas), contra-íons (para ácido-nucléicos e moléculas carregadas), metais essenciais como zinco, cálcio, etc (para proteínas que contém ou utilizam metais) tem provado ser bastante eficiente para a melhora da qualidade do cristal. (BERGFORDS, 1999)

7. RESULTADOS POSSÍVEIS

Os resultados dos experimentos de cristalização devem ser avaliados periodicamente com o uso de um microscópio ou de uma lupa. Existem alguns resultados tipicamente observados nas gotas num experimento de cristalização. Alguns exemplos mais comuns estão na figura 6-14.

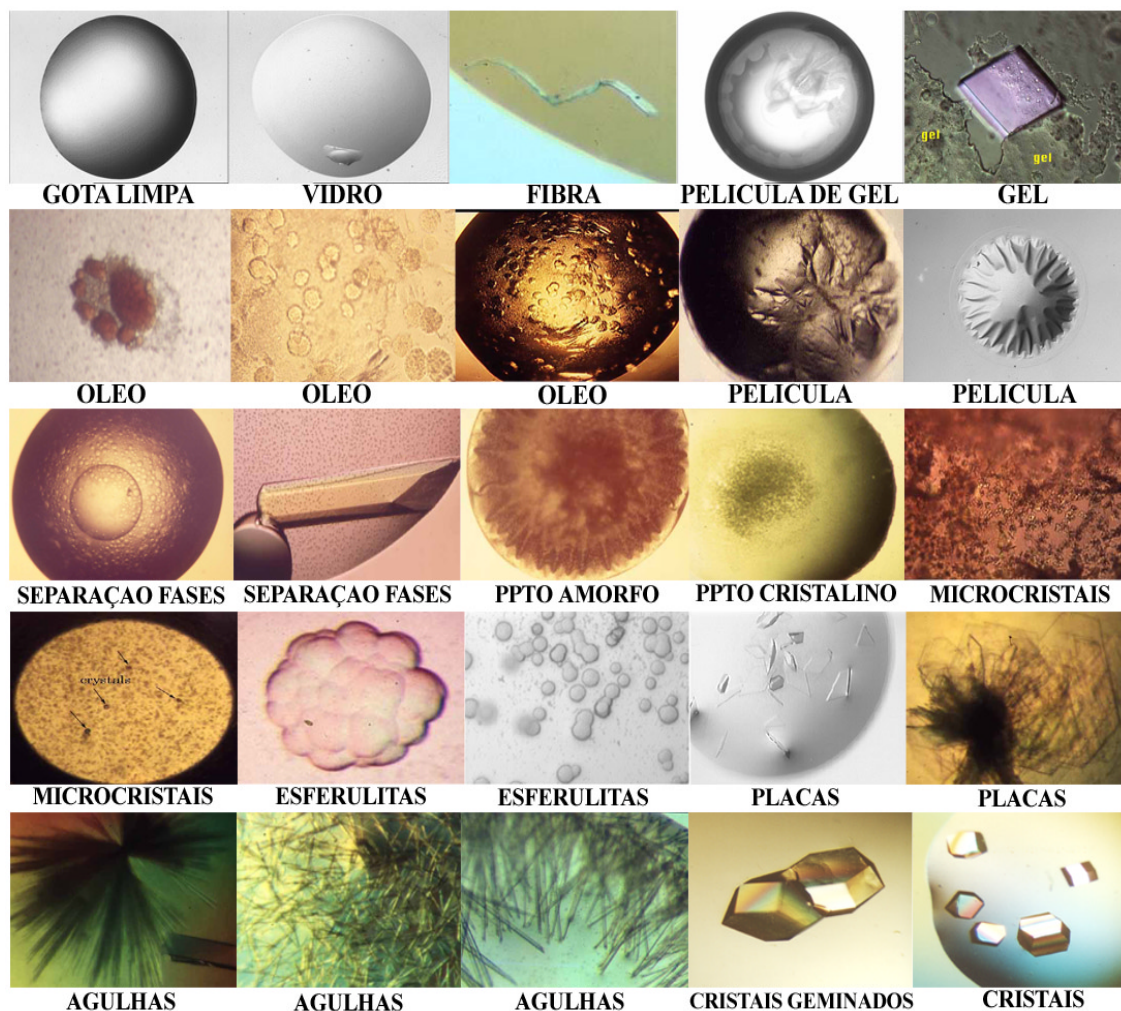


Figura 6-14: resultados tipicamente obtidos num experimento de cristalização.

Tem-se observado desde gotas completamente limpas até *sujeiras* ou *objetos estranhos* na gota, como vidro por exemplo, que tanto podem ser frutos de um descuido do pesquisador como podem ser arranjos formados pela própria proteína ou pelos outros compostos químicos presentes na gota. É o caso de fibras, camadas, gotas ou aglomerados de géis e óleos, películas e separações de fases que formam muitas vezes desenhos curiosos na gota. Embora não se saiba exatamente como são formados, eles podem servir como um

indício de que a proteína sofreu desnaturação, como é o caso de películas na superfície da gota, apesar de que polímeros como polietilenoglicol ou 2-metil-2,4-pentanediol, que são precipitantes comuns, também podem formar essas películas. Um outro resultado que indica desnaturação da proteína é a formação de um precipitado amorfo que é caracterizado por uma coloração marrom.

Dentre os resultados promissores são observados precipitados cristalinos, microcristais, esferulitas, agulhas, placas bidimensionais, cristais que não difratam, cristais que difratam a média resolução, cristais que difratam a alta resolução mas com simetria ou cela unitária desfavorável e cristais que difratam a alta resolução com simetria e cela unitária favoráveis. A partir da condição em que esses resultados foram obtidos pode-se, através de um refinamento, melhorar os resultados até que cristais adequados à difração sejam obtidos.

8. PREPARANDO OS CRISTAIS PARA A AQUISIÇÃO DOS DADOS

A partir do momento que cristais adequados aos experimentos de difração são obtidos, eles devem ser testados e se os resultados preliminares forem positivos, os dados de difração podem ser coletados. Seleciona-se um bom cristal e o primeiro passo é montar-se esse cristal. Esse passo deve ser feito cuidadosamente pois um bom cristal é fundamental para se coletar um bom conjunto de dados. Existem dois jeitos de se montar um cristal: usando um capilar, que é o método mais antigo, e usando um loop de raion, que é o método mais utilizado atualmente.

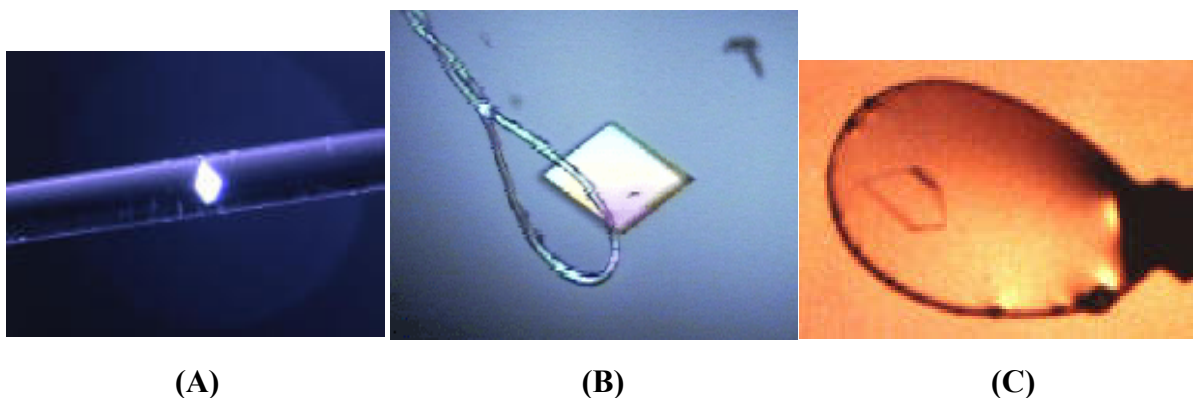


Figura 6-15: cristais de proteína montados em (A) capilar de vidro, (B) e (C) loops de raion.

O segundo passo consiste em se encontrar uma solução crio-protetora adequada para o cristal para impedir que haja a formação de gelo no cristal durante o congelamento no feixe de crio-nitrogênio.

A crio-cristalografia é agora amplamente usada pois permite um significativo aumento no tempo de vida de alguns cristais de proteína devido à redução dos danos provocados pela radiação durante a coleta dos dados. Em alguns casos também permite um aumento na resolução dos dados.

Os crio-protetores mais comumente usados são glicerol, etileno glicol, 2-metil-2,4-pentano diol (MPD) algumas vezes usado em combinação com polietileno glicóis de baixo peso (200-600), xilitol, eritrol, inositol, rafinose, trehalose, glicose, etc. O cristal montado no loop é mergulhado na solução crio-protetora e retirado rapidamente, somente para que o solvente crio-protetor forme uma camada na superfície do cristal. Alternativamente, para diminuir possíveis danos no cristal, o solvente crio-protetor é misturado ao líquido mãe em várias concentrações que são testadas até que uma concentração adequada seja obtida (geralmente em torno de 30-40%). (DUCRUIX & GIEGÉ, 1999)

9. PREPARAÇÃO DE DERIVADOS PARA A SUBSTITUIÇÃO ISOMÓRFICA

O método da substituição isomórfica (veja capítulo 8) tem sido central para a cristalografia de proteínas desde o trabalho inicial com a hemoglobina. Nessa técnica, alguns átomos de um metal pesado (com um grande número de elétrons) são introduzidos em cada macromolécula, como um adicional ou substituto aos átomos externos que circundam a macromolécula. Os metais pesados mais comumente usados são os compostos com Pt, Hg, U e Au. Esses compostos freqüentemente se ligam em sítios específicos na superfície da proteína sem causar mudanças conformacionais ou alterar o retículo cristalino, formando complexos isomorfos. Se o retículo cristalino é perturbado, os parâmetros da cela unitária usualmente mudam, indicando não-isomorfismo, o que inviabiliza o uso dessa técnica. Essa adição de átomos pesados na estrutura causa mudanças significativas nas intensidades do padrão de difração, as quais são então usadas para obter uma estimativa das fases para cada reflexão, e essas fases são usadas para o cálculo dos mapas de densidade eletrônica e a solução da estrutura. Embora a resolução máxima obtida para os dados do derivado seja geralmente menor do que para o cristal nativo, é possível fasear um conjunto

nativo até a sua resolução completa começando dos dados de baixa resolução dos derivados.

O cristal ao qual foram adicionados os átomos pesados é chamado de derivado e o cristal original sem átomos pesados é o cristal nativo. Os métodos tradicionais de preparação de derivados consistem em colocar o cristal nativo em uma solução de baixa concentração de átomos pesados por um período prolongado de tempo. Assim espera-se que apenas alguns sítios de ligação específicos sejam preenchidos pelos átomos pesados. Mas muitas vezes durante esse processo, os cristais são afetados sofrendo danos ou mesmo sendo destruídos. Embora a obtenção de um bom derivado seja essencialmente um processo de tentativa e erro, existem algumas considerações gerais que podem aumentar as chances de sucesso. (DUCRUIX & GIEGÉ, 1999)

Assim como para o cristal nativo, a preparação dos cristais derivados dependerá do pH, composição do líquido mãe e da temperatura, bem como da concentração e do tempo de permanência do cristal na solução de metal pesado.

pH

De uma forma geral, os valores de pH para obtenção de um bom derivado estão na faixa de 6-8. Se o pH estive muito abaixo de 6, a maioria dos grupos reativos os quais poderiam ligar os íons metálicos estarão protonados e bloqueados. Por outro lado, em pHs muito elevados, os íons metálicos tendem a formar hidróxidos insolúveis.

Temperatura

A temperatura pode mudar a taxa de ligação e em alguns casos o grau de ligação. De uma forma geral a diminuição da temperatura diminui a taxa de ligação. Na maioria dos casos a temperatura do banho com metais pesados deve ser a mesma da cristalização.

Composição do líquido mãe

Exceto em pHs abaixo de 6, sulfato de amônio é um líquido mãe não apropriado para a ligação de metais pesados devido à produção de NH_3 que é um bom nucleófilo e vai reagir com os metais pesados formando complexos. Assim, se for possível, os cristais devem ser transferidos para sulfatos de magnésio ou sódio ou para fosfatos de sódio e potássio. Entretanto, um excesso de fosfato não é bom para a ligação de urânio e metais terras raras pois também existe a tendência de formação de complexos.

Concentração

A concentração necessária depende da solubilidade do composto de metal pesado escolhido. Tipicamente, 12mM é um valor inicial apropriado.

Tempo de permanência do cristal na solução com metal pesado

O tempo de imersão pode variar de 20 minutos até meses, mas para uma tentativa inicial 4 a 18 horas são suficientes. O cristal é observado continuamente pelos primeiros 10 minutos, então de hora em hora para as primeiras quatro horas e novamente no final do processo. Se aparentemente o cristal não sofreu danos, e apenas poucas mudanças entre o cristal derivado e o nativo são encontradas na redução dos dados, é necessário aumentar a concentração de metal pesado e o tempo de imersão. Por outro lado, se o cristal quebra ou se a resolução de difração diminui dramaticamente, uma concentração e tempo de imersão menores devem ser tentados.

Não-isomorfismo

Os cristais não devem apresentar não-isomorfismo. Alguns aspectos devem ser cuidadosamente monitorados durante a coleta dos dados para o derivado. Os parâmetros da célula unitária não devem mudar (por exemplo mudanças de 0,5% na célula unitária correspondem a mudanças de 15% nas intensidades a uma resolução de 3Å). Um aumento na mosaicidade e no fator de temperatura também podem indicar a introdução de desordem no cristal e a presença de não-isomorfismo.

Selecionando o metal pesado

Muitos metais pesados podem ser usados para a obtenção de um derivado. A tabela 6-1 mostra alguns exemplos dos compostos mais comuns que podem ser usados como primeira escolha. Problemas de solubilidade no líquido mãe podem necessitar de uma escolha diferente. Complexos grandes podem dar um sinal forte mas também fornecem um Patterson mais difícil de interpretar.

Muitos compostos de metais pesados são bastante fotorreativos, por isso o processo deve ser feito em ambiente com pouca luz ou mesmo no escuro. Cobrir a caixa de cristalização geralmente é suficiente. Também se deve trabalhar com soluções de metal pesado recém-preparadas sempre que possível para evitar deterioração. Frequentemente muitos compostos de átomos pesados (entre 10 e 50) necessitam ser testados antes que um

bom derivado isomorfo seja encontrado. Esse estágio exige muito trabalho e uma boa dose de paciência. Uma outra maneira de introduzir íons de metal pesado é co-cristalizando a proteína e o composto de metal pesado, mas na prática isso não funciona bem pois os cristais são frequentemente não isomorfos. Pode-se ainda modificar quimicamente a proteína e produzir sítios geneticamente engenheirados. É o caso da introdução de selenometioninas (substituição de enxofre por selênio) em proteínas com alto conteúdo de resíduos com enxofre que são usados para MAD (do inglês *Multiple Anomalous Dispersion*), que é a técnica na qual é utilizado o sinal anômalo ou a diferença anômala entre o cristal nativo e o derivado ou o sinal advindo dos átomos pesados (enxofres) aderidos à macromolécula.

Tabela 6-1: Alguns exemplos de compostos de metal pesado que são comumente usados para a obtenção de um derivado para a substituição isomórfica.

Elemento	Elétrons	Reagentes	Alvos prováveis
Zr	40	Zr(NO ₃) ₄	
Pd	46	K ₂ PdCl ₄ , K ₂ PdBr ₄ , K ₂ PdI ₄ , PdCl ₂ , Pd(NO ₃) ₂	
Ag	47	AgNO ₃ , KAgCN ₂	
Sm	62	SmAc ₃ , Sm(NO ₃) ₃ , SmCl ₄	Substitui Ca ²⁺
Eu	63	EuAc ₃ , Eu ₂ O ₃	Substitui tampão fosfato
Gd	64	GdAc ₃ , Gd(NO ₃) ₂	Substitui tampão fosfato
Yb	70	YbCl ₃ , YbAc ₃	Substitui tampão fosfato
Re	75	ReCl ₃	
Pt	78	K ₂ PtCl ₄ , K ₂ PtCl ₆ , K ₂ PtI ₆ , K ₂ Pt(NO ₂) ₄ , Pt(NH ₃) ₂ Cl ₂	Substitui tampão Tris
Au	79	KAu(CN) ₂ , NaAuCl ₄ , KAuCl ₄ , KAuI ₄	His, Cys
Hg	80	EtHgCl ₂ , Hg(CN) ₂ , PhHgAc, HgCl ₂ , HgAc ₂ , HgSO ₄	His, Cys
Pb	82	PbAc ₂ , PbCl ₂ , Pb(NO ₃) ₂ , MePbAc	
U	92	UO ₂ Ac ₂ , K ₃ UO ₂ F ₅ , UO ₂ (NO ₃) ₂ , UO ₂ SO ₄	Substitui tampão fosfato

10. O MÉTODO DE QUICK CRYOSOAKING

Recentemente um novo procedimento para a obtenção de derivados, chamado de “quick cryo-soaking” (DAUTER *et al.*, 2000) foi proposto. De acordo com esse método, um rápido mergulho do cristal nativo (cerca de 15–300 segundos são suficientes) numa solução crio-protetora contendo ânions brometo e iodeto, antes de congelar o cristal no feixe de nitrogênio criogênico, leva à incorporação desses espalhadores anômalos nas

regiões ordenadas de solvente ao redor das moléculas de proteína. Os haletos promovem uma forma rápida e conveniente para a derivatização de cristais. Os haletos são ânions monovalentes que irão se ligar preferencialmente às áreas positivamente carregadas da superfície da proteína, competindo com as águas. A derivatização com haletos pode ser usada com sucesso na solução rápida de estruturas de macromoléculas (DAUTER *et al.*, 2000). Além dos haletos, também podem ser usados cátions monovalentes, como Cs^{1+} e Rb^{1+} , e cátions polivalentes, como os lantanídeos (Gd^{3+} , Eu^{3+} , Sm^{3+} ou Ho^{3+}), com o mesmo sucesso (NAGEM *et al.*, 2001). Os cátions se ligam preferencialmente aos sítios negativos da superfície das macromoléculas, de uma forma complementar aos haletos (NAGEM *et al.*, 2001). A concentração dos ânions e cátions é bem mais elevada em relação ao método tradicional de derivatização, utilizando-se em torno de 250mM–1000mM.

A vantagem mais óbvia desse novo método em relação ao tradicional é a considerável diminuição do tempo e do trabalho para a obtenção de um derivado útil para a substituição isomorfa, que é normalmente um dos passos mais demorados na determinação da estrutura. Além disso, também se pode controlar facilmente a deterioração do cristal que é visualmente monitorado durante todo o processo com os resultados sendo vistos rapidamente. Como a solução já possui agentes crio-protetores em sua composição, o derivado se encontra pronto para ser congelado e usado para a coleta. Na grande maioria dos casos, não se verificou não-isomorfismo.

Esse procedimento foi usado para a obtenção de dois derivados, um com cátions cério e outro com ânions iodeto, para a resolução da estrutura do inibidor extraído das sementes de *Copaifera langsdorffii* por substituição isomórfica.

AQUISIÇÃO E TRATAMENTO DOS DADOS DE DIFRAÇÃO

1. APARATO EXPERIMENTAL

O aparato usado nos experimentos de difração consiste de uma fonte de radiação e seus componentes óticos, um instrumento para posicionar e orientar a amostra, um sistema de detecção e um computador que controle a coleta. Além disso, quase todos os laboratórios de difração modernos usam um dispositivo de baixa-temperatura para resfriar a amostra (crio-cristalografia).

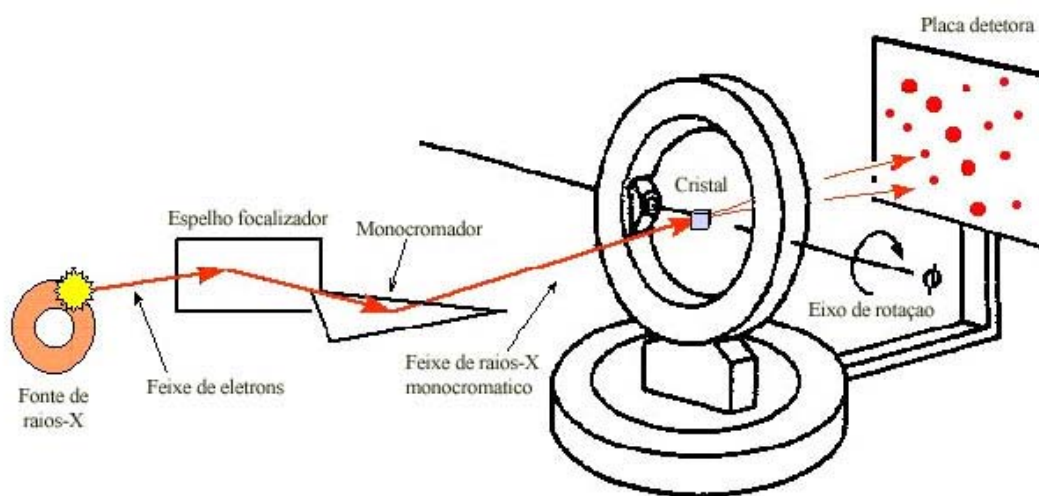


Figura 7-1: Esquema ilustrando os principais componentes do aparato experimental usado para a coleta dos dados de difração de um cristal. (Adaptado de <http://www-structure.llnl.gov/Xray/xrayequipment.htm>)

1.1. FONTES DE RAIOS-X

Raios-x característicos para os experimentos de difração podem ser produzidos por fontes convencionais, também chamadas *caseiras*, como um tubo selado ou um ânodo rotatório e por uma fonte de radiação síncrotron.

Em um tubo de raios-x selado, o ânodo é geralmente uma placa de cobre sobre o qual bombardeia-se um feixe de elétrons de alta energia promovendo-se emissões em comprimentos de onda característicos chamadas bandas de emissão. Uma dessas bandas é chamada $\text{CuK}\alpha$ e tem um comprimento de onda de $1.5418\text{\AA}$ sendo selecionada com um monocromador para os experimentos de difração. Uma das maiores dificuldades desse aparato é que a maior parte da energia do feixe de elétrons incidente é convertido em calor. O aquecimento produz três efeitos: enrugamento da superfície, fusão do alvo de cobre e estresse térmico provocado pela expansão diferencial do alvo de cobre na borda do ponto focal. Esse calor é removido pelo resfriamento do ânodo, geralmente com água. É justamente esse aquecimento do ânodo que limita o máximo poder do tubo. Esse limite é ampliado quando o ânodo é um cilindro rotatório ao invés de uma peça fixa de metal. O ânodo rotatório pode aumentar de 7–45 vezes o poder do tubo selado. Tem-se dessa maneira uma fonte pequena, 0,1–0,2mm, com uma *brilhança* muito alta.

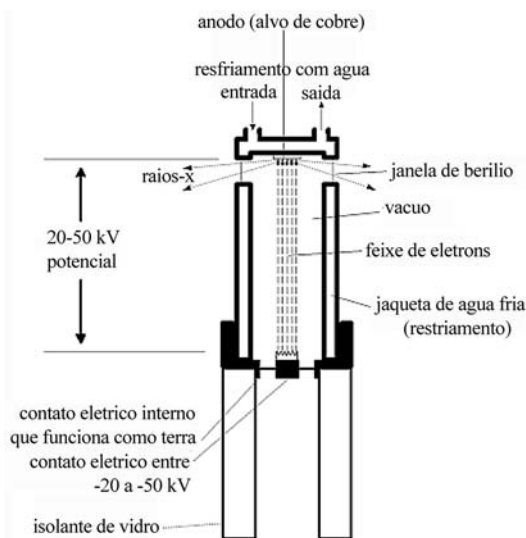


Figura 7-2: Esquema mostrando os principais componentes do ânodo rotatório.

Os raios-x numa fonte síncrotron são gerados por um feixe de elétrons que é gerado e passa inicialmente por um acelerador linear sendo então acelerado dentro de um anel até

atingir energias adequadas aos experimentos de difração. A trajetória curva dos elétrons dentro do anel é sustentada por ímãs magnéticos ou por dispositivos de inserção como *multipole wigglers* ou *undulators*. Do espectro da radiação síncrotron assim gerada pode-se selecionar, com um monocromador, qualquer comprimento de onda adequado. Suas propriedades podem ser usadas para os experimentos de difração que usam um único comprimento de onda até dispersão anômala a múltiplos comprimentos de onda (MAD), sendo especialmente úteis para a cristalografia de proteínas. Além disso, permite comprimentos de onda da ordem de 1Å ou menores, que apresentam menor absorção ao longo da trajetória até o cristal e dentro dele. A maior vantagem da radiação síncrotron é a alta intensidade que permite a coleta de dados com cristais muito finos, que difratam fracamente, ou cristais com célula unitária muito grande. Outra vantagem é que a radiação síncrotron é altamente polarizada. A polarização do feixe tem um efeito no espalhamento anômalo de raios-x dos átomos, o qual ocorre quando o comprimento de onda dos raios-x se aproxima da borda de absorção dos átomos (DRENTH, 1994).

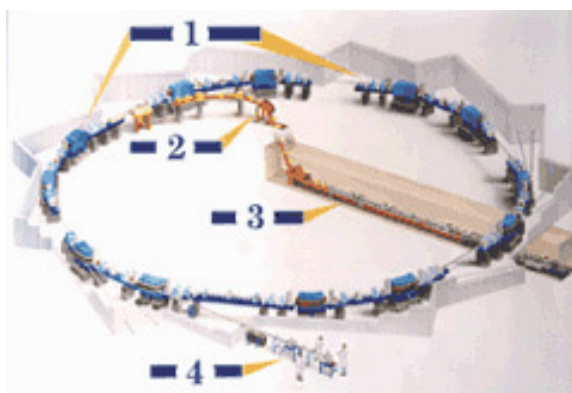


Figura 7-3: Esquema mostrando os principais componentes de uma fonte síncrotron de raios-x: (1) anel de armazenamento de elétrons; (2) linha de transporte (3) acelerador linear de elétrons; (4) linha de luz onde o feixe de raios-x é utilizado para os experimentos. (Figura retirada de <http://www.lnls.br>)

A fonte de raios-x usada para a coleta de dados de CTI e de DrTI foi a do Laboratório Nacional de Luz Síncrotron (Campinas, SP, Brasil) na linha de Cristalografia de Proteínas.

1.2. FILTROS

Os experimentos de difração de raios-x requerem que a energia da radiação, ou o seu comprimento de onda, seja limitado a uma banda tão fina quanto possível – radiação

monocromática ou com comprimento de onda único seria o ideal. A fim de se selecionar o comprimento de onda desejado de forma que preencha esses requisitos são usados os filtros e monocromadores.

Os filtros são usados no caso das fontes convencionais como o ânodo rotatório. Essa fonte possui duas linhas de emissão chamadas $K\alpha$ e $K\beta$. Para se confeccionar um filtro é escolhido um material cuja banda de absorção seja próxima à energia da linha do feixe de fótons que se quer absorver. Nesse caso, esse material reduz significativamente a intensidade da linha $K\beta$ em relação a $K\alpha$. A grossura do filtro é usualmente escolhida para reduzir a intensidade da linha $K\beta$ por um fator de 100 enquanto reduz a intensidade da linha $K\alpha$ por um fator de 10 ou menos.

1.3. MONOCROMADORES

Uma forma alternativa para produzir um feixe de raios-x com uma distribuição fina de comprimentos de onda é difratar o feixe incidente por um cristal cujas dimensões do retículo cristalino sejam muito bem conhecidas. Fótons de raios-x de diferentes comprimentos de onda são difratados a partir de um dado conjunto de planos desse cristal em diferentes ângulos de espalhamento de acordo com a lei de Bragg. Conseqüentemente, uma banda estreita de comprimentos de onda pode ser escolhida selecionando um ângulo de espalhamento particular para o cristal monocromador. Os cristais monocromadores necessitam ter as seguintes propriedades:

1. O cristal deve ser mecanicamente forte e estável no feixe de raios-x.
2. O cristal deve ter uma forte intensidade de difração num ângulo de espalhamento razoavelmente baixo para o comprimento de onda da radiação considerado.
3. A mosaicidade do cristal, a qual determina a divergência do feixe difratado e a resolução do cristal, deve ser pequena.

Uma variedade de geometrias é possível para cristais monocromadores. A maioria dos monocromadores é cortada com uma face paralela a um grande conjunto de planos cristalinos. Alguns monocromadores são cortados a fim de produzir um feixe difratado com

pouca divergência. Curvando-se o cristal monocromador apropriadamente, o feixe difratado pode ser focado em uma área muito pequena. Monocromadores curvos são geralmente reservados para aplicações especiais nos síncrotrons.

Cristais de grafite cortados na face (002) são os monocromadores mais comumente usados nos laboratórios de difração de raios-x. Outros materiais usados em monocromadores são germânio e fluoreto de lítio. Em todos os instrumentos disponíveis comercialmente que utilizam mono-cristal, o monocromador é colocado no caminho do feixe incidente. Cristais monocromadores alteram sistematicamente a polarização do feixe incidente e requerem diferentes correções geométricas.

1.4. COLIMADORES

Colimadores são objetos inseridos no caminho do feixe incidente ou difratado para dar forma ao feixe de raios-x. Tubos de metal são tipicamente usados. O raio interno do colimador é escolhido de forma a ser um pouco maior que o tamanho das amostras que ficarão imersas no feixe incidente. Colimadores do feixe incidente são construídos com duas regiões estreitas. A região próxima ao feixe de raios-x faz a colimação do feixe determinando o seu diâmetro. A segunda região tem um diâmetro um pouco maior que a primeira e é usada para remover a radiação *parasítica* que é devida à interação com as bordas da primeira região fina do colimador. Colimadores colocados no caminho do feixe difratado, funcionam apenas para remover qualquer radiação perdida impedindo que essa chegue ao detector.

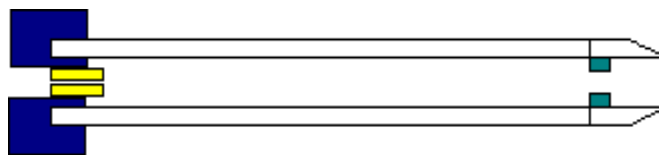


Figura 7-4: Esquema de um colimador típico.

Quando uma fonte pontual muito intensa e com diâmetro muito pequeno é necessária, como para cristalografia de proteínas, espelhos de raios-x devem ser usados para moldar o feixe incidente. Os espelhos são algumas vezes feitos com materiais que

atuam como filtros para a radiação em uso. Espelhos são primariamente usados em fontes de raios-x de intensidade muito alta como os anodos rotatórios ou síncrotrons.

1.5. GONIÔMETROS

Os goniômetros são dispositivos que posicionam ou rotacionam a amostra numa variedade de orientações.

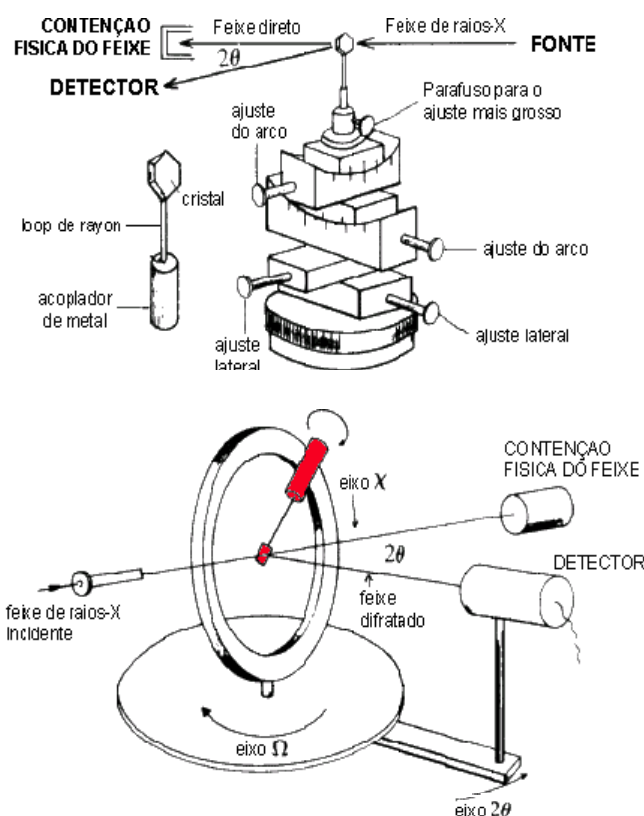


Figura 7-5: Esquema de um goniômetro. Acima estão as partes que compõem a cabeça goniométrica onde o loop contendo o cristal é fixado e abaixo está o goniômetro propriamente mostrando o posicionamento da cabeça goniométrica.

Esses dispositivos freqüentemente incluem um suporte para a fonte de radiação e quase sempre incluem um suporte para o detector. Os goniômetros usam dois ou três círculos para rotar a amostra em torno de um ponto fixo no espaço. Existe sempre um eixo de rotação no plano horizontal (ao redor do eixo vertical) chamado de eixo ω . Muitos

goniômetros incluem um movimento no plano vertical. Uma rotação sobre esse eixo é chamada de rotação χ . O eixo χ pode ser um círculo completo, algumas vezes um quarto de círculo ou um braço fixo. Todos os instrumentos incluem um movimento ao redor da cabeça goniométrica. Esse eixo é chamado de eixo φ . Além de movimentar a amostra, o goniômetro também movimenta o detector. Esse eixo é chamado de eixo 2θ quando um detector pontual é usado. Quando um detector de área é usado esse eixo é nomeado como o ângulo de movimentação do detector. Todos os movimentos da amostra e do detector devem ocorrer ao redor de um ponto fixo no espaço, então um microscópio ou uma câmera de vídeo amplificadora é providenciada para ajudar no posicionamento do cristal nesse ponto central do instrumento.

1.6. DETECTORES DE RAIOS-X

No experimento de difração de raios-x as intensidades de todos os raios difratados numa dada resolução devem ser medidos. Detectores comuns na cristalografia de pequenas moléculas são os contadores de cintilação 1D. Para medir as intensidades difratadas na cristalografia de proteínas os clássicos contadores simples e os filmes fotográficos foram sendo substituídos por detectores 2D muito mais rápidos, como as placas de imagem, câmera de raios-x com fósforo externo (ou detector de área), CCD sensível a raios-x ou as chamadas *multiwire proportional chamber (MWPC)*.

Placas de imagem são o tipo de detector mais usado, sendo muito populares devido a sua velocidade, sensibilidade, conveniência de uso e manutenção. São compostas por uma camada fina de um fósforo inorgânico que recobre uma base. Os fótons dos raios-x excitam os elétrons do material inorgânico para níveis mais altos de energia. Uma parte da energia é emitida como luz fluorescente normal na região de comprimento de onda visível, mas outra parte é retida no material por elétrons que ficam aprisionados em centros de cor. A placa de imagem é lida varrendo-a com um laser e medindo a luminescência emitida pelos centros de cor na região do azul. Elas têm uma ampla faixa de sensibilidade em relação ao comprimento de onda dos raios-x, o que proporciona alta eficiência de contagem em energias maiores (menores comprimentos de onda) e a vantagem para aplicação com radiação síncrotron. As intensidades das reflexões fortes e fracas podem ser coletadas com

apenas uma exposição. Após cada exposição a placa pode ser apagada expondo-a à luz branca intensa e usada novamente (DRENTH, 1994).

Raios-x difratados numa câmera de raios-x são absorvidos numa placa de fósforo e convertidos em luz visível. A imagem é então intensificada e focada ou num tubo de televisão sensível ou numa CCD para luz visível. Detectores desse tipo têm alta eficiência de absorção, alta taxa de contagem e leitura rápida, embora a faixa de intensidade dinâmica seja limitada em alguns casos.

As chamadas CCD de depleção profunda são sensíveis aos raios-x e detectam diretamente os feixes difratados. Essa é uma tecnologia nova e promissora a qual apresenta características similares a câmera de raios-x e, embora tenha uma menor eficiência de absorção, o fator de distribuição de pontos é melhor.

Os MWPCs contêm um grande número de fios que formam o ânodo que atuam independentemente com contagem proporcional. O cátodo geralmente consiste de dois conjuntos de fios onde o pulso positivo é induzido.

O detector usado para a coleta dos dados de CTI e DrTI foi uma placa de imagem MAR345.

2. ESTRATÉGIAS PARA A COLETA E TRATAMENTO DOS DADOS

Todos os dados simétricos únicos devem ser coletados para o cristal. Dados simétricos únicos de Laue são todos os dados absolutamente necessários para estruturas de cristais centrossimétricos. Dados relacionados de Friedel são também necessários para qualquer amostra que cristaliza num grupo espacial não centrossimétrico. Sempre se devem coletar medidas redundantes ou repetidas para cada reflexão (hkl) única. Medidas repetidas aumentam significativamente a qualidade dos dados e adicionam um pequeno esforço extra quando a coleta é feita usando um detector de área.

O fluxograma da figura 7-6 descreve um procedimento geral usado para coletar dados. Como visto no capítulo anterior, a seleção de um cristal adequado e o alinhamento desse cristal no instrumento devem ser feitos cuidadosamente a fim de se obter os melhores resultados para os dados.

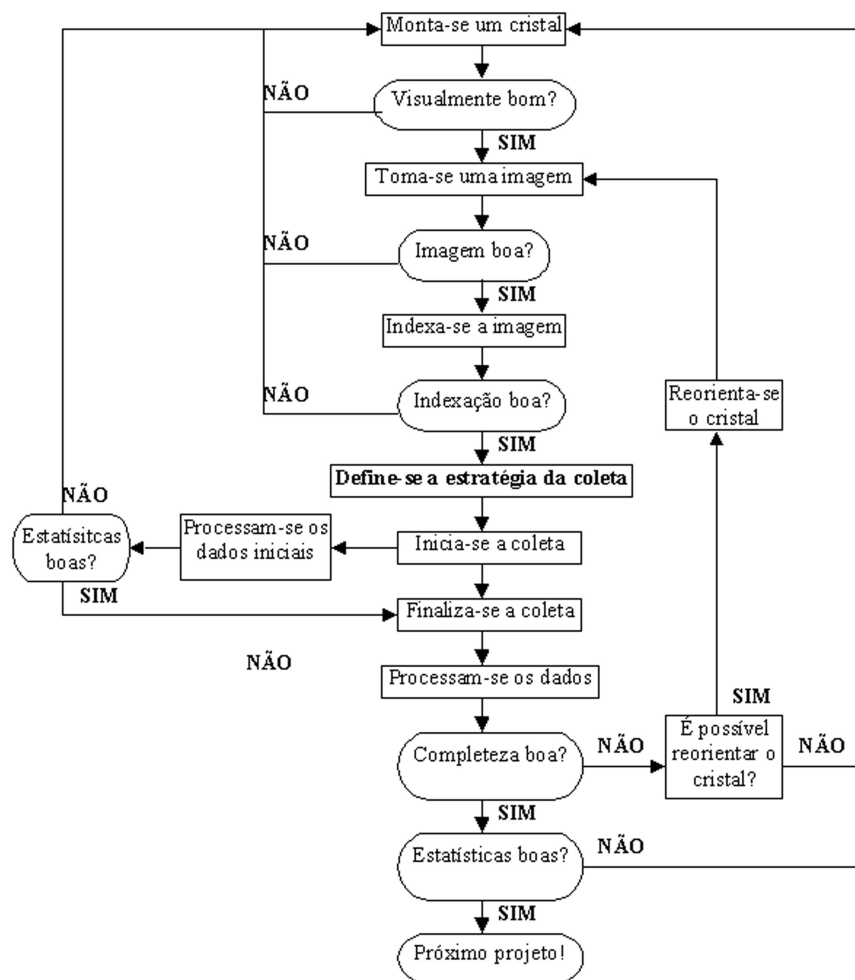


Figura 7-6: Fluxograma com as etapas e direções que definem as estratégias para uma coleta de dados eficiente.

2.1. ORIENTAÇÃO E CHECAGEM INICIAL DO CRISTAL

O cristal montado como já descrito (ver capítulo 6) é então afixado no goniômetro (Figura 7-5). E o centro de massa do cristal é alinhado no centro dos círculos do goniômetro. Com o auxílio do goniômetro o cristal é adequadamente orientado no feixe de raios X.

Após orientar adequadamente o cristal, toma-se o seu padrão de difração. Frequentemente toma-se uma imagem inicial com um tempo de exposição relativamente curto. A partir dessa imagem vários aspectos podem ser checados a fim de se acertar a simetria do cristal, os parâmetros da sua célula unitária, a sua orientação e o limite de resolução. A partir dessas informações é definida a estratégia que irá maximizar a resolução e a completeza e assim determinam-se os parâmetros para a coleta das imagens de difração.

Individualidade do cristal: se o padrão de difração não for formado por luas bem definidas de pontos, isso indica que, provavelmente, o cristal está quebrado, é múltiplo ou geminado. Nesse caso, transladando o cristal ao longo do feixe pode-se encontrar uma região em que ele seja individual.

Mosaicidade: mais pontos do que o esperado para o intervalo de oscilação indicam uma alta mosaicidade do cristal. A mosaicidade pode ser corrigida mas isso deve ser checado após indexar a imagem.

Saturação dos pontos: pontos saturados são distinguíveis dos outros na tela do computador – eles aparecem coloridos na tela da placa de imagem e em preto na tela do CCD. Não deve haver mais do que uma pequena percentagem de reflexões saturadas na faixa de resolução de interesse.

Regiões de sombra: pode acontecer do equipamento bloquear parte da superfície do detector, freqüentemente mau posicionamento do beam-stop. Geralmente isso é obvio, mas no caso de uma exposição curta com relativamente poucos pontos pode ser necessário olhar atentamente para detectar uma região onde dados estejam faltando.

Razão sinal ruído: picos adequados ao background são necessários para um bom conjunto de dados. O escalonamento que o programa de visualização faz pode fazer uma imagem parecer adequada quando na verdade não é. Uma verificação nos valores de background pode revelar o problema – geralmente valores acima de 1000 para placa de imagem ou 5000 para CCD são inadequados.

Separação dos pontos: o sucesso na integração das reflexões requer separação suficiente entre elas. Como exemplo a figura 7-7 mostra uma pequena região de um padrão de difração contendo um tanto de pontos.

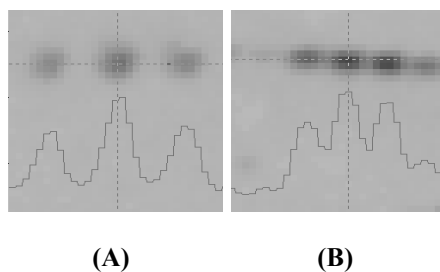


Figura 7-7: Uma pequena região de um padrão de difração contendo alguns pontos. É mostrado um gráfico com os valores de pixels de cada ponto. Em (A) a separação entre os pontos é de cerca de 10 pixels e em (B) essa separação é de 6 pixels.

Um gráfico dos valores de pixels ao longo das linhas pontilhadas horizontais é mostrado. A distância de separação requerida entre os pontos depende do tamanho do ponto, mas é tipicamente cerca de 10 pixels como na figura 7-7A. A separação de 6 pixels da figura 7-7B causará dificuldades na integração e deve ser evitada se possível, ou movendo-se o detector para trás ou diminuindo-se a faixa angular de oscilação.

Se houver um (ou mais) problema com o cristal e esse não puder ser resolvido, volta-se à etapa anterior escolhendo-se outro cristal. Uma vez que uma imagem visualmente satisfatória tenha sido obtida, o cristal deve ser rotado, geralmente de 90°, e outra exposição tomada. Os mesmos aspectos devem ser checados a fim de se verificar qualquer quebra, rachadura ou defeito do cristal que não tenha ficado evidente na primeira imagem e também qualquer problema de posicionamento do cristal. Se a segunda imagem também é boa, é o momento de indexar uma imagem.

2.2. INDEXANDO UMA IMAGEM INICIAL

Inicialmente uma parte dos pontos de difração é usada na auto-indexação. Esse processo é baseado na busca completa de todos os vetores possíveis no espaço real. Cada eixo (**a,b,c**) é independentemente indexado e três vetores linearmente independentes são utilizados na determinação de uma célula unitária. Essa célula é comparada com cada uma das 14 redes de Bravais e o erro associado ao desvio de cada rede é utilizado para a escolha da simetria da rede. Os parâmetros do cristal e do detector são refinados para se obter uma atribuição mais acurada das posições dos pontos de difração.

Isso pode ser facilmente feito usando o programa DENZO, parte do pacote de programas HKL (OTWINOWSKI & MINOR, 1993), sendo que os únicos parâmetros necessários são a posição do feixe direto e a distância do cristal ao detector. A partir da célula unitária determinada, o programa calcula as posições das reflexões e faz uma superposição com as experimentalmente medidas. Com a concordância entre os parâmetros com os dados experimentais é determinada a mosaicidade do cristal.

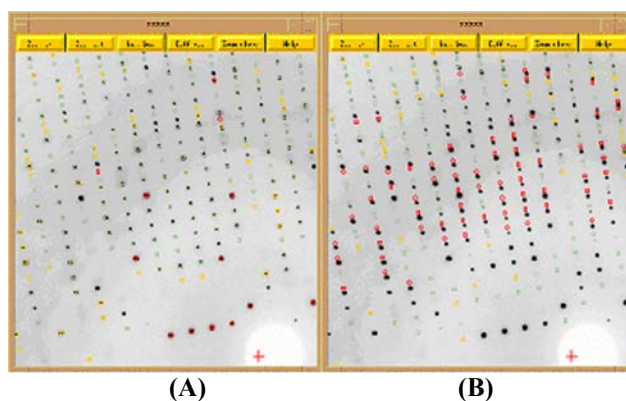


Figura 7-8: Porções de uma imagem de difração medida superposta com as posições das reflexões previstas. Os círculos coloridos em verde indicam reflexões completas, em amarelo as parciais e em vermelho as reflexões com problemas. Em (A) a indexação foi bem sucedida e em (B) a distância do cristal ao detector dada está incorreta.

2.3. COLETANDO OS DADOS DE DIFRAÇÃO PARA O CRISTAL

Procede-se então a coleta das imagens de difração e o método mais extensivamente usado é o método de rotação, também chamado método de oscilação. A fotografia de rotação ou imagem é tirada enquanto o cristal é rotado sobre um eixo. Para se coletar um conjunto de dados completo, as imagens devem ser tomadas até que o cristal tenha sido rotado através do espaço recíproco o suficiente para coletar todas as reflexões únicas. Para se definir o intervalo angular a ser coletado a fim de se obter a maior completeza dos dados pode-se usar o programa ESTRATEGY (OTWINOWSKI & MINOR, 1993). E a escolha do ângulo de rotação depende dos parâmetros já determinados: tamanho da célula unitária – quanto maior a célula unitária menor o ângulo de rotação – mosaicidade e resolução – desejando-se uma resolução maior diminui-se o ângulo de rotação (MCREE, 1993). O ângulo de oscilação é tipicamente um valor em torno de 1° .

Definido o ângulo de oscilação de cada imagem, a exposição pode ser definida por tempo ou por dose. Quando a fonte de raios-x usada produz um feixe de raios-x de intensidade constante (ânodo) a exposição por tempo pode ser feita sem problemas. Mas quando o feixe perde gradualmente a intensidade ao longo do tempo (síncrotron) a exposição por dose é mais adequada pois garante que a mesma quantidade de fótons atinja a amostra em todas as imagens coletadas.

2.4. TRATAMENTO DOS DADOS DE DIFRAÇÃO

Também chamado de redução dos dados, o processo de tratamento das medidas de difração envolve para cada máximo de intensidade uma integração do pico incluindo correções para a forma do ponto, a subtração da intensidade relativa do background, correções para a geometria do instrumento, correções para o decaimento do cristal e *merging* ou fusão dos dados.

2.4. INTEGRAÇÃO DOS DADOS

O primeiro passo envolve tanto a integração do pico quanto a subtração do background. Como as imagens são coletadas girando-se o cristal em intervalos angulares pequenos, uma parte das reflexões será parcialmente medida em imagens distintas. Para que essas reflexões parciais sejam convenientemente somadas é estimado um perfil para cada reflexão baseado nos perfis das outras reflexões. Então a intensidade do *background* ou ruído é avaliada e subtraída do perfil do ponto.

O background é formado pelo espalhamento do material no qual o cristal foi montado, espalhamento do ar, radiação de fluorescência do cristal ou do material no qual ele foi montado e radiação cósmica. Para se estimar o background toma-se uma imagem antes e outra depois que o cristal tenha sido montado ou então se considera simplesmente a área ao redor do pico.

Esse passo produz as intensidades I e seus respectivos desvios padrões $\sigma(I)$ estimados. Mas verifica-se que os desvios padrões, ou incertezas, assim calculados são subestimados em relação aos seus valores verdadeiros. Por isso devem incluir pequenos termos de correção devido à ‘instabilidade instrumental’. Esse processo de correção também é chamado de escalonamento dos dados.

2.4.2. ESCALONAMENTO DOS DADOS

Para um pequeno cristal completamente imerso num feixe uniforme de intensidade I_0 , a intensidade integrada é dada por:

$$I(hkl) = \frac{\lambda^3}{\omega} \left(\frac{e^2}{mc^2} \right)^2 \frac{V_{cristal}}{V^2} I_0 LPA |F(hkl)|^2 \quad [7.1]$$

A quantidade $e^2/mc^2=2,82\times10^{-13}$ cm é o raio clássico de um elétron, λ é o comprimento de onda da radiação incidente, V é o volume da célula unitária, $V_{cristal}$ é o volume do cristal, ω é a velocidade angular com que o pico atravessa a esfera de Ewald. Os termos de correção incluem a correção de Lorentz L , a correção de polarização P e a correção de absorção A .

Geralmente as constantes na expressão são combinadas e negligenciadas até o refinamento. Durante o refinamento, um fator de escalonamento geral entre os dados observados e calculados é aplicado. Dessa forma, a seguinte expressão simplificada é usualmente aplicada para calcular as amplitudes dos fatores de estruturas:

$$I(hkl) = |F(hkl)|^2 LPA \quad [7.2]$$

Fator de Lorentz:

Alguns picos, mais próximos ao eixo de rotação, gastam mais tempo para atravessar a esfera de Ewald do que outros. Essa diferença de tempo é corrigida por um termo chamado fator de Lorentz que leva em conta a diferença de tempo que cada plano de Bragg entra em condição de difração.

Fator de polarização:

Em diferentes ângulos de espalhamento o feixe espalhado será atenuado pela polarização do feixe pela amostra. Se a radiação incidente é polarizada no plano (orientação aleatória do vetor elétrico da radiação) então a correção de polarização é dada pelo fator $p = (1 + \cos^2 2\theta)/2$. Se um monocromador é inserido no feixe incidente, então os raios-x já serão parcialmente polarizados pelo cristal do monocromador e uma expressão diferente que dependerá da geometria do monocromador deverá ser aplicada.

Decaimento do cristal:

Durante a coleta é verificada uma perda de intensidade de difração devido aos danos sofridos pelo cristal durante a exposição aos raios-x. Se o cristal é congelado e os dados são coletados a baixas temperaturas (crio-cristalografia) então esse decaimento é desprezível.

Fator de absorção:

A absorção de raios-x pela amostra é frequentemente a correção mais difícil de fazer. A extensão da absorção depende do tamanho e forma do cristal bem como dos tipos e quantidades dos diferentes átomos na amostra e do comprimento de onda da radiação usada no experimento. Além disso a absorção do material no qual o cristal foi montado pode precisar ser incluída na correção. Os efeitos da absorção podem ser reduzidos montando-se o cristal adequadamente, usando um cristal pequeno ou usando radiação de alta energia. Existem quatro classes gerais de correções de absorção: analítica, empírica, geométrica e de Fourier. O método que usa correção analítica depende de uma cuidadosa indexação das faces do cristal. Esse método é acompanhado pela divisão matemática da amostra em pedaços muito pequenos e calcula-se a transmitância para cada pedacinho do cristal para cada reflexão medida. Ele é preferido para amostras que absorvem fortemente mas não corrige os efeitos de absorção do material no qual o cristal foi montado. O método empírico requer que os dados sejam coletados com alta redundância pois ele compara as intensidades das medidas coletadas diversas vezes e calcula uma superfície de absorção para a amostra. Esse método corrige tanto a absorção da amostra quanto do material no qual o cristal foi montado. Sendo originalmente desenvolvido para amostra de proteínas quando essas eram montadas em capilares de vidro, é o método de correção de absorção escolhido para a maioria das amostras de pequenas moléculas devido à sua simplicidade e sucesso. O método geométrico depende da forma da amostra que é aproximada para algum objeto geométrico conhecido, geralmente uma esfera ou cilindro, e a partir das dimensões do cristal e da transmitância da amostra correções aproximadas são aplicadas aos dados. O método de Fourier utiliza parâmetros de deslocamento isotrópico assumindo que qualquer diferença entre os dados calculados e observados é devida à absorção. Como no método empírico, uma superfície de absorção é calculada e a correção baseada nessa superfície é aplicada aos dados. Esse método não é recomendado porque ‘amassa’ os dados para que se ajustem ao modelo e assim qualquer informação interessante que não esteja incorporada ao modelo será perdida.

O escalonamento pode ser feito com o programa SCALEPACK (OTWINOWSKI & MINOR, 1993). Uma escala isotrópica simples é calculada para cada imagem e um fator de correção é aplicado a cada I e $\sigma(I)$ sendo dado por:

$$S = \frac{e^{2B\left(\frac{2\sin\theta}{\lambda}\right)^2}}{escala} \quad [7.3]$$

B é o fator de temperatura de Boltzman e λ é o comprimento de onda da radiação utilizada.

2.4.3. MERGING OU FUSÃO DOS DADOS

Então as intensidades das reflexões simetricamente equivalentes são somadas num conjunto de reflexões únicas num processo conhecido como *merging*. Uma medida do quão bem os dados se somam é dada pelo termo R_{merge} . O R_{merge} é uma medida da dispersão das reflexões simetricamente relacionadas e é matematicamente expresso por:

$$R_{merge} = \frac{\sum_{hkl} \sum_i \left| I_i(h,k,l) - \overline{I_i(h,k,l)} \right|}{\sum_{hkl} \sum_i I_i(h,k,l)} \quad [7.4]$$

Sendo i as reflexões simetricamente equivalentes.

O conjunto de dados resultante contém as intensidades I e os respectivos desvios padrões $\sigma(I)$ para cada reflexão (h,k,l).

CAPÍTULO 8

OS MÉTODOS DE FASEAMENTO: SUBSTITUIÇÃO ISOMÓRFICA (SIR/SIRAS/MIR/MIRAS), DISPERSÃO ANÔMALA (MAD), SUBSTITUIÇÃO MOLECULAR E MÉTODOS DIRETOS

1. O PROBLEMA DAS FASES

Como foi visto, quando as ondas são difratadas por um cristal, elas formam os pontos de difração. Cada ponto de difração corresponde a um ponto na rede recíproca e representa uma onda com uma amplitude e uma fase relativa.

O vetor – amplitude e fase ou, mais apropriadamente, o número complexo – representando o espalhamento total de um conjunto particular de planos de Bragg é chamado fator de estrutura e é comumente denotado **F**. Os fatores de estrutura para vários pontos na rede recíproca correspondem à transformada de Fourier da distribuição da densidade eletrônica na célula unitária do cristal. Uma propriedade muito conveniente da transformada de Fourier é que ela é reversível, ou seja, se aplicarmos uma transformada de Fourier inversa aos fatores de estrutura, teremos a distribuição da densidade eletrônica.

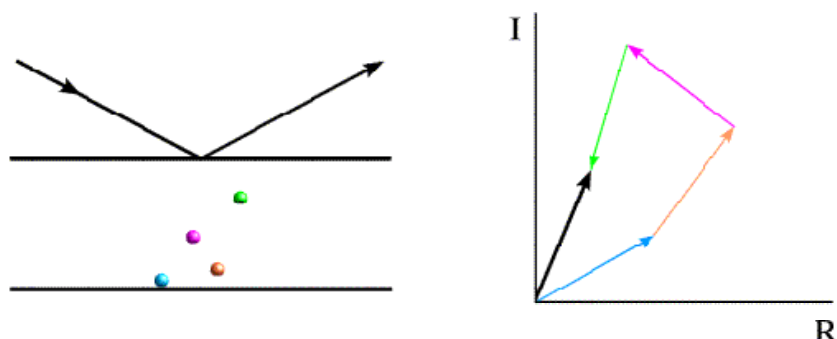


Figura 8-1: Dois planos de Bragg são mostrados junto com quatro átomos. A fase relativa (de 0 a 360 graus) depende da distância relativa dos átomos entre os planos que define um ângulo de fase de zero. Os átomos e suas respectivas contribuições para o espalhamento, representadas como vetores, são mostrados em cores correspondentes. A onda do espalhamento total é representada como o vetor preto, o qual é a soma dos outros vetores.

Assim bastaria medir o padrão de difração que contém as intensidades dos pontos de difração e calcular os fatores de estrutura que estão diretamente relacionados a essas intensidades $I(hkl)$ pela equação 7.3 já apresentada no capítulo anterior:

$$I(hkl) = |F(hkl)|^2 LPA \quad [7.2]$$

sendo L , P e A fatores de correção aplicados durante o processamento dos dados.

Entretanto, não se pode calcular o vetor $\mathbf{F}$ a partir das intensidades, mas apenas a sua amplitude $|\mathbf{F}|$. Se as fases fossem conhecidas seria possível calcular $\mathbf{F}$ e assim se poderia computar uma figura da molécula, mas essa informação é perdida no experimento de difração porque que se faz na prática é a contagem dos fótons que formam os pontos de difração. Os fótons são refletidos pelo cristal em diferentes direções com uma probabilidade proporcional ao quadrado da amplitude dessa onda e qualquer informação sobre a fase relativa dessa onda é perdida. Isso é conhecido como o problema das fases e uma grande parte da cristalografia é dedicada a resolvê-lo.

2. A IMPORTÂNCIA DAS FASES

Alguns cristalógrafos dizem que os dados medidos fornecem apenas a metade da informação necessária para calcular um mapa de densidade eletrônica. De fato, as fases são tão importantes quanto as amplitudes para se determinar a densidade eletrônica.

A figura a seguir dá uma idéia da importância da fase. No topo estão as fotografias de JEROME KARLE, à esquerda, e de HERB HAUPTMAN, à direita, que são os ganhadores do prêmio Nobel por seu trabalho na resolução do problema das fases para cristais de pequenas moléculas. Podemos tratar as fotografias como se fossem mapas de densidade eletrônica e calcular as suas transformadas de Fourier para obter as suas amplitudes e fases. (DRENTH, 1994)



Figura 8-2: A importância das fases na resolução de uma estrutura macromolecular.

Se combinarmos as fases da foto de HAUPTMAN com as amplitudes da foto de KARLE, teremos a figura embaixo à esquerda. A figura embaixo à direita combina as fases de KARLE com as amplitudes de HAUPTMAN. Como se vê claramente, as fases são dominantes.

Isso é particularmente perturbador quando se considera que, quando se usa a substituição molecular, *emprста-se* as fases de um modelo atômico previamente conhecido para se estimar as fases do modelo desconhecido. Felizmente já se têm evidências

suficientes, comparando-se modelos resolvidos por diferentes métodos, de que os modelos estão bem próximos da realidade, diferente dessas fotos. Ainda assim é preciso se preocupar com a imparcialidade do modelo.

3. RESOLVENDO O PROBLEMA DAS FASES: OS MÉTODOS DE FASEAMENTO

Existem quatro técnicas para resolver o problema das fases em cristalografia de proteínas e cada uma delas se aplica melhor a uma determinada situação:

(1) A técnica da substituição isomórfica, que é empregada para se resolver estruturas de proteínas *novas*, aquelas que não apresentam homologia seqüencial com nenhuma proteína cuja estrutura já seja conhecida. É necessário coletar-se um conjunto de dados de difração para um cristal nativo e para pelo menos um cristal derivado.

(2) A técnica de difração anômala a múltiplos comprimentos de onda, que também é usada para resolver estruturas novas. Nesse caso, é necessária a presença de átomos espalhadores na estrutura da proteína que produzam um sinal anômalo suficientemente forte.

(3) A substituição molecular, que é a técnica mais rápida para se determinar a estrutura de uma proteína, mas para a qual é necessário que essa proteína tenha uma homologia razoável com alguma proteína cuja estrutura seja conhecida.

(4) Métodos diretos, nos quais as fases são obtidas matematicamente de forma direta. São métodos que ainda estão num estágio de desenvolvimento para aplicação em proteínas mas que no futuro permitirão a rápida resolução do problema das fases em cristalografia de proteínas.

Em todas essas técnicas, a função de Patterson desempenha um papel fundamental e por isso primeiro essa função será apresentada e a sua interpretação física discutida.

4. A FUNÇÃO DE PATTERSON

Como mencionado, se aplicarmos uma transformada de Fourier inversa aos fatores de estrutura (amplitudes e fases) teremos uma figura da densidade eletrônica. Patterson perguntou-se o que resultaria se aplicasse uma transformada de Fourier às intensidades, o que requer apenas o conjunto de dados medidos. O mapa assim resultante, que é agora chamado função de Patterson ou mapa de Patterson, tem alguns aspectos interessantes e úteis.

A função de Patterson $P(\mathbf{u})$ ou $P(uvw)$ é uma somatória de Fourier com intensidades como coeficientes e sem ângulos de fase ou com todos os ângulos de fase iguais a zero. (DRENTH, 1994)

$$P(uvw) = \frac{1}{V} \sum_{hkl} |F(hkl)|^2 \cos[2\pi(hu + kv + lw)] \quad [8.1]$$

Ou de forma mais curta:

$$P(u) = \frac{1}{V} \sum_S |F(S)|^2 \cos[2\pi \mathbf{u} \cdot \mathbf{S}] \quad [8.2]$$

Para evitar confusão com as coordenadas x , y e z da célula real, são usados u , v e w na célula de Patterson, a qual entretanto tem dimensões idênticas a da célula real. Note que os coeficientes da somatória são $|F(hkl)|^2$ e não $|F(hkl)|$ porque todos os ângulos de fase são zero na função de Patterson e ela pode ser calculada sem qualquer conhecimento prévio da estrutura. Além disso, a função de Patterson também pode ser escrita como uma integral:

$$P(u) = \int_{r_1} \rho(r_1) \times \rho(r_1 + u) d\mathbf{v} \quad [8.3]$$

A integração é feita para r_1 em todas as posições na célula unitária real. Essa forma de $P(\mathbf{u})$ ajuda a entender a sua interpretação física: movendo-se através da célula unitária real com um vetor $\mathbf{u}$, multiplica-se para cada posição de $\mathbf{u}$ a densidade eletrônica ρ no início de $\mathbf{u}$ (na posição $\mathbf{r}_1$) e a densidade eletrônica no final do vetor $\mathbf{u}$ (na posição $\mathbf{r}_1 + \mathbf{u}$) e toma-se a integral sobre esses valores. Se o vetor $\mathbf{u}$ definir a distância entre dois átomos

começando no átomo 1 e terminando no átomo 2, então a função de Patterson multiplicará as densidades eletrônicas desses dois átomos, o que resultará num pico no mapa de Patterson correspondente à distância vetorial entre eles. Ou seja, um pico no mapa de Patterson na posição $\mathbf{u}$ (ou uvw) indica que na célula real existem átomos nas posições (x, y, z) e $(x+u, y+v, z+w)$ ou $(x-u, y-v, z-w)$. Embora as suas posições atômicas reais não sejam conhecidas, a distância vetorial entre eles é clara no mapa de Patterson. (DRENTH, 1994)

Para um número relativamente pequeno de átomos, é possível determinar as posições originais dos átomos que originaram os picos observados no mapa de Patterson. Isso é chamado deconvolucionar o mapa de Patterson. Mas torna-se rapidamente impossível deconvolucionar um mapa de Patterson para moléculas maiores. Se tiver N átomos numa célula unitária e a resolução dos dados for alta o suficiente, serão observados N picos separados no mapa de densidade eletrônica. No mapa de Patterson, cada um desses N átomos terá um vetor para todos os outros N átomos e assim serão N^2 vetores, sendo que N desses serão auto-vetores de um átomo para ele mesmo que se acumularão como um grande pico na origem mas, ainda serão $N^2 - N$ picos aleatoriamente espalhados pelo mapa de Patterson. Se N é um número pequeno, digamos 10, então existirá um número grande mas ainda plausível de picos de Patterson fora da origem (90 para $N=10$). Mas para valores maiores de N , 1000 por exemplo (o que está na faixa para cristais de proteínas), então serão 999.000 picos de Patterson fora da origem, o que torna inviável a deconvolução. Apesar disso, a função de Patterson é muito útil como parte de outros métodos para resolver estruturas. (DRENTH, 1994)

Se um limitado número de átomos pesados numa célula unitária grande deve ser localizado (como numa das etapas da substituição isomórfica ou da dispersão anômala), a função de Patterson é extremamente útil. Na substituição molecular, a função de Patterson desempenha um papel fundamental pois é pela superposição dos Pattersons da molécula conhecida com a desconhecida que a posição correta do modelo na célula unitária é determinada.

A função de Patterson tem as seguintes propriedades:

1. O mapa de Patterson tem picos nos pontos onde os vetores $\mathbf{u}$ terminam iguais aos vetores entre átomos na célula real;
2. Para cada par de átomos na célula real existe um único pico no mapa de Patterson;
3. Um mapa de Patterson é sempre centrossimétrico;
4. Eixos helicoidais na célula real correspondem a eixos normais na célula de Patterson;
5. Os elementos de simetria podem causar uma concentração de picos em certas linhas ou planos conhecidos como *linhas de Harker* ou *planos de Harker*;
6. A altura de um pico é proporcional ao produto dos números atômicos (=número de elétrons) dos átomos responsáveis pelo pico. (Isso é particularmente importante na localização dos picos dos átomos pesados na técnica da substituição isomórfica ou da dispersão anômala.)

As figuras 8.16 e 8.17 mostram seções de dois mapas de Patterson típicos.

5. A SUBSTITUIÇÃO ISOMÓRFICA

O método da substituição isomórfica pode ser considerado como uma pedra-angular na cristalografia de proteínas: ele foi o método usado para resolver o problema das fases para as primeiras estruturas de proteínas e ainda permanece, com poucas exceções (dispersão anômala), a única técnica disponível para resolver a estrutura de proteínas *novas*, com nenhuma ou muito pouca homologia com qualquer proteína cuja estrutura já se conheça. O método envolve as seguintes etapas:

1. Preparação de pelo menos um mas, preferencialmente mais cristais derivados além do cristal nativo para a proteína. (Um derivado = SIR de *Single Isomorphous Replacement* e mais derivados = MIR de *Multiple Isomorphous Replacement*).
2. Coleta e processamento dos dados de difração para o cristal nativo, bem como para cada um dos derivados. Todos os conjuntos de dados devem ser escalonados juntos (mesma escala).

3. Aplicação da função de Patterson para a determinação das coordenadas dos átomos pesados.
4. Refinamento dos parâmetros dos átomos pesados e cálculo dos ângulos de fases para a proteína.
5. Cálculo dos mapas de densidade eletrônica para a proteína.

A preparação de cristais nativos bem como de derivados é tratada no capítulo 6, o qual descreve a cristalização de macromoléculas biológicas. E a coleta dos dados de difração é tratada no capítulo 7, que descreve desde a escolha de um cristal adequado com testes preliminares e indexação, até a coleta das imagens de difração e a redução dos dados para a obtenção de um conjunto de dados de difração para o cristal.

O termo ‘derivado isomorfo’ indica idealmente um cristal nativo no qual algumas moléculas de água tenham sido substituídas por átomos com mais elétrons, sem qualquer alteração na estrutura da proteína ou do retículo cristalino em si. Daí o nome substituição isomórfica. A figura a seguir ilustra o efeito da adição de um átomo pesado na estrutura considerada.

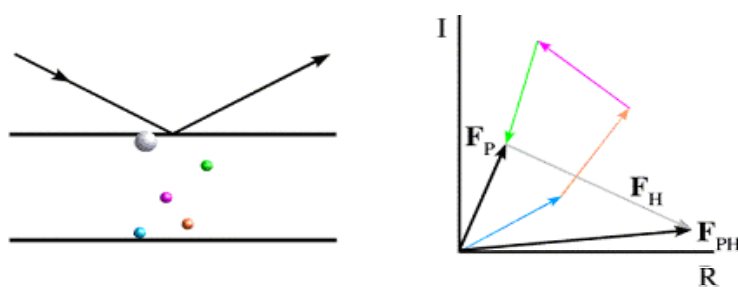


Figura 8-3: Mesma figura 8-1 com dois planos de Bragg e quatro átomos e suas respectivas contribuições para o espalhamento, representadas como vetores em cores correspondentes. A onda de espalhamento F_P representa a soma das contribuições dos quatro átomos nativos, F_{PH} representa os nativos mais a contribuição do átomo pesado, representado em tamanho maior à esquerda, e F_H representa a contribuição do átomo pesado isoladamente.

A introdução de um átomo pesado muda significativamente a intensidade do espalhamento. Como se pode ver na figura 8-3, as contribuições dos átomos leves tenderão

a se cancelarem porque eles espalham com diferentes ângulos de fase. Por outro lado, todos os elétrons no átomo pesado espalham essencialmente em fase. Por causa desse efeito, átomos pesados contribuem para a intensidade do espalhamento na proporção do quadrado do número de elétrons que eles contém. Por exemplo, um átomo de urânio contém 15 vezes mais elétrons do que um átomo de carbono então a sua contribuição para a intensidade será equivalente àquela de 225 átomos de carbono. Como resultado, a mudança na intensidade provocada pela adição de um átomo pesado à uma proteína pode ser facilmente medida.

Se assumirmos um isomorfismo perfeito entre o cristal nativo e o cristal derivado, as diferenças nas intensidades medidas refletirão exclusivamente a contribuição dos átomos pesados no espalhamento. Assim o fator de estrutura para o cristal derivado (F_{PH}) é igual à soma do fator de estrutura da proteína nativa (F_P) e do fator de estrutura dos átomos pesados (F_H), isso resulta na seguinte soma vetorial:

$$F_{PH} = F_P + F_H \quad [8.4]$$

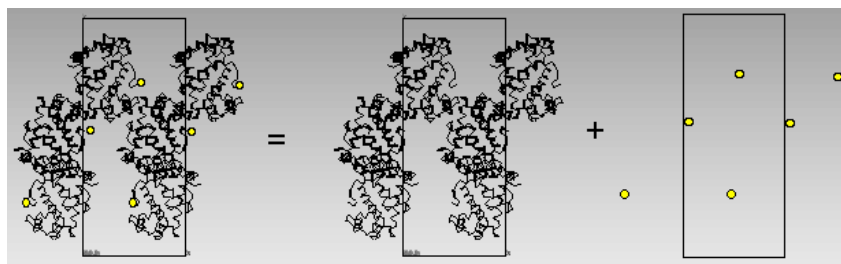


Figura 8-4: Derivado usado na substituição isomórfica.

Infelizmente, na ausência da informação das fases não se pode simplesmente aplicar a subtração vetorial $F_H = F_{PH} - F_P$ que resultaria na obtenção dos fatores de estrutura (com posições e fases) dos átomos pesados em um passo. Entretanto, as diferenças entre os fatores de estrutura do nativo e do derivado podem ser usadas para computar um mapa de Patterson. Como existem poucos átomos pesados, o mapa de Patterson será relativamente simples e de fácil deconvolução. Assim, uma vez sabendo onde os átomos pesados estão localizados no cristal, a sua contribuição nos fatores de estrutura pode ser ponderada e isso permite fazer algumas deduções sobre possíveis valores para os ângulos de fases da proteína.

5.1. A DETERMINAÇÃO DAS POSIÇÕES DOS ÁTOMOS PESADOS

Para as pequenas moléculas, as posições dos átomos pesados podem ser determinadas por meio de um mapa de Patterson calculado usando as amplitudes medidas para o derivado. O mesmo mapa não seria útil no caso de uma proteína pois a quantidade de átomos é tão grande que um mapa de Patterson calculado com os coeficientes do derivado será absolutamente ininteligível. Mas se a amplitude dos fatores de estrutura de cristais isomorfos são conhecidas, vários tipos de diferenças de Pattersons podem ser calculadas. A diferença de Pattersons que interessa é calculada usando como coeficientes $|F_{PH}^2 - F_P^2|$. O mapa assim obtido representa a diferença entre o Patterson do derivado e o Patterson da proteína nativa. A partir da equação 8.4: (GIACOVAZZO, 1992)

$$F_{PH}^2 - F_P^2 = (F_P + F_H)(F_P^* + F_H^*) - F_P F_P^*$$
$$F_{PH}^2 - F_P^2 = F_H^2 + F_H F_P^* + F_P F_H^* \quad [8.5]$$

O lado direito da equação 8.5 mostra que o mapa de Patterson calculado dessa maneira contém picos correspondentes aos vetores entre átomos pesados e átomos pesados (F_H^2) e aos vetores entre átomos pesados e átomos da proteína (termos misturados).

Entretanto é mais comum calcular-se um mapa de Patterson usando os coeficientes $(F_{PH} - F_P)^2$. Esse mapa é comumente conhecido como *Patterson da diferença isomórfica* e é mais representativo da situação dos átomos pesados. Em particular no caso centrossimétrico (que indica reflexões cujas fases são restritas a dois valores que diferem entre si por 180°), no qual F_{PH} e F_P são colineares, esses coeficientes representam uma estimativa verdadeira do valor de F_H : (GIACOVAZZO, 1992)

$$F_H = |F_{PH} \pm F_P| \quad [8.6]$$

Como em geral o valor de F_H é pequeno se comparado com F_P e com F_{PH} , a equação 8.7 será verdadeira.

$$F_H = |F_{PH} - F_P| \quad [8.7]$$

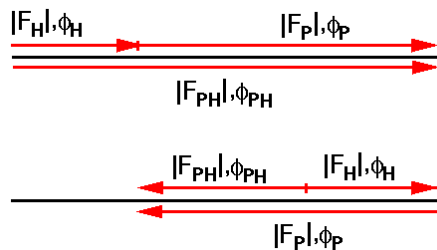


Figura 8-5: Diagrama de Argand ilustrando os fatores de espalhamento para reflexões cêntricas.

Em alguns casos, quando F_{PH} e F_P são ambos muito pequenos, pode acontecer que $F_H = F_{PH} + F_P$. Isso é chamado uma *intersecção* e, por efeito prático, pode ser removida do cálculo.

Para reflexões não-centrossimétricas (a grande maioria) a equação 8.7 não dá uma estimativa correta para F_H como já foi mencionado. A partir do triângulo da figura 8-6 a relação 8.8 pode ser escrita: (GIACOVAZZO, 1992)

$$F_H^2 = F_P^2 + F_{PH}^2 - 2F_{PH}F_P \cos(\varphi_P - \varphi_{PH}) \quad [8.8]$$

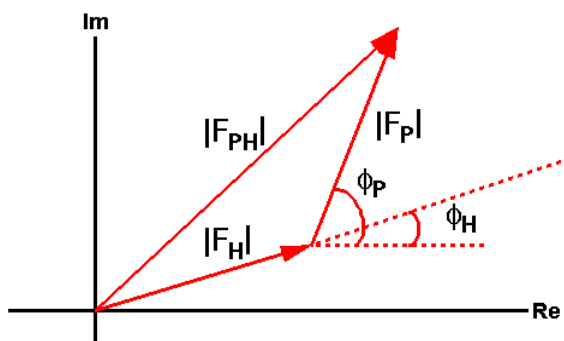


Figura 8-6: Diagrama de Argand que ilustra a relação entre os fatores de espalhamento do nativo, do derivado e dos átomos pesados para as reflexões acêntricas.

Se os ângulos dos fatores de estrutura da proteína nativa e do derivado são muito similares, como é o caso se os dois vetores F_{PH} e F_P são aproximadamente colineares, o

termo cosseno é muito próximo de 1 e a equação 8.8 fica igual à equação 8.7: (GIACOVAZZO, 1992)

$$F_H^2 \approx F_P^2 + F_{PH}^2 - 2F_{PH}F_P = (F_{PH} - F_P)^2$$

Na prática, um mapa de Patterson calculado com os coeficientes da equação 8.7 tende a representar o mapa da diferença de Patterson apenas para os átomos pesados para o caso não-centrossimétrico também e, pelo menos as primeiras posições para os átomos pesados são calculadas dessa forma.

Além da interpretação visual das seções de Harker dos mapas de Patterson, a localização dos átomos pesados também pode ser feita computacionalmente por programas. Isso pode ser especialmente útil quando a quantidade de átomos pesados na estrutura da proteína é grande, o que torna a interpretação do mapa uma tarefa muito complicada. A aproximação mais bem sucedida desenvolvida até agora para localizar um grande número de átomos pesados é o método do *espaço dual* implementado nos programas SHAKE-N-BAKE (BLESSING & SMITH, 1999) e SHELXD (SHELDRICK, 1992). Esses programas alternam ciclos de trabalho no espaço real com interpretação de Patterson e localização de posições atômicas teste e trabalho no espaço recíproco com refinamento por tangente. O programa SHAKE-N-BAKE foi usado com sucesso na localização das posições dos átomos de césio e de iodo no caso do inibidor extraído das sementes de *Copaifera langsdorffii*.

5.2. O REFINAMENTO DOS PARÂMETROS DOS ÁTOMOS PESADOS

Uma vez que um ou mais átomos pesados tenham sido localizados, o problema que surge imediatamente é o refinamento das suas coordenadas, fatores de temperatura e ocupância. O refinamento é dificultado pela falta de um valor correto de $F_{H(obs)}$, o que se tem é a estimativa do seu valor. Lembre-se que a estrutura dos átomos pesados é apenas imaginária formada por um cristal idêntico ao isomorfo mas contendo apenas os átomos pesados circundados por espaço vazio onde é preenchido pela proteína (como mostra a figura 8-4).

Historicamente, o primeiro método de refinamento dos parâmetros dos átomos pesados foi proposto por ROSSMANN (ROSSMANN, 1961) e trabalha com minimização usando o método clássico dos mínimos-quadrados (detalhes desse método são discutidos no próximo capítulo sobre refinamento): (GIACOVAZZO, 1992)

$$S = \sum w(F_{H(obs)} - F_{H(calc)})^2 \quad [8.9]$$

$F_{H(calc)}$ pode ser calculado normalmente a partir das posições dos átomos pesados mas, uma estimativa correta de $F_{H(obs)}$ é encontrada apenas para reflexões cêntricas, assim a princípio a equação 8.9 deve ser empregada para refinar apenas essas reflexões e os parâmetros refinados usados para calcular todas as fases. Entretanto, como a equação 8.7 pode ser considerada uma estimativa aproximada de $F_{H(obs)}$, a equação 8.9 é algumas vezes usada para refinar os dados acêntricos também.

Possivelmente, o método de refinamento mais usado e mais popular é o chamado refinamento de fases, introduzido pela primeira vez por MATTHEWS (MATTHEWS, 1968) no refinamento da mioglobina. A quantidade minimizada é: (GIACOVAZZO, 1992)

$$S = \sum w |F_{PH(obs)} - F_{PH(calc)}|^2 \quad [8.10]$$

A qual representa a minimização dos mínimos quadrados da diferença que é conhecida como *falta de fechamento*, a qual representa que o triângulo formado pelos vetores F_P , F_{PH} e F_H da figura 8-6 não fecha quando os erros são levados em conta (como mostra a figura 8-7).

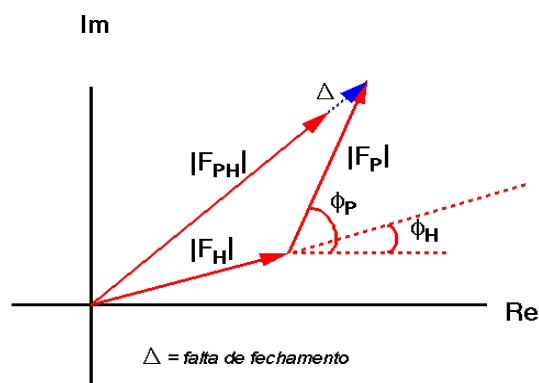


Figura 8-7: Diagrama de Argand mostrando o significado físico da falta de fechamento.

A *falta de fechamento* depende das fases da proteína, uma vez que:

$$F_{PH(calc)} = |F_{P(obs)} \exp(i\phi_P) + F_{H(calc)}| \quad [8.11]$$

Em geral w é escolhido como o recíproco do erro total E , calculado com a equação 8.12 para o derivado sendo refinado:

$$E_j = \left\langle \left(F_{PH(obs)} - F_{PH(calc)} \right)^2 \right\rangle \quad [8.12]$$

Qualquer que seja o método de refinamento selecionado, alguns procedimentos gerais podem ser seguidos: refinar apenas reflexões cêntricas pode ser seguro se elas representam uma boa porção do total. O número de variáveis também é importante: em geral, três parâmetros posicionais, a ocupância e um fator de temperatura isotrópico são variados para cada átomo mas, como os dois últimos parâmetros são altamente correlacionados, eles devem ser refinados separadamente. Se reflexões gerais são usadas, a fim de se evitar que as fases fiquem tendenciosas, é aconselhável refinar os parâmetros de um derivado omitindo-o do cálculo das fases, isto é, refinar um derivado e calcular as fases com todos os outros. Obviamente isso só é possível quando vários derivados diferentes estão disponíveis.

Uma das maiores dificuldades no refinamento dos átomos pesados é checar se os parâmetros usados são corretos. Isso é particularmente verdade no começo do refinamento, quando algumas vezes é difícil saber se a solução encontrada é correta ou não. Muitas quantidades podem ser usadas para monitorar o progresso do refinamento: nenhuma delas sozinha dá uma indicação absoluta, mas juntas elas podem ser consideradas uma checagem consistente. Diferentes tipos de fatores R foram definidos.

$$R_{Cullis} = \left(\sum |F_{PH} \pm F_P| - F_{H(calc)} \right) / \left(\sum |F_{PH} - F_H| \right) \quad [8.13]$$

$$R_{Kraut} = \left(\sum |F_{PH(obs)} - F_{PH(calc)}| \right) / \left(\sum F_{PH(obs)} \right) \quad [8.14]$$

$$R_{(módulo)} = \left(\sum |F_{PH(obs)} - F_{PH(calc)}| \right) / \left(\sum |F_H| \right) \quad [8.15]$$

O fator R de Cullis é calculado apenas para reflexões cêntricas. Um valor entre 0,40 e 0,60 é geralmente considerado bom, o único cuidado é um baixo número de reflexões usadas no cálculo. Para reflexões gerais, o fator R de Kraut é usado. Como o seu numerador é a quantidade sendo minimizada no refinamento, ele poder ser considerado um bom parâmetro para monitorar o curso do refinamento. Infelizmente um alto grau de substituição do derivado implica um R_{Kraut} estatisticamente grande e uma substituição muito baixa um valor pequeno para R_{Kraut} , tanto que ele não pode ser considerado uma indicação de que as variáveis sendo refinadas estão corretas.

Uma idéia da importância relativa de um único derivado no cálculo das fases é dada pelo $R_{(módulo)}$, o qual é a *falta de fechamento* dividida pelo módulo do fator de estrutura dos átomos pesados. Se o seu valor é maior do que 1, os círculos da figura 8-8 nunca irão se interseccionar de modo que nunca haverá uma solução para as fases da proteína. Se $R_{(módulo)}$ for calculado em camadas de resolução, servirá como um índice da utilidade do derivado em diferentes resoluções. Frequentemente o recíproco do $R_{(módulo)}$, o qual é chamado de *poder das fases* é avaliado. O seu valor deve ser maior que 1.

A última quantidade importante usada no refinamento é a *figura de mérito*:

$$m = N \int_0^{2\pi} P(\varphi_P) \exp(i\varphi_P) d\varphi_P \quad [8.16]$$

na qual N é o fator de normalização $N^{-1} = \int_0^{2\pi} P(\varphi_P) d\varphi_P$ e $P(\varphi_P)$ é a distribuição de probabilidade das fases φ_P . A figura de mérito dá uma estimativa direta dos erros nas fases, mas é bastante dependente do valor do erro total E sendo que uma super-estimativa de E dará uma figura de mérito super-otimista.

5.3. CALCULANDO AS FASES PARA A PROTEÍNA

Experimentalmente o que se mede são as intensidades para o nativo e o derivado, com as quais se calcula as amplitudes do nativo $|F_P|$ e do derivado $|F_{PH}|$.

Uma vez que as posições dos átomos pesados são razoavelmente bem conhecidas, os valores calculados para F_H podem ser usados para estimar os ângulos de fase dos fatores de estrutura para a proteína nativa. A do triângulo da figura 8-8 pode se escrever a equação 8.17 e derivar o valor para as fases da proteína, como mostra a equação 8.19: (GIACOVAZZO, 1992)

$$F_{PH}^2 = F_P^2 + F_H^2 + 2F_P F_H \cos(\varphi_P - \varphi_H) \quad [8.17]$$

$$\cos(\varphi_P - \varphi_H) = \frac{(F_{PH}^2 - F_P^2 - F_H^2)}{2F_P F_H} \quad [8.18]$$

$$\varphi_P = \varphi_H + \cos^{-1} \left[\frac{(F_{PH}^2 - F_P^2 - F_H^2)}{2F_P F_H} \right] \quad [8.19]$$

Como a equação 8.19 contém um termo cosseno, conseqüentemente existem duas soluções para φ_P . Essa ambigüidade pode ser mais bem ilustrada representando os fatores de estrutura para o nativo e o derivado no plano complexo como círculos de raios $|F_P|$ e $|F_{PH}|$ respectivamente, como na figura 8-8 à esquerda.

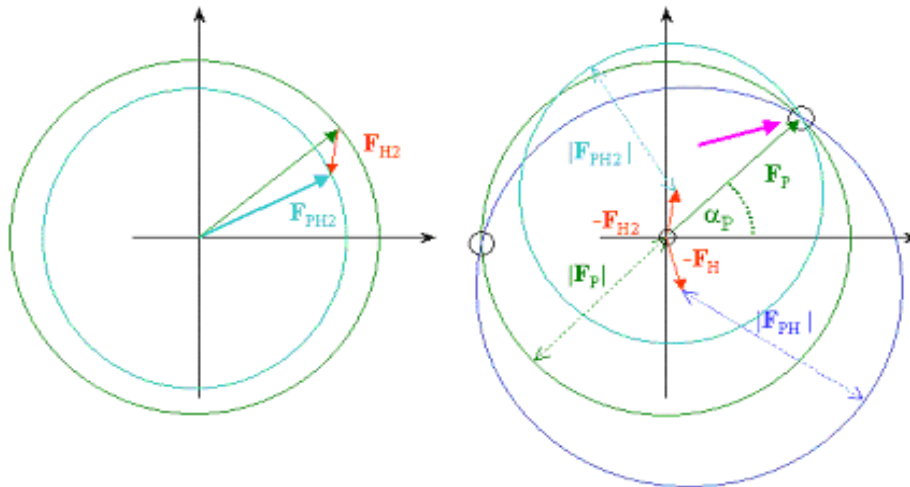


Figura 8-8: Representação dos vetores dos fatores de estrutura como círculos de raio F_P para a proteína e F_{PH} para o derivado; os fatores de estrutura para os átomos pesados são representados pelo vetor F_H . À direita os círculos são ponderados pelos valores de F_H e duas soluções são observadas para os possíveis valores das fases de F_P .

Como o valor de F_H é agora conhecido, esses círculos podem ser desenhados ponderados pelo vetor F_H e são obtidas duas soluções para os valores de fases possíveis para F_P , como indicado pelas setas na figura 8-8 à direita.

5.4. RESOLVENDO A AMBIGÜIDADE DAS FASES: A SUBSTITUIÇÃO ISOMÓRFICA MÚLTIPLA

A substituição isomórfica múltipla foi a primeira técnica usada para resolver essa ambigüidade das fases e ainda é a mais comum sendo amplamente adotada. Consiste na preparação de um segundo derivado que tenha átomos pesados ligados à proteínas em posições diferentes com relação ao primeiro derivado. Determinando-se as posições dos átomos pesados para os dois derivados com relação à mesma origem, calculando-se F_{H1} e F_{H2} e usando-se os coeficientes medidos para os dois derivados F_{PH1} e F_{PH2} , pode-se chegar à solução correta para o valor das fases para a proteína. O diagrama de Harker na figura 8-9 ilustra isso. Se os derivados forem perfeitamente isomorfos, os três círculos devem se interseccionar em um ponto único que indica a solução correta.

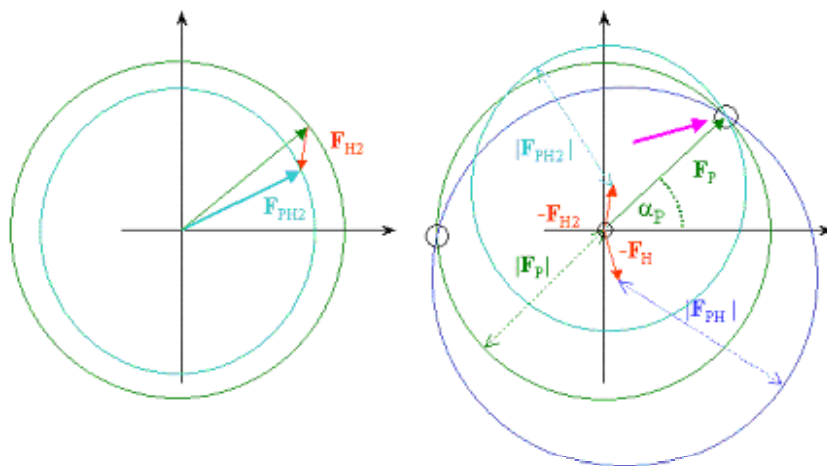


Figura 8-9: Agora um segundo derivado é introduzido e o valor correto para as fases de F_P pode ser obtido.

Essa construção gráfica representa o equivalente em resolver um par de equações como 8.11:

$$\varphi_P = \varphi_{H1} + \cos^{-1}[(F_{PH1}^2 - F_P^2 - F_{H1}^2)/2F_P F_{H1}] \quad [8.20]$$

$$\varphi_P = \varphi_{H2} + \cos^{-1}[(F_{PH2}^2 - F_P^2 - F_{H2}^2)/2F_P F_{H2}] \quad [8.21]$$

Assim tem-se, pelo menos em teoria, um valor exato para as fases de F_P .

Mas a teoria assume que não existem erros no modelo para os átomos pesados no cristal derivado, nem nas amplitudes dos fatores de estrutura medidas e assume também que os cristais são perfeitamente isomorfos. O efeito dessas fontes de incertezas é distorcer os círculos, de forma que as regiões de sobreposição são muito mais difusas e se percebe muito mais ambigüidade. O resultado é que a solução não será necessariamente um ponto, e a presença de mais derivados pode ser necessária.

5.5. USANDO SINAL ANÔMALO PARA RESOLVER A AMBIGÜIDADE DAS FASES NA SUBSTITUIÇÃO ISOMÓRFICA (SIRAS/MIRAS)

Como apresentado nas seções seguintes, o espalhamento anômalo (também chamada dispersão anômala ou sinal anômalo) pode ser usado numa técnica alternativa para resolver o problema das fases. Além disso, o sinal anômalo também é usado para determinar o enantiomorfismo ou a configuração absoluta da macromolécula. Mas uma das aplicações mais importantes para o espalhamento anômalo em cristalografia de proteínas é a possibilidade de resolver a ambigüidade das fases na substituição isomórfica.

Se o derivado apresenta sinal anômalo, no diagrama da figura 8-6 o vetor F_{PH} é desdobrado em F_{PH}^+ e F_{PH}^- , que são os chamados pares de Bijvoet (descritos na seção a seguir) como mostrado na figura 8-10.

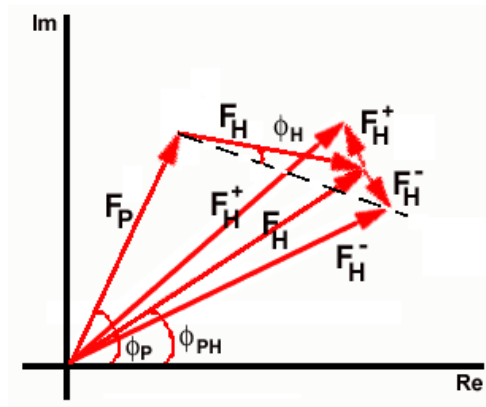


Figura 8-10: Diagrama de Argand que ilustra a relação entre os fatores de espalhamento do nativo, do derivado e dos átomos pesados, agora considerando a contribuição anômala que resulta nos pares de Bijvoet com o desdobramento de F_{PH} em F_{PH}^+ e F_{PH}^- .

A partir do diagrama da figura 8-10, a equação 8.22 pode ser escrita e resulta nos ângulos de fase para o derivado como mostra a equação 8.23.

$$F_{PH}^+ - F_{PH}^- = 2F_H'' \sin(\varphi_{PH} - \varphi_H) \quad [8.22]$$

$$\varphi_{PH} = \varphi_H \pm \sin^{-1}[(F_{PH}^+ - F_{PH}^-)/2F_H''] \quad [8.23]$$

A equação 8.23 permite novamente duas soluções possíveis para φ_{PH} mas, devido ao termo de seno a sua informação é complementar à da equação 8.20 e o ângulo de fase pode ser agora completamente determinado. Assim, usando tanto a substituição isomórfica quanto o sinal anômalo é possível determinar as fases sem ambigüidade, o papel do segundo derivado é suprido pela informação anômala.

5.6. O USO DO SINAL ANÔMALO NA DETERMINAÇÃO DA CONFIGURAÇÃO ABSOLUTA DA MACROMOLÉCULA.

Dois derivados isomorfos são em princípio suficientes para determinar o valor correto dos ângulos de fase dos fatores de estrutura da proteína mas, o método MIR apenas, não permite selecionar o enantiomorfismo correto para a proteína. Uma vez que o sistema de mão-direita, por exemplo, tenha sido escolhido na indexação da rede recíproca, o arranjo inicial dos átomos pesados para o primeiro derivado é arbitrário. Em outras palavras, a fase de F_H pode ser φ_H ou $-\varphi_H$. Uma vez que a mão tenha sido arbitrariamente selecionada para

o primeiro derivado, todos os outros devem ser consistentes com ele, o que não impede que todos eles sejam relativos ao enantiomorfismo errado. Nesse caso, o mapa de densidade eletrônica obtido representará a imagem especular da estrutura correta, o que significa todos os aminoácidos com a configuração D. Quando um mapa de densidade eletrônica a alta resolução pode ser produzido, alguns aspectos podem indicar que o enantiomorfismo errado foi escolhido, por exemplo, as α -hélices aparecerão no sentido errado (mão esquerda ao invés de mão-direita) e, nesses casos, o problema pode ser imediatamente corrigido recalculando-se as fases com os átomos pesados arranjados no outro sentido.

Entretanto, as diferenças anômalas no espalhamento podem ser usadas para identificar o enantiomorfismo correto desde o início. Como foi visto, pelo menos a princípio, um derivado com átomos pesados com a presença de sinal anômalo pode resolver a ambigüidade das fases para a proteína. Mas isso não é mais válido se duas possibilidades são tomadas em conta para os valores de φ_H , pois agora existirão dois possíveis valores para φ_P cada um associado com cada valor de φ_H . Dessa vez a ambigüidade surge no começo, desde que a escolha de φ_H foi arbitrária entre os dois valores possíveis φ_H e $-\varphi_H$ que definem a seleção do enantiomorfismo.

Mas se dois derivados com átomos pesados, e a presença de sinal anômalo para um deles, estão disponíveis, então se pode selecionar a configuração correta para a molécula. Usando as fases de um dos derivados com a contribuição anômala e os átomos pesados arranjados nos dois sentidos possíveis, dois mapas de diferença de Fourier do segundo derivado são calculados. Os picos no mapa de diferença calculado com os átomos nas posições corretas serão reforçados enquanto no mapa calculado a partir das posições erradas para os átomos pesados os picos serão enfraquecidos. E assim, determina-se o enantiomorfismo correto para a molécula.

6. DISPERSÃO ANÔMALA

Essa técnica é conhecida como dispersão anômala a múltiplos comprimentos de onda (MAD do inglês *Multiple Anomalous Dispersion*) e, para a sua aplicação, é necessário que a proteína contenha em sua estrutura elementos capazes de produzir sinal anômalo suficientemente forte. Esses elementos podem ser átomos pesados introduzidos através do processo de derivatização de cristais nativos já descrito ou podem representar átomos de

selênio introduzidos utilizando-se um substrato que contenha selenometioninas ao invés de metioninas, no caso de proteínas clonadas e expressas. Se a proteína tiver um alto conteúdo de átomos de enxofre (cisteínas, metioninas) em sua estrutura, pode-se usar o sinal anômalo produzido por esses átomos para se resolver a sua estrutura, mas isso depende bastante da qualidade dos dados.

Essa técnica explora a diferença na intensidade do espalhamento entre os pares de Bijvoet (BIJVOET, 1954), $|F_H(hkl)|^2$ e $|F_H(-h-k-l)|^2$, causada pela presença de espalhadores anômalos para determinar as fases para a proteína. Nessa técnica o comprimento de onda em que os dados são coletados depende do espalhador anômalo usado. Apesar do princípio desse método ser conhecido há muito tempo, somente com o desenvolvimento das fontes de radiação síncrotron nas quais pode-se mudar o comprimento de onda selecionado é que ele pode ser efetivamente utilizado para a determinação de estruturas de macromoléculas. Hendrickson (HENDRICKSON & TEETER., 1981) foi o primeiro a empregar esse método para resolver a estrutura de uma proteína.

6.1. A INTERAÇÃO DOS RAIOS-X COM OS ELÉTRONS ESPALHADORES

Como os átomos de H, C, N e O presentes nos cristais de proteína podem ser considerados leves, considera-se que os elétrons desses átomos espalhadores são elétrons livres, isto é, fracamente ligados aos núcleos atômicos. Entretanto, para um átomo mais pesado introduzido na estrutura da proteína os elétrons não podem mais ser considerados como livres. De fato, quanto mais pesado for o núcleo atômico mais fortemente ligados a ele estarão os elétrons, principalmente aqueles das camadas K e L mais internas. Devido a isso, esses elétrons apresentam um comportamento diferente no processo de espalhamento de raios-x.

Quando os raios-x atingem os átomos do cristal de proteína, o campo elétrico da onda eletromagnética induz uma oscilação nos elétrons. Se a frequência de oscilação da radiação incidente é muito diferente da frequência natural de oscilação dos elétrons, eles irão oscilar todos com a mesma fase. Isso equivale ao espalhamento causado por elétrons livres e é verdadeiro para a maioria dos elétrons no cristal. Mas se a frequência de oscilação

da radiação incidente é próxima à frequência natural de oscilação do sistema elétron-núcleo, então existirá um pequeno deslocamento tanto na amplitude quanto na fase da oscilação induzida. Esse é um efeito de ressonância e é verdadeiro para os elétrons de camadas internas nos átomos pesados introduzidos na estrutura da proteína. Esse deslocamento na amplitude e fase é chamado *espalhamento anômalo* e pode ser descrito como:

$$f = f_{ano} = f^0 + f' + if'' \quad [8.24]$$

A contribuição anômala no espalhamento consiste de duas partes: uma parte real f' e uma parte imaginária if'' . Quando não existe espalhamento anômalo esses componentes são desprezíveis ficando apenas f^0 que é o fator de espalhamento normal.

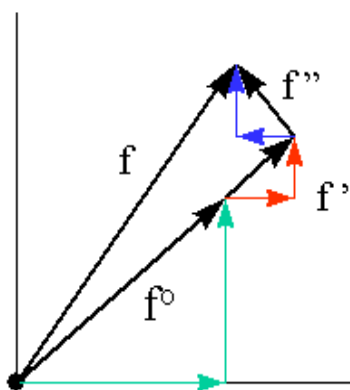


Figura 8-11: Diagrama com a representação das quantidades f, f^0, f' e f'' .

Os valores de f' e f'' são característicos para cada espalhador anômalo (átomo pesado) e com eles se determinam os comprimentos de onda adequados para a coleta dos dados.

6.2. ESCOLHENDO OS COMPRIMENTOS DE ONDA ADEQUADOS

Considerando a presença de apenas um tipo de átomo espalhador anômalo, necessita-se de no mínimo dois comprimentos de onda. Mas se três conjuntos de dados forem coletados em comprimentos de onda diferentes se terá maior chance de sucesso

durante o processo de faseamento. Entretanto, se formos por esse caminho, 4 é melhor do que 3, etc, mas pode ser que o tempo para a coleta de dados extras não justifique o melhoramento que eles podem trazer.

Os comprimentos de onda escolhidos devem produzir o maior efeito anômalo possível, assim o primeiro comprimento de onda escolhido será aquele que corresponde ao valor máximo de f'' e é representado por λ_1 na figura 8-12. O segundo comprimento de onda é geralmente escolhido para ter o máximo de $|f'|$, sendo representada por λ_2 na figura 8-12. Note que λ_1 e λ_2 são muito próximos requerendo uma grande precisão do equipamento de seleciona o comprimento de onda durante a coleta dos dados. Comprimentos de onda adicionais (λ_3 e λ_4) são escolhidos em pontos distantes da banda de absorção. O sinal aumenta lentamente com a distância a partir dos dois primeiros comprimentos de onda. Entretanto, a escolha desses comprimentos de onda é limitada pelas condições do equipamento usado. Tipicamente eles estão entre 100eV e 1000eV distantes da banda de absorção.

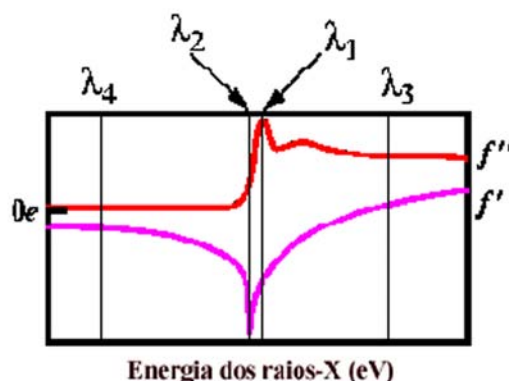


Figura 8-12: A escolha dos comprimentos de onda adequados é feita de acordo com a banda de absorção do átomo pesado usado.

6.3. O ESPALHAMENTO ANÔMALO QUEBRA A LEI DE FRIEDEL

A maioria dos elétrons nos átomos que compõem o cristal interage de forma idêntica com os raios-x, difratando com a mesma fase relativa. Devido a isso, pares de pontos de difração obedecem à chamada lei de Friedel, a qual é ilustrada na figura 8-13 à esquerda. Os chamados pares de Friedel são reflexões de Bragg relacionadas por uma

inversão através da origem. Segundo a lei de Friedel os membros de um par de Friedel têm amplitudes iguais e fases opostas.

$$|F_{hkl}| = |F_{\bar{h}\bar{k}\bar{l}}| \quad \varphi_{hkl} = -\varphi_{\bar{h}\bar{k}\bar{l}}$$

O deslocamento de fase no espalhamento anômalo leva à uma quebra na lei de Friedel, como ilustrado na figura 8-13 à esquerda. É conveniente representar o deslocamento de fase causado pelo átomo pesado, adicionando-se um vetor a 90 graus com o espalhamento normal. Como o termo f'' é sempre positivo e afeta da mesma maneira os pares de Friedel, o vetor está a +90 graus em ambos os casos o que provoca uma quebra na simetria. Se todos os átomos espalham igualmente, então as amplitudes permanecem iguais e as fases mudam.

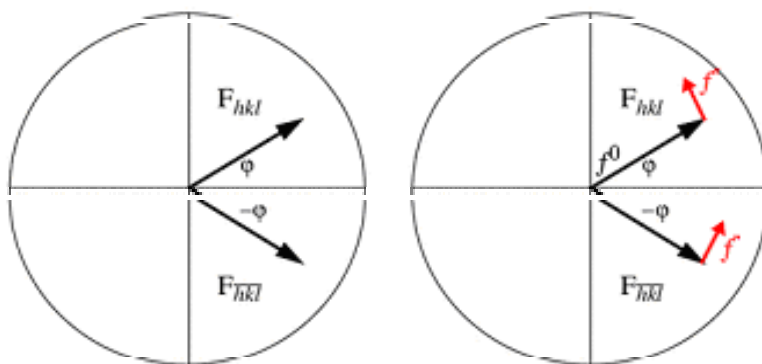


Figura 8-13: À esquerda estão os pares de Friedel para o espalhamento normal e à direita os pares de Bijvoet para o espalhamento que contém um componente anômalo.

Se alguns átomos espalham anormalmente e outros não, então tanto as relações das amplitudes quanto das fases são quebradas.

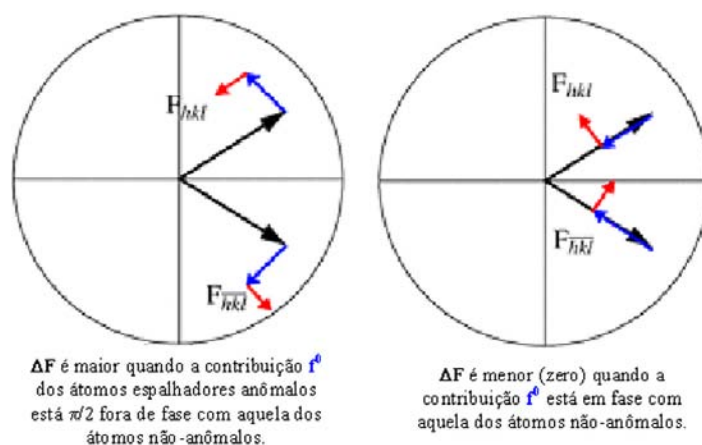


Figura 8-14: Aqui são mostrados os pares de Bijvoet e as contribuições de f^0 além de f'' .

Quando existe a quebra na lei de Friedel, os pares de Friedel recebem o nome de pares de Bijvoet.

6.4. OS PARES DE BIJVOET

Os pares de Bijvoet são reflexões de Bragg simetricamente equivalentes aos dois membros de um par de Friedel. Na presença de espalhamento anômalo, as verdadeiras reflexões de Bragg simetricamente equivalentes ainda têm amplitudes iguais. Os dois membros de um par de Bijvoet podem ter amplitudes diferentes, pelas mesmas razões que o par de Friedel.

$$|F^+| \equiv |F_{hkl}| \equiv |F_{\bar{h}\bar{k}\bar{l}}| \quad |F^-| \equiv |F_{\bar{h}kl}| \equiv |F_{h\bar{k}\bar{l}}|$$

A diferença na amplitude medida para um par de Bijvoet é chamada de diferença de Bijvoet ou ΔF .

$$\Delta F = |F^+| - |F^-| \quad [8.25]$$

Um mapa de Patterson calculado com coeficientes ΔF^2 contém apenas picos correspondentes aos vetores interatômicos entre pares de espalhadores anômalos. Esse é

freqüentemente o primeiro passo no faseamento por MAD pois a localização dos espalhadores anômalos é necessária para o desenvolvimento das estimativas das fases.

Se as fases estimadas para a estrutura são conhecidas, podem-se também usar as diferenças de Bijvoet para calcular um mapa de Fourier. Os coeficientes para esse mapa são (ΔF ; $\phi+90^\circ$). As fases usadas no calculo do mapa são 90° graus diferentes das fases estimadas para a estrutura como um todo pois a diferença de Bijvoet é calculada com a contribuição imaginária do espalhamento f'' total. Quando se adiciona 90° às fases, pode-se se usar as estimativas das fases para a estrutura como um todo ao invés de apenas para os espalhadores anômalos. Isso funciona porque ΔF é grande exatamente quando o ângulo de fase de f para os espalhadores anômalos é 90° distante do resto da estrutura (veja figura 8-14). As reflexões para as quais $\phi+90^\circ$ é uma estimativa ruim são exatamente as reflexões para as quais ΔF é pequeno e portanto não contribui para o mapa de qualquer forma.

6.5. ESTIMATIVAS INICIAIS PARA O FASEAMENTO POR MAD

As equações para a determinação das fases usando o método da dispersão anômala a múltiplos comprimentos de onda (MAD) foram primeiramente deduzidas por KARLE em 1980 sendo reformuladas por HENDRICKSON (HENDRICKSON, 1985). O fator de estrutura total observado para uma reflexão hkl medida num determinado comprimento de onda λ é dado por:

$${}^\lambda F(hkl) = F_T^0(hkl) + \sum_k F_{A_k}^0(hkl) \left[\frac{f_k' + if_k''}{f_k^0} \right] \quad [8.26]$$

Na qual ${}^\lambda F$ é o fator de estrutura que contém as fases que se quer determinar, F_T^0 é o componente normal do espalhamento de todos os átomos, F_A^0 é a contribuição anômala do espalhamento de todos os átomos. Como dado na equação 8.25, f^0 , f' e if'' são os componentes normal, real e imaginário, respectivamente, do espalhamento anômalo para cada espalhador k . Cada um desses componentes do fator de estrutura total observado pode ser representado num diagrama vetorial como na figura 8-15.

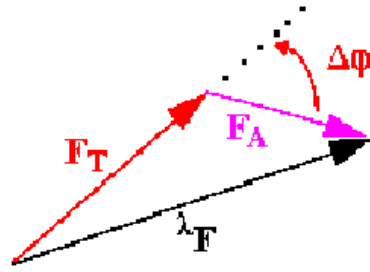


Figura 8-15: Diagrama vetorial mostrando as relações entre os componentes do fator de espalhamento total.

A idéia básica é que se os espalhadores anômalos puderem ser localizados na célula unitária, então F_A e o seu correspondente ângulo de fase φ_A podem ser calculados. As equações de Karle/Hendrickson podem ser usadas para gerar uma estimativa para $\Delta\varphi$ e calcular F_T . No caso mais simples, em que apenas um tipo de átomo espalhador anômalo está presente, pode-se estimar as fases de F_T como $\Delta\varphi + \varphi_A$. Assim, uma transformada de Fourier das amplitudes F_T e fases $\Delta\varphi + \varphi_A$ resulta num mapa de densidade eletrônica correspondente a todos os átomos na estrutura.

Então, considerando que apenas um tipo de átomo espalhador anômalo está presente e expressando a amplitude da intensidade medida $|\lambda F(hkl)|^2$ em termos de seus componentes:

$$\begin{aligned}
 |\lambda F(hkl)|^2 &= |F_T^0(hkl)|^2 \\
 &+ a(\lambda) |F_A^0(hkl)|^2 \\
 &+ b(\lambda) |F_T^0(hkl)| |F_A^0(hkl)| \cos[\varphi_T^0(hkl) - \varphi_A^0(hkl)] \\
 &+ c(\lambda) |F_T^0(hkl)| |F_A^0(hkl)| \sin[\varphi_T^0(hkl) - \varphi_A^0(hkl)] \quad [8.27]
 \end{aligned}$$

Sendo que:

$$a(\lambda) = [(f'^2)^2 + (f''^2)^2] / (f^0)^2 \quad [8.28]$$

$$b(\lambda) = 2(f' / f^0) \quad [8.29]$$

$$c(\lambda) = 2(f'' / f^0) \quad [8.30]$$

O mais importante é que toda a informação que depende do comprimento de onda no qual os dados são coletados é expressa pelos coeficientes $a(\lambda)$, $b(\lambda)$ e $c(\lambda)$. Usando esses coeficientes pode-se reescrever:

$$|^\lambda F(hkl)|^2 = F_T^2 + a(\lambda)F_A^2 + b(\lambda)F_T F_A \cos \Delta\varphi + c(\lambda)F_T F_A \sin \Delta\varphi \quad [8.31]$$

Cada conjunto de dados coletado num comprimento de onda, resulta num exemplo de equação 8.29. Assim tem-se um sistema de equações simultâneas das quais se quer obter as quantidades F_A , F_T e $\Delta\varphi$. Na verdade têm-se duas medidas para cada comprimento de onda porque a observação para $F^-(hkl)$ pode ser tratada como uma observação de $F^+(hkl)$ com o valor de f'' invertido. Então, para se obter os valores das três incógnitas, necessita-se de pelo menos três observações que são fornecidas coletando-se conjuntos de dados em dois comprimentos de ondas distintos. Para que se tenha maior certeza para os valores calculados, é melhor coletar dados em três ou mais comprimentos de onda.

Na prática, ou pelo menos no código implementado em programas como MADLSQ (HENDRICKSON, 1988), a equação 8.27 é escrita na forma linear com 4 incógnitas.

$$|^\lambda F(hkl)|^2 = P_1 + a(\lambda)P_2 + b(\lambda)P_3 + c(\lambda)P_4 \quad [8.32]$$

E para impor a identidade fundamental ($\sin^2 + \cos^2 = 1$) adiciona-se a imposição Lagrangiana:

$$0 = P_1 * P_2 - P_3 * P_3 - P_4 * P_4 \quad [8.33]$$

E as quantidades que realmente se quer determinar:

$$|F_T| = \sqrt{P_1} \quad [8.34]$$

$$|F_A| = \sqrt{P_2} \quad [8.35]$$

$$\Delta\varphi = \arctan g(P_4 / P_3) \quad [8.36]$$

O tratamento dos dados coletados para cristais com mais de um tipo de átomo espalhador anômalo é feito exatamente de forma paralela, com a adição de duas novas quantidades para serem estimadas para cada novo espalhador anômalo tipo k : $|F_{Ak}|$ e $\Delta\phi_k$. Assim, para dois tipos de espalhadores anômalos, existirão 5 incógnitas e serão necessários dados coletados em pelo menos três comprimentos de onda (F^+ e F^- em cada um dos 3 comprimentos de onda = 6 observações para calcular 5 incógnitas).

6.6. LOCALIZANDO OS ESPALHADORES ANÔMALOS

Até aqui, a partir dos conjuntos de dados de MAD coletados foram produzidas as estimativas iniciais para as três quantidades F_T , F_A e $\Delta\phi$. Agora, a partir dessas quantidades é que as fases para a proteína ϕ_T serão estimadas.

Como $\phi_T = \phi_A + \Delta\phi$, se o valor de ϕ_A for conhecido então as fases da proteína podem ser calculadas. Por isso é necessário calcular as posições dos átomos espalhadores anômalos na célula unitária. O jeito mais simples de fazer isso é através da construção de mapas de Patterson. Vários mapas de Patterson podem ser calculados usando os dados coletados mas, o mais útil talvez seja o mapa de Patterson calculado usando as diferenças de Bijvoet ΔF^2 como já mencionado. A figura 8-16 exemplifica um mapa desse tipo.

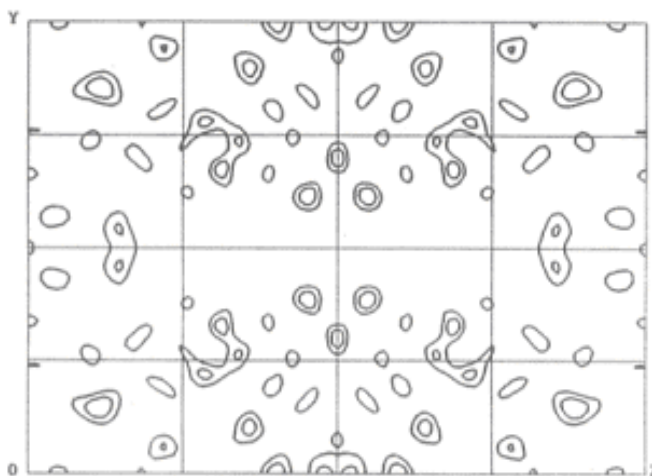


Figura 8-16: Aqui é mostrado um exemplo de localização de átomos de cobre na metaloproteína CBP que possui 96 resíduos (GUSS, et al., 1989). É mostrada a seção de Harker $u=1/2$ do mapa de Patterson calculado com coeficientes ΔF^2 usando dados coletados para $\lambda_{\text{pico}}=1,3771\text{\AA}$ no qual $f''=4,17e$. Note que a razão sinal/ruído é bastante elevada devido ao baixo valor de f'' .

Para aumentar a razão sinal/ruído do mapa, trabalha-se com os valores de ΔF calculados com os dados coletados no comprimento de onda em que as diferenças entre F^+ e F^- sejam as maiores possíveis, isto é, aquele com maior f'' . Como exemplificado na figura 8-17.

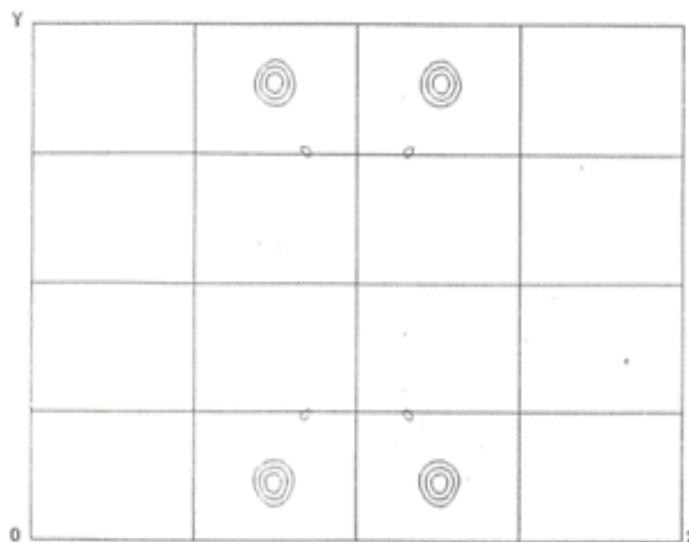


Figura 8-17: Aqui é mostrada a mesma seção de Harker $u=1/2$ mas, ao invés de usar um subconjunto de dados coletados, foram usados os valores estimados para $|F_A|$ em todos os quatro comprimentos de onda para calcular o mapa de Patterson com coeficientes $|F_A|^2$. Como todos os dados coletados contribuem simultaneamente para esse mapa, ele terá bem menos ruído sendo bem mais claros os sítios de cobre.

Como já foi mencionado quando se tratou da substituição isomórfica, existem alguns programas (SHAKE-N-BAKE e SHELXD) que podem auxiliar bastante na localização dos átomos pesados. Particularmente se muitos átomos pesados estão presentes na estrutura da proteína, o que pode tornar a interpretação dos resultados dos mapas de Patterson uma tarefa bem difícil (cerca de 20 átomos pesados é o limite para um mapa interpretável manualmente).

7. A SUBSTITUIÇÃO MOLECULAR

A substituição molecular baseia-se no fato de que o nível de semelhança entre as estruturas de duas proteínas se correlaciona muito bem com o nível de identidade entre as suas seqüências. Isso significa que se pode ter uma boa idéia se a substituição molecular irá ou não ser bem sucedida antes de tentar. A substituição molecular pode ser usada quando se tem um bom modelo para uma fração razoavelmente grande da estrutura no cristal. Como

uma regra, ela será provavelmente bastante satisfatória se o modelo da estrutura conhecida estiver suficientemente completo e sua seqüência compartilhar de pelo menos 30% de identidade com a da proteína desconhecida. Isso se torna progressivamente mais difícil conforme o modelo fica menos completo ou a identidade seqüencial compartilhada diminui. Mas, como o número de estruturas resolvidas nos bancos de dados fica cada vez maior, este método tem-se tornado cada vez mais útil na resolução de novas estruturas.

Na substituição molecular, a função de Patterson desempenha um papel fundamental. Calcula-se um mapa de Patterson para o modelo previamente conhecido (modelo de busca) e compara-se com o mapa de Patterson calculado a partir dos dados experimentais. Quando o modelo está corretamente orientado e posicionado na célula unitária, os dois Patterson devem ser similares. Nessa aproximação simplista, a substituição molecular é um problema de seis dimensões – três parâmetros para especificar a orientação e três para especificar a posição – o que já a torna um grande problema. Felizmente, o mapa de Patterson pode ser dividido em partes que são sensíveis apenas a alguns desses parâmetros – a estratégia é considerar os parâmetros separadamente, determinando primeiro a orientação e então a translação.

Como já foi discutido no início desse capítulo, o mapa de Patterson é um mapa vetorial, com picos nas posições dos vetores entre os átomos na célula unitária. Embora não seja possível resolver esses vetores para uma estrutura do tamanho de uma proteína, a forma com que eles são acumulados fornece uma assinatura para a estrutura da proteína. A figura 8-18 ilustra um mapa de Patterson correspondente a uma célula unitária contendo duas moléculas de quatro átomos cada.

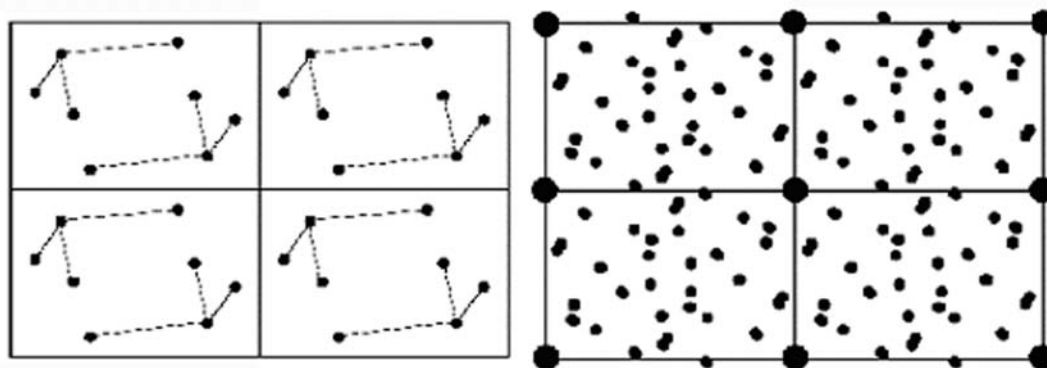


Figura 8-18: À esquerda são mostradas quatro células unitárias de um cristal contendo duas moléculas com quatro átomos. E à direita é mostrado o mapa de Patterson resultante de todos os vetores entre esses átomos.

Os vetores no mapa de Patterson podem ser divididos em duas categorias. Vetores intramoleculares, que são aqueles entre um átomo na molécula e outro átomo na mesma molécula, e dependem apenas da orientação da molécula e não da sua posição na célula unitária, sendo explorados na função de rotação. E vetores intermoleculares, que são aqueles entre átomos de moléculas diferentes, e dependem tanto da orientação da molécula quanto da sua posição na célula mas, uma vez que a orientação é conhecida, eles podem ser explorados na função de translação. Na figura 8-19 os vetores intramoleculares estão em azul e os intermoleculares em vermelho.

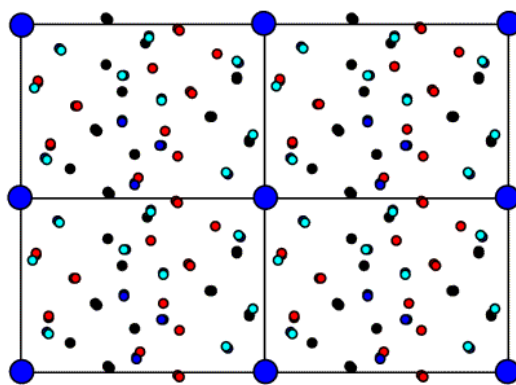


Figura 8-19: Mesmo mapa de Patterson da figura 8-18 mas com os vetores intramoleculares (azul) e os intermoleculares (vermelho) mostrados separadamente.

7.1. A FUNÇÃO DE ROTAÇÃO

Antes de definir matematicamente a função de rotação, é preciso considerar como descrever uma rotação. Para se descrever uma rotação são necessários três parâmetros. Uma maneira de definir uma orientação é definir um eixo de rotação e um ângulo de rotação sobre o eixo. Dois parâmetros definem um eixo, o qual pode ser representado como um vetor do centro de uma esfera até um ponto na superfície dessa esfera. Uma convenção comum é mostrada a seguir. O eixo de rotação começa paralelo ao eixo z e é rotado ao redor do eixo y por um ângulo ψ . Então ele é rotado ao redor do eixo z por um ângulo ϕ . Esses especificam um ponto na superfície de uma esfera unitária, análogo a definir um ponto na superfície da terra por latitude e longitude. Finalmente, o objeto que está sendo orientado é rotado por um ângulo κ ao redor do eixo de rotação. A descrição com $\kappa\psi\phi$ é particularmente útil se formos analisar rotações com um ângulo de rotação fixo.

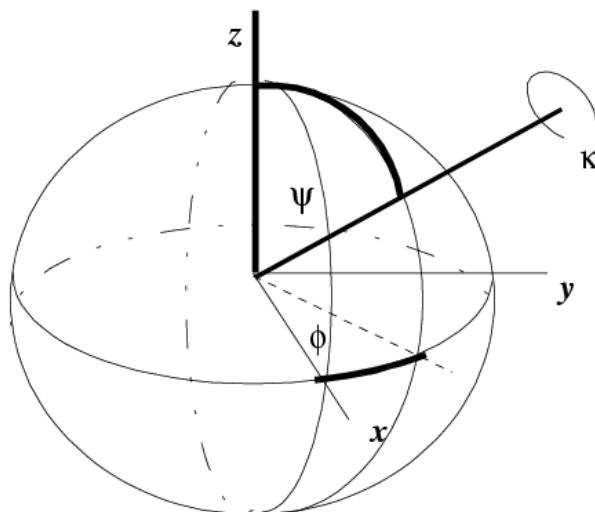


Figura 8-20: Descrevendo a rotação com $\kappa\psi\phi$.

Na equação a seguir, a função de rotação é expressa em termos de (κ, ψ, ϕ) .

$$R(\kappa, \phi, \psi) = \int_{r_{\min}}^{r_{\max}} P_{\text{experimental}}(u) P_{\text{modelo}}(\kappa, \phi, \psi, u) du \quad [8.37]$$

na qual $P_{\text{experimental}}$ é a função de Patterson para a molécula em estudo e P_{modelo} é a função de Patterson para o modelo de busca. Na integração sobre uma camada esférica, a região próxima à origem é tipicamente omitida para excluir o pico de Patterson da origem, o qual adiciona um grande termo constante. O raio da esfera é limitado porque os vetores intramoleculares estão mais concentrados próximos à origem.

Outra convenção, mais comum na substituição molecular, é usar os ângulos de Euler. Esses podem ser definidos de várias formas, mas a que é usada em programas que fazem substituição molecular, como o AMoRe (NAVAZA, 1994), é a convenção zyz . Nessa convenção, o sistema de coordenadas é rotado por um ângulo α ao redor do eixo original z , então por um ângulo β ao redor do novo eixo y e então por um ângulo γ ao redor do eixo final z . A descrição com ângulos de Euler é particularmente útil pois agiliza a matemática da função de rotação.

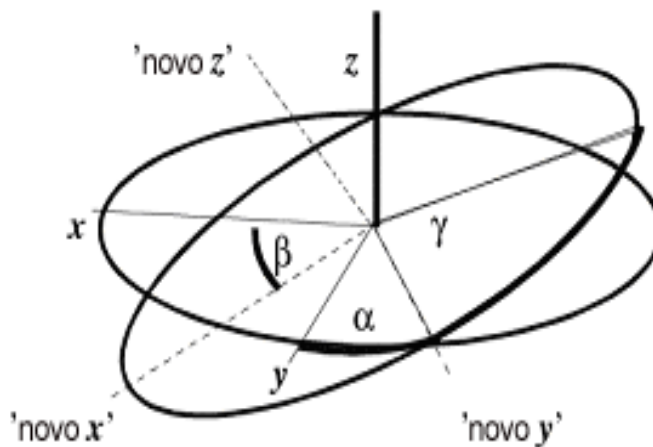


Figura 8-21: Descrição da rotação com ângulos de Euler utilizando a convenção zyz.

A partir disso, a função de rotação é comumente definida como uma função produto. O Patterson do modelo de busca é rotado e superposto no Patterson experimental. O produto dos dois Pattersons para todos os pontos numa camada esférica é somado e o resultado é a função de rotação.

Uma maneira mais rápida porém mais complexa de se determinar a função de rotação é através da chamada função de rotação rápida. Essa foi desenvolvida por Tony Crowther (NAVAZA, 1994) que percebeu que a função de rotação poderia ser rapidamente computada com transformadas de Fourier se as funções de Patterson fossem expressas em coordenadas polares (r, θ, φ) como descrito na equação 8.38.

$$R(r, \theta, \varphi) = \int_V P_{nat}(r, \theta, \varphi) RP_{mol}(r, \theta, \varphi) dr^2 \sin \theta dr d\theta d\varphi \quad [8.38]$$

A função de rotação no AMoRe é baseada na mesma teoria, mas Jorge Navaza percebeu que a implementação original da função de rotação rápida apresentava algumas instabilidades numéricas durante o processo computacional. A função acima pode ser expandida usando-se a função de Bessel e o resultado final é que ela pode ser estimada como a soma de dois termos, um deles independente da rotação, o que diminui o tempo computacional necessário. Além disso, as melhoras nas análises numéricas aumentaram significativamente a razão sinal/ruído.

7.2. A FUNÇÃO DE TRANSLAÇÃO

Desde que a orientação da molécula foi encontrada numa célula desconhecida, o próximo passo é a determinação da sua posição absoluta. Quando uma molécula de referência, já orientada, é transladada na célula desconhecida, moléculas simetricamente relacionadas movem-se da mesma maneira e todos os vetores intermoleculares mudam.

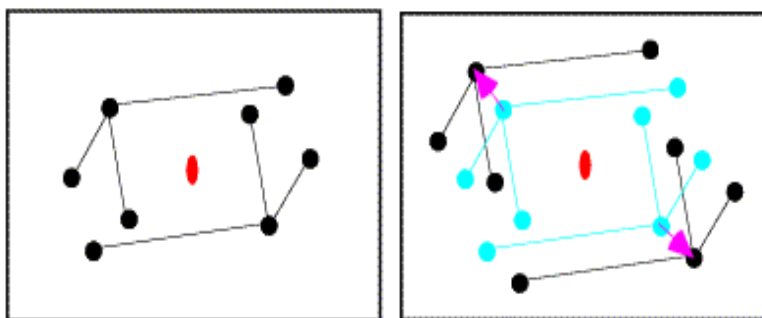


Figura 8-22: À esquerda são mostradas duas moléculas posicionadas arbitrariamente na célula unitária. A molécula inferior é gerada pelo eixo cristalográfico de ordem 2 em vermelho. Quando esta molécula é transladada para a posição correta, o que é indicado pelas setas rosas na figura da direita, a cópia se move na direção oposta (espelho) em relação ao eixo de ordem 2.

Somente quando todas as moléculas na cela unitária estão nas posições corretas, os vetores intermoleculares do Patterson do modelo de busca sobrepõe-se aos do Patterson experimental. A figura 8-22 ilustra uma translação aplicada à uma das células unitárias da figura 8-18 com duas moléculas com quatro átomos.

Como a função de rotação, a definição tradicional para a função de translação foi uma função produto. Primeiro é preciso definir matematicamente a função de Patterson calculada correspondente aos vetores intermoleculares. O mapa de Patterson é uma função de autocorrelação, pois é uma função que correlaciona a densidade com ela mesma. Assim o conjunto de vetores intermoleculares é uma função de correlação entre a densidade eletrônica para uma molécula e a densidade para a outra molécula, o que é mostrado pela figura 8-23 e pela equação 8.39.

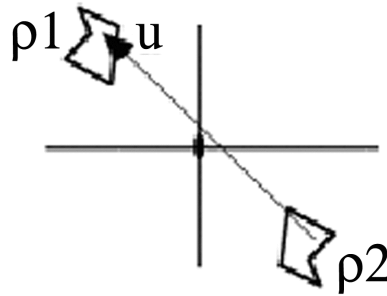


Figura 8-23: Translação do pico de densidade eletrônica de ρ_1 (molécula 1) para ρ_2 (molécula 2) em relação ao eixo de ordem 2.

$$P_{2 \rightarrow 1}(u) = \int_{\text{célula}} \rho_1(x) \rho_2(x-u) dx = \frac{1}{V} \sum_h F_1(h) F_2^*(h) \exp(-2\pi i h \cdot u) \quad [8.39]$$

A translação requerida pode ser deduzida transladando os vetores intermoleculares sobre o mapa de Patterson observado e calculando a função produto. Quando a translação correta é escolhida, as duas funções se sobrepõem e é observado um máximo para a função de translação (equações 8.40 a seguir).

$$T(t) = \int_{\text{célula}} P_{2 \rightarrow 1}(u-t) P_{\text{exp}}(u) du \quad [8.40]$$

$$T(t) = \frac{1}{V} \sum_h \left(F_1(h) F_2^*(h) \right)^* \left| F_0(h) \right|^2 \exp(-2\pi i h \cdot t) \quad [8.41]$$

$$T(t) = \frac{1}{V} \sum_h F_1^*(h) F_2(h) \left| F_0(h) \right|^2 \exp(-2\pi i h \cdot t) \quad [8.42]$$

Note que essa função de translação fornece apenas o componente da translação perpendicular ao eixo de rotação. Para grupos espaciais com maiores simetrias, como o P222, podem ser calculadas funções de translação para cada eixo de simetria, assim são calculados todos os componentes do vetor de translação. Se forem consideradas todas as operações de simetria simultaneamente, a função fornece a translação diretamente e a razão sinal/ruído pode ser melhorada. Uma das maneiras de se fazer isso é calcular o **coeficiente de correlação** entre o quadrado das amplitudes observadas e calculadas como uma função

da translação do modelo. A translação correta deve resultar num pico nesse mapa. A função de correlação é expressa pela equação 8.43.

$$C(t) = \frac{\sum (|F_0|^2 - \overline{|F_0|^2}) (|F_C(t)|^2 - \overline{|F_C(t)|^2})}{\sqrt{\sum (|F_0|^2 - \overline{|F_0|^2})^2 \sum (|F_C(t)|^2 - \overline{|F_C(t)|^2})^2}} \quad [8.43]$$

Por sua vez, devido ao teorema de Parseval, esse coeficiente de correlação é equivalente ao coeficiente de correlação entre dois mapas de Patterson cuja origem tenha sido removida. Por isso, é freqüentemente chamado de correlação de Patterson (PC). O programa AMoRe dá os valores de PC como uma maneira primária para avaliar a qualidade das soluções obtidas para a substituição molecular.

8. MÉTODOS DIRETOS

Não se entrará em detalhes aqui, mas se assumirmos que um cristal é composto por átomos com formas semelhantes e todos têm densidade eletrônica positiva, então existem relações estatísticas entre conjuntos de fatores de estrutura. Essas relações estatísticas podem ser usadas para deduzir possíveis valores para as fases. Os métodos diretos exploram essas relações e podem ser usados para resolver estruturas de pequenas moléculas. Infelizmente, as relações estatísticas ficam mais fracas conforme o número de átomos aumenta, e os métodos diretos estão limitados a estruturas com no máximo poucas centenas de átomos na célula unitária. Embora existam avanços que estendem esse limite, particularmente para cristais que difratam até resoluções altas (1.2Å ou melhor), os métodos diretos geralmente não são aplicáveis para a maioria das estruturas cristalinas de proteínas.

CAPÍTULO 9

REFINAMENTO E VALIDAÇÃO DO MODELO

1. INTRODUÇÃO

Os métodos de determinação estrutural fornecem um modelo freqüentemente incompleto com apenas uma parte dos átomos da célula unitária. Como visto, a partir da substituição molecular é obtido um modelo inicial baseado no modelo de busca usado, por isso deve-se trocar a seqüência de aminoácidos de acordo com os dados experimentais do modelo. Então é possível verificar as partes faltantes do modelo, geralmente regiões mais lábeis e variáveis como loops e terminais apresentam as maiores distorções, embora em alguns casos partes maiores da proteína podem estar faltando.

Esse conjunto parcial de átomos contém informação suficiente para que os átomos restantes sejam localizados. A partir do tipo e das posições relativas dos átomos no modelo inicial um conjunto de fatores de estrutura pode ser calculado. Então um mapa de densidade eletrônica pode ser preparado usando os ângulos de fase calculados e os fatores de estrutura observados (F_o) obtidos a partir das intensidades experimentais. Porém, os mapas de densidade com coeficientes de $|F_o|$ fornecem apenas as partes do modelo que já são conhecidas. Por isso é mais útil preparar um mapa de densidade eletrônica da diferença, usando coeficientes de $(|F_o|-|F_c|)$ e os ângulos de fase calculados. Nos mapas da diferença

podem ser observados picos onde uma quantidade insuficiente de densidade eletrônica tenha sido incluída no modelo (átomos ausentes) e são observados buracos com densidade negativa onde muita densidade eletrônica tenha sido incluída no modelo (um átomo mais pesado ou um resíduo maior do que o real).

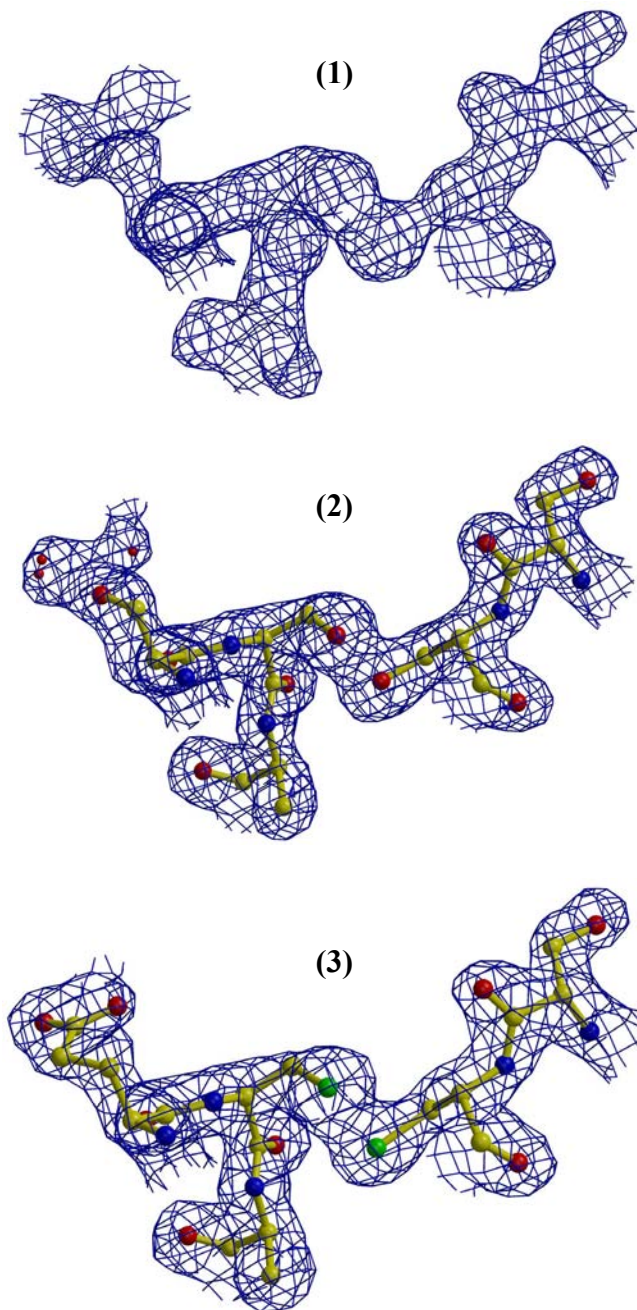


Figura 9-1: Em (1) é mostrada uma porção do mapa de densidade eletrônica obtido para CTI usando substituição isomórfica; em (2) é mostrada essa mesma porção do mapa superposta com a porção correspondente do modelo construído automaticamente (arp/warp) para CTI; e em (3) a mesma porção após interferência manual trocando os resíduos pelos resíduos corretos. A porção mostrada corresponde à única ponte dissulfeto presente em CTI.

Pela interpretação desses mapas, átomos faltantes podem ser localizados e incluídos no modelo e átomos deslocados podem ser corrigidos. Dessa maneira o modelo inicial parcial é melhorado. Então o modelo é submetido a alguns ciclos de refinamento, novos fatores de estrutura são calculados e novos mapas de densidade eletrônica da diferença são preparados. Esse processo é repetido até que o maior número possível de átomos tenha sido determinado.

Já no caso da substituição isomórfica ou da dispersão anômala, a partir das posições dos átomos pesados, são obtidos os mapas de densidade eletrônica parciais para o modelo. A partir da interpretação desses mapas o modelo vai sendo construído manualmente ou com a ajuda de métodos automáticos.

Esses mapas de densidade eletrônica não são auto-explicativos, necessitando de inteligência humana ou da ajuda de métodos automáticos para a sua interpretação. Os principais problemas com os mapas podem ser divididos em três categorias: os mapas podem ser indistintos, o que é devido principalmente à resolução, podem estar incorretos, o que é causado por erros no modelo ou podem ser observadas regiões de densidade não esperada, o que pode indicar a presença de moléculas de água, cofatores, açúcares ou outros tipos de ligantes. A interpretação dos mapas de densidade eletrônica é feita com base nas informações químicas, estereoquímicas, seqüência de aminoácidos e natureza de qualquer ligante presente (se conhecido) sobre a macromolécula em estudo.

2. A RESOLUÇÃO

Quanto melhor a resolução, mais fácil será localizar e ajustar os átomos individuais, e o problema passa a ser juntar os pontos.

No outro extremo, apenas pedaços grandes da macromolécula podem ser localizados e construídos. Por exemplo, a 6Å α -hélices podem ser localizadas mas folhas- β s não são claras. Em resoluções maiores do que 8Å apenas moléculas inteiras podem ser distinguidas.

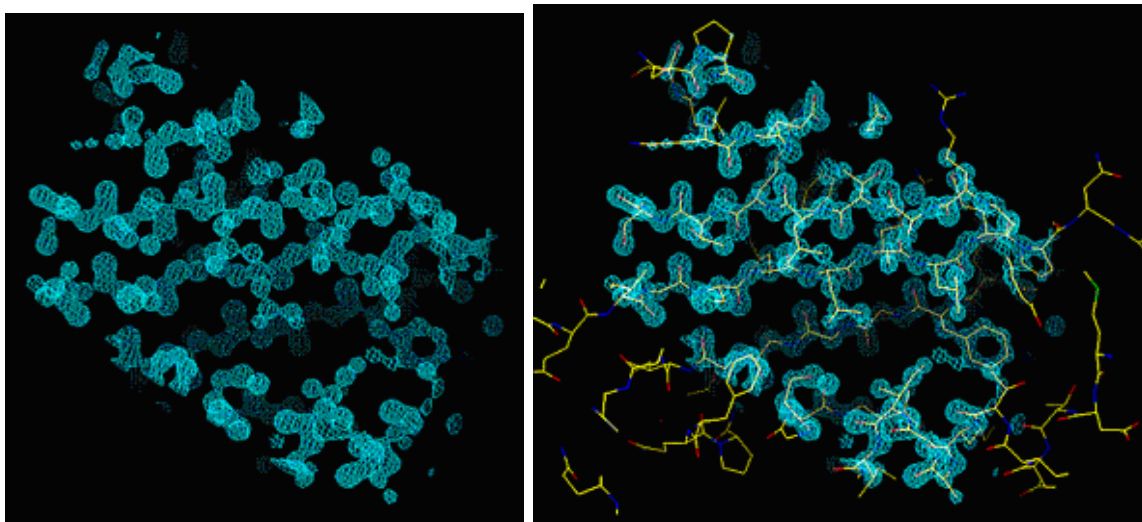


Figura 9-2: À esquerda o mapa de densidade a 1Å de resolução e à direita o modelo construído e ajustado à esse mapa. A densidade é contínua com picos para os átomos.

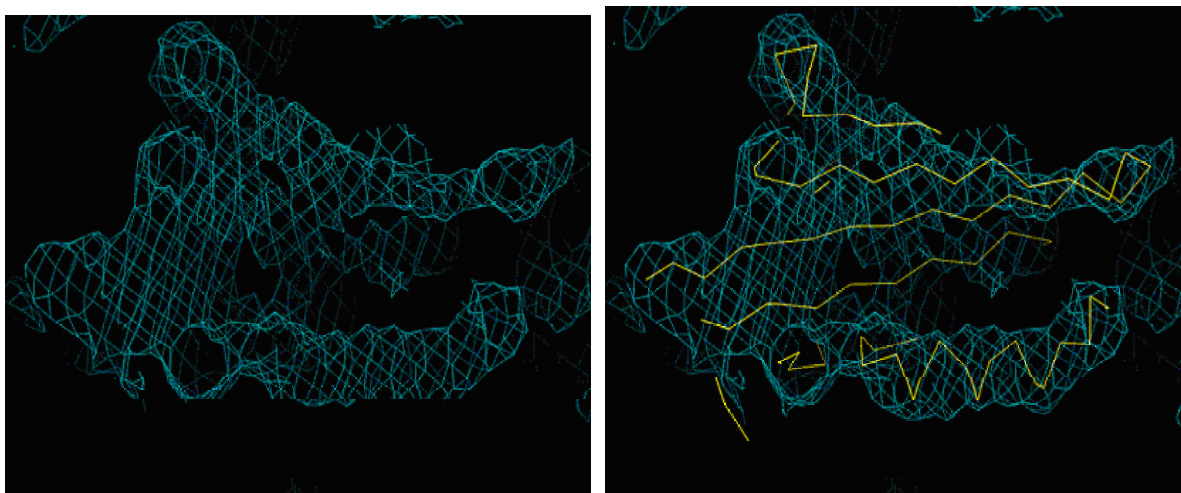


Figura 9-3: Aqui a resolução é de 6Å e apenas a cadeia principal pode ser vista de forma clara no mapa de densidade, então apenas os carbonos- α foram ajustados à densidade.

Numa resolução de 1Å os átomos individuais podem ser perfeitamente visualizados e o átomo de nitrogênio é maior que os de carbonos. A 2,5Å a cadeia principal e a maioria das cadeias laterais podem ser facilmente localizadas e ajustadas, os anéis são bastante claros. A 3Å essa tarefa se torna mais difícil e a 4Å a localização e o ajuste dos átomos é bastante impreciso.

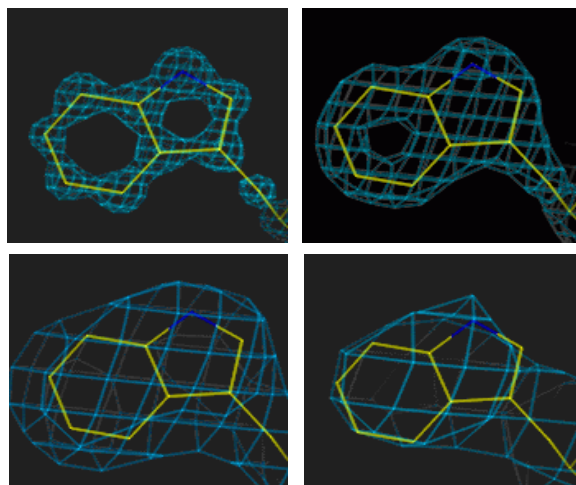


Figura 9-4: Comparação da qualidade da densidade eletrônica em várias resoluções.

3. ESTRATÉGIAS PARA A CONSTRUÇÃO E OS AJUSTES DO MODELO

A construção do modelo é bastante complicada para ser feita num único passo, principalmente quando se usa a construção manual, por isso ela pode ser dividida em estágios. Lembre-se que o que está sendo construído é um modelo, uma hipótese, e qualquer coisa pode ser mudada desde que os comprimentos e ângulos das ligações bem como a planaridade e estereoquímica dos grupos sejam respeitadas, como veremos. Em linhas gerais podem-se seguir os seguintes passos:

- (1) Traçar a cadeia polipeptídica tão completa quanto possível seguindo o mapa de densidade eletrônica inicial. Podem ser deixados buracos em regiões mais difíceis. Isso servirá como um caminho a se seguido e completado nos estágios seguintes.
- (2) Identificar pelo menos um ponto da sequência usando a densidade ou marcadores conhecidos. Podem ser usados, por exemplo, resíduos de cisteína que formam pontes dissulfeto ou o sinal anômalo produzido por elas ou por metioninas.
- (3) Localizar e nomear resíduos com a sequência correta identificando-os se possível. Nos lugares onde existem buracos mas alguma densidade pode ser vista podem ser construídas poli-Ala ou poli-Gly.
- (4) Ir preenchendo a cadeia adicionando átomos de acordo com a estereoquímica conhecida e bibliotecas de conformações comuns ao esqueleto carbônico, e adicionando-se as cadeias laterais nas conformações mais comuns para os rotâmeros.

(5) Alterar cada resíduo para melhorar o ajuste à densidade mantendo ou re-impondo a estereoquímica.

Existe um crescente número de programas que automatizam parte ou todo esse processo. Por exemplo, para construir fragmentos como hélices, folhas, etc, podem ser usados os programas ESSENS (KLEYWEGT & JONES, 1994) ou FFFEAR (COWTAN, 1998), para construir cadeias localizando átomos podem ser usados ARPWARP (warpNtrace) (PERRAKIS *et al.*, 1997) ou QUANTA (OLDFIELD *et al.*, 1992).

Nos casos favoráveis em que a resolução é alta o suficiente (warpNtrace necessita de uma resolução bastante boa) esses programas podem ser muito úteis sendo capazes de construir a maioria da estrutura.

Intercaladamente a esses estágios de construção e ajustes, o modelo é submetido a ciclos de refinamento que podem ser fundamentais para indicar pontos onde átomos estejam faltando ou onde eles tenham sido colocados erradamente, além de corrigir a estereoquímica onde necessário. A seguir são apresentados os métodos de refinamento mais utilizados.

4. MÉTODO DA MÁXIMA VEROSSIMILHANÇA (MAXIMUM LIKELIHOOD)

A idéia básica do método da máxima verossimilhança é bem simples: o melhor modelo é o mais consistente com as observações. Essa consistência é medida estatisticamente pela probabilidade com que as observações devem ter sido feitas. Se o modelo é mudado para fazer as observações mais prováveis, a verossimilhança aumenta, indicando que o modelo é melhor. Também se pode dizer que o modelo concorda melhor com os dados experimentais, mas a idéia de probabilidade define “concordância” mais precisamente.

Matematicamente, a verossimilhança (ou o likelihood) é definida como a probabilidade de fazer um conjunto de medidas. Se forem feitas N observações de diferentes quantidades x_i por exemplo, então a verossimilhança (L para likelihood) é definida por:

$$L = p(x_1, x_2, x_3, \dots, x_N) \quad [9.1]$$

Note que, como indicado pelas vírgulas, trata-se de uma combinação da distribuição de probabilidades de todas as medidas. Frequentemente as medidas serão independentes umas das outras e nesse caso a combinação de probabilidades é simplesmente o produto de todas as probabilidades individuais:

$$L = \prod_{j=1}^N p(x_j) \quad [9.2]$$

Assumir que as probabilidades são independentes certamente simplifica, pois combinar distribuições pode ser extremamente difícil, mas informações podem ser perdidas se os erros são correlacionados, então é importante se tomar cuidado.

Geralmente a função de verossimilhança é descrita como uma soma logarítmica, não só porque a soma é mais fácil de manipular do que o produto mas também porque o produto de uma grande quantidade de números muito pequenos fica ainda menor para se trabalhar computacionalmente. O log varia conforme o argumento, isto é, $\ln(x)$ aumenta quando x aumenta e diminui quando x diminui, assim o log de uma função terá o máximo na mesma posição.

$$LL = \ln L = \sum_{j=1}^N \ln[p(x_j)] \quad [9.3]$$

Conforme o próprio nome do método expressa, estamos interessados em encontrar a posição do máximo da função de verossimilhança. A fim de tornar a analogia com o método dos mínimos quadrados mais óbvia, frequentemente é feita a minimização da função menos log da verossimilhança, o que é equivalente.

Uma forma de pensar sobre a verossimilhança é imaginar que nenhuma medida dos dados tenha sido feita ainda. Tem-se um modelo com vários parâmetros para ajustar (que são as coordenadas e B-fatores para cada átomo do modelo) e alguma idéia das fontes de erro e como eles podem se propagar. Com isso, calcula-se a probabilidade de qualquer possível conjunto de medidas. Então se compara com as medidas experimentais e verifica-se a concordância entre elas.

Embora a idéia básica por trás do método seja simples, a parte difícil é trabalhar com a probabilidade com que uma observação é feita. As probabilidades têm que levar em conta os efeitos de todas as fontes de erros, incluindo não apenas os erros nas medidas mas também os erros no modelo em si. Mas conforme o modelo fica melhor, esses erros ficam claramente menores, o que faz com que a curva de probabilidades fique mais “pontuda”, significando que as probabilidades estão mais concentradas ao redor do valor correto que é o pico da curva. Esse efeito de tornar a curva mais pontuda aumenta a verossimilhança, desde que o valor do pico esteja realmente correto. O exemplo a seguir irá esclarecer melhor esse efeito.

5. UM EXEMPLO DE VEROSSIMILHANÇA

A fim de se entender mais facilmente o comportamento de uma função de verossimilhança pode-se usar esse exemplo. Toma-se uma série de vinte números que seguem a distribuição de probabilidades Gaussiana com uma média de 5 e um desvio padrão de 1. Usando esses dados, pode-se deduzir a estimativa da média e do desvio padrão com a função de máxima verossimilhança. A distribuição de probabilidades é muito importante para definir a função de verossimilhança e uma distribuição Gaussiana é freqüentemente uma boa escolha, afinal a grande maioria dos sistemas trabalhados na pesquisa científica segue essa distribuição.

A função de verossimilhança será então a probabilidade de gerar os vinte pontos dos dados. Os parâmetros dessa função de verossimilhança serão a média e o desvio padrão da distribuição de probabilidades Gaussiana.

No gráfico a seguir, é representada a distribuição de probabilidades Gaussiana para uma média de 5 e um desvio padrão de 1. As vinte barras verticais correspondem aos vinte pontos dos dados, a altura de cada barra representa a probabilidade daquela medida considerando a média e o desvio padrão assumidos. A função de verossimilhança é o produto de todas as alturas. Nesse caso nenhuma das probabilidades é particularmente baixa e elas são maiores no centro da distribuição, que é mais populada pelos dados.

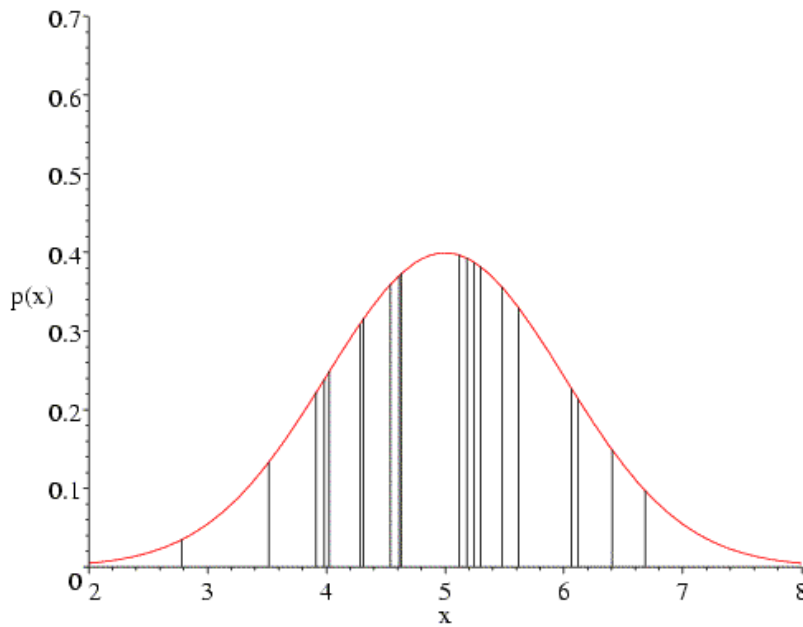


Figura 9-5: Uma distribuição gaussiana de probabilidades para uma média de 5 e um desvio padrão de 1.

O que acontece se a média assumida para a distribuição for mudada? Na figura 9-6A a seguir se pode ver que, se a média for mudada para 4, os dados nas extremidades da distribuição ficam muito improváveis o que reduz significativamente a função de verossimilhança. Além disso, diminui a quantidade de dados na região de pico. O mesmo acontece se a média for mudada para 6 (figura 9-6B).

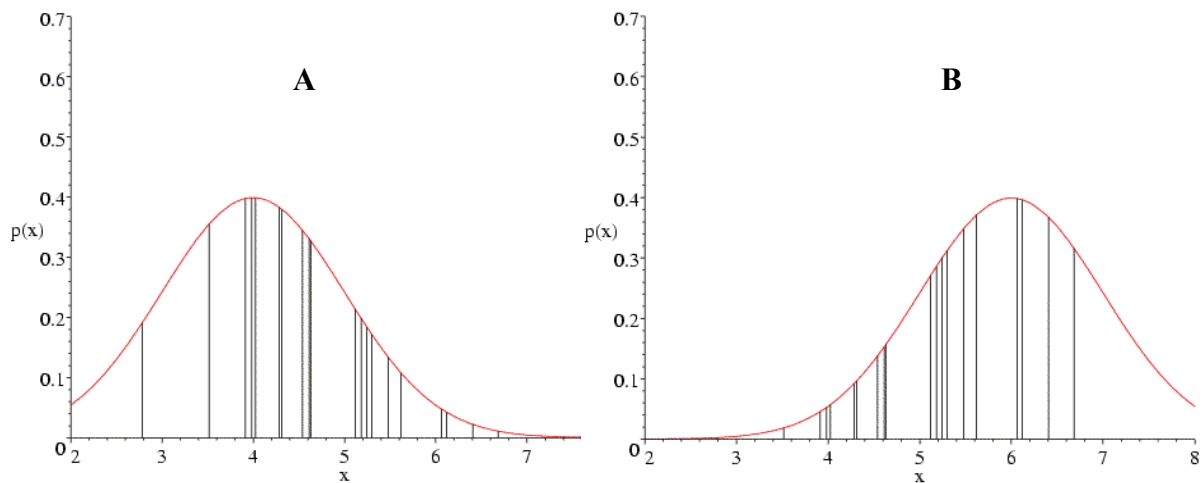


Figura 9-6: A mesma distribuição da figura 9-1 mas em (A) a média foi mudada para 4 e em (B) para 6.

Agora o que acontece se a média estiver correta mas o desvio padrão estiver errado? Na figura 9-7A o valor de 0,6 é usado para o desvio padrão. Na região próxima ao pico, que é densamente populada, os valores de probabilidade aumentam, mas os valores nas

extremidades diminuem numa proporção maior e dessa forma o valor global da verossimilhança é reduzido. Similarmente se o valor assumido para o desvio padrão for muito alto, como o valor 2 na figura 9-7B, as probabilidades nas extremidades aumentam mas a diminuição na região densamente populada do pico faz com que haja uma diminuição na verossimilhança.

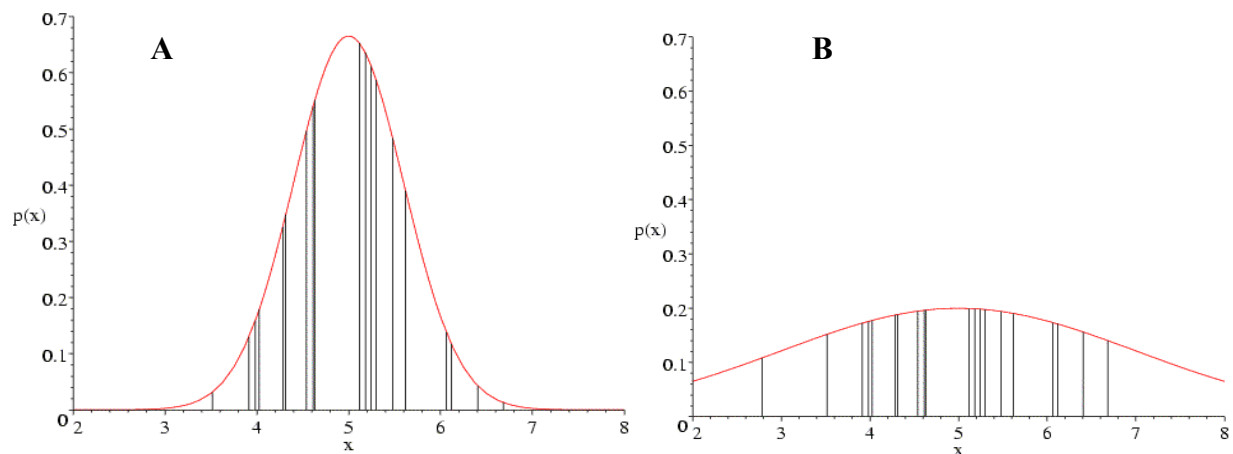


Figura 9-7: A mesma distribuição da figura 9-1 mas em (A) o desvio padrão foi mudado para 0,6 e em (B) para 2.

Se a média estiver correta, a função de verossimilhança será balanceada pela influência das extremidades esparsamente populadas e do centro densamente populado indicando o desvio padrão correto.

Pode-se fazer assim o cálculo para todos os possíveis pares de média e desvio padrão, como mostrado no gráfico da figura 9-8. O pico nessa distribuição é próximo à média e desvio padrão que foram usados para gerar os dados a partir da distribuição Gaussiana.

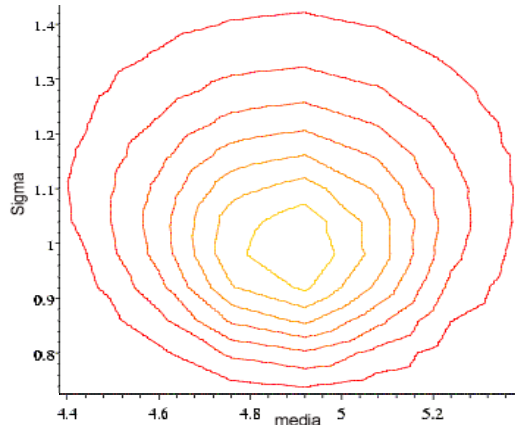


Figura 9-8: A função de verossimilhança vista em 2D (uma vista de cima da representação 3D). Todos os possíveis pares de média e desvio padrão calculados para a distribuição do exemplo formam os círculos quase concêntricos e o centro desses círculos representa o máximo da função de verossimilhança e se aproxima bastante da média (5) e do desvio padrão (1) verdadeiros.

6. O MÉTODO DOS MÍNIMOS QUADRADOS

O método dos mínimos quadrados é um caso particular do método da máxima verossimilhança. No método dos mínimos quadrados, as observações têm valores fixos e os parâmetros são variados até que os valores calculados se aproximem o máximo possível dos observados. Nessas condições, o método da máxima verossimilhança dá os mesmos resultados que o dos mínimos quadrados caso as observações tenham uma distribuição Gaussiana. Primeiro lembrando a equação para a distribuição de probabilidades de Gaussian: (DRENTH, 1994)

$$p(x_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left[-\frac{1}{2\sigma_j^2}(x_j - \hat{x}_j)^2\right] \quad [9.4]$$

Sendo que $\hat{x}_j$ é o valor predito de x_j . A forma logarítmica da equação da verossimilhança é dada pelo log dessa distribuição de probabilidades:

$$\ln[p(x_j)] = -\frac{1}{2}\ln(2\pi) - \ln\sigma_j - \frac{1}{2\sigma_j^2}(x_j - \hat{x}_j)^2 \quad [9.5]$$

Agora a verossimilhança é a probabilidade de fazer o conjunto inteiro de medidas. Como antes, assume-se que as observações são independentes, de tal forma que a verossimilhança é o produto de todas as probabilidades individuais e o log da verossimilhança é a soma dos logs. Como a comparação será feita com o método dos mínimos quadrados, usa-se menos o log da verossimilhança:

$$-LL = \sum_{j=1}^N \frac{1}{2} \ln(2\pi) + \ln \sigma_j + \frac{1}{2\sigma_j^2} (x_j - \hat{x}_j)^2 \quad [9.6]$$

Comparando-se com a soma ponderada dos quadrados residuais usada no método dos mínimos quadrados:

$$WSSQ = \sum_{j=1}^N \frac{1}{\sigma_j^2} (x_j - \hat{x}_j)^2 \quad [9.7]$$

O termo $\ln(2\pi)$ é constante em $-LL$. Se assumirmos que os desvios padrões (erros das medidas) são também constantes, então $-LL$ difere de $WSSQ$ apenas por um fator de 2 e obviamente resultará na mesma solução de otimização.

Mas o método dos mínimos quadrados apresenta alguns inconvenientes, como vai se ver, a probabilidade de distribuição para os erros não é gaussiana e os desvios padrões não são constantes.

Poderia-se argumentar que as distribuições de Gaussian são tão comuns que o método dos mínimos quadrados deveria ser a escolha correta. De fato, várias importantes distribuições do fator de estrutura seguem a forma Gaussiana. Mas os erros nessas distribuições do fator de estrutura estão na forma Gaussiana no plano complexo, enquanto são medidas as intensidades (com as quais são calculadas as amplitudes dos fatores de estrutura) são feitas no plano real. Então para usar a distribuição dos fatores de estrutura para prever as amplitudes, é preciso mediar todos os possíveis valores das fases e isso muda a natureza da distribuição dos erros. No método dos mínimos quadrados, os desvios quadráticos devem ser ponderados pelo inverso da variância (=quadrado do desvio padrão) da observação. Mas os refinamentos ponderados pelas medidas experimentais dos erros não convergem bem.

$$SS = \sum_h \frac{1}{\sigma^2} (|F_0(h)| - |F_C(h)|)^2 \quad [9.8]$$

O método dos mínimos quadrados parece ser independente das fases, mas de fato ele implicitamente assume que as fases estão essencialmente corretas.

$$SS = \sum_h \frac{1}{\sigma^2} \| |F_0(h)| \exp(i\alpha_C) - |F_C(h)| \|^2 \quad [9.9]$$

Ignorando o termo da variância, o método dos mínimos quadrados é equivalente a minimizar a diferença entre a densidade eletrônica do modelo (=transformada de Fourier de F_C) e um mapa preparado usando as amplitudes dos fatores de estrutura experimentais e as fases calculadas. Por isso, o refinamento pode convergir pobremente, afinal o mapa computado com as amplitudes experimentais e fases calculadas é tendencioso para se parecer com o modelo. Por isso, o uso do método dos mínimos quadrados deve ser feito com cautela para que não seja inapropriado, por exemplo quando existe a falta de isomorfismo ou quando a qualidade dos dados medidos é baixa, o que pode introduzir erros no modelo afetando as fases calculadas.

7. O MÉTODO DO MAXIMUM LIKELIHOOD PARA O REFINAMENTO

Como visto, a probabilidade de distribuição dos fatores de estrutura tem a forma gaussiana no plano complexo. Mas, como o que se medem são as intensidades ou as amplitudes dos fatores de estrutura, é necessário converter as probabilidades dos fatores de estrutura numa forma que aplique as amplitudes. Os fatores de estrutura podem ser considerados como números complexos com uma parte real e uma parte imaginária ou como tendo uma amplitude e uma fase. Para se ter uma distribuição expressa em termos de amplitudes, é preciso converter primeiro a descrição em termos de amplitude/fase e então integrar para eliminar a fase. Em geral, variáveis aleatórias secundárias numa distribuição de probabilidades combinadas podem ser eliminadas por integração. Essa integração é mostrada na equação 9.10 e representada esquematicamente na figura 9-9 a seguir.

$$p(|F|; F_C) = \int_0^{2\pi} p(|F|, \alpha; F_C) d\alpha \quad [9.10]$$

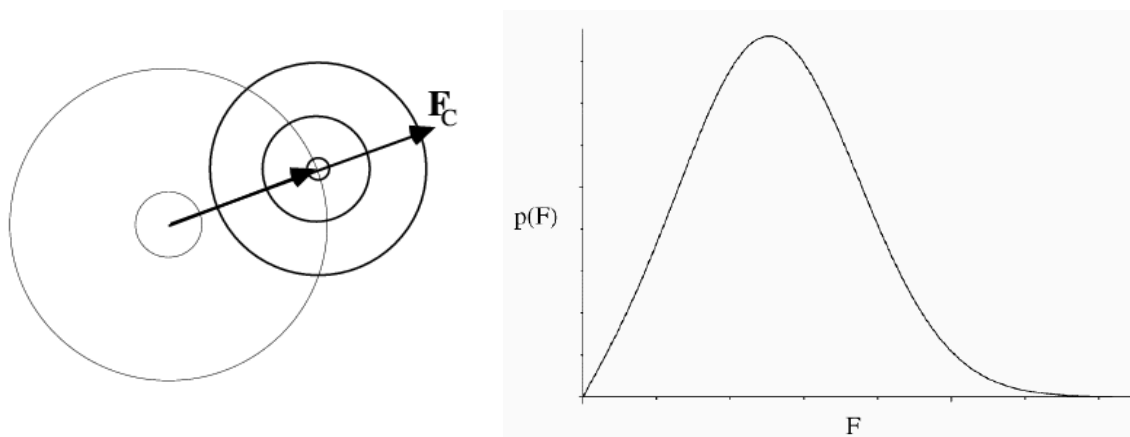


Figura 9-9: Representação esquemática do vetor fator de estrutura F_C e do respectivo círculo de fase e à direita a distribuição de probabilidades desses fatores de estrutura que representa a equação a seguir.

Na teoria estatística, o resultado dessa integração é chamado de distribuição de Rice, mas a mesma distribuição é usada em cristalografia por Luzzati e Sim e seus sucessores. A distribuição de Rice leva em conta os efeitos dos erros do modelo para a predição da amplitude observada, mas ainda não leva em conta os efeitos dos erros observados.

Três métodos são usados na prática. De uma forma bastante simplificada:

(1) Aumentar a variância da gaussiana 2D. Isso modela o efeito de um erro na amplitude como um erro no plano complexo, o que à primeira vista parece uma má idéia. Mas como o erro da medida é pequeno comparado à amplitude, apenas o componente paralelo ao vetor fator de estrutura desse erro adicionado irá influenciar a distribuição de amplitudes. O componente perpendicular ao vetor do fator de estrutura não terá realmente muito impacto. O benefício dessa aproximação é que ela ainda resulta numa distribuição de Rice, a qual é fácil de manusear nos programas. Essa é a aproximação usada no programa REFMAC e BUSTER/TNT.

(2) Converter a distribuição de Rice na sua aproximação gaussiana, então aumentar a variância para essa aproximação gaussiana. Essa aproximação é fácil de se fazer na prática

e geralmente dá bons resultados. Essa é a base das opções MLF e MLF1 nos programas CNS e X-PLOR (BRÜNGER *et al.*, 1998)

(3) Transformar a distribuição de Rice para trabalhar em termo dos quadrados das amplitudes, então adicionar um erro gaussiano para a intensidade. Numericamente, essa é a aproximação mais difícil de se manipular. Entretanto, ela mostra vantagens significantes. Primeiro, o que se medem são as intensidades, essencialmente quadrados das amplitudes, e os erros nas medidas são mais bem tratados como erros gaussianos nas intensidades. Segundo, quando os dados são processados, freqüentemente termina-se com uma quantidade substancial de intensidades líquidas negativas após subtrair o background do pico. Essas são observações perfeitamente legítimas, mas o refinamento baseado nas amplitudes (mínimos quadrados) enfrenta problemas porque não se pode tirar a raiz quadrada de um número negativo. Os programas que tratam o refinamento usando intensidades, manipulam esses casos naturalmente, é o caso das opções MLI e MLF2 nos programas CNS e X-PLOR. (BRÜNGER *et al.*, 1998)

Na prática os três métodos são similares, embora MLI parece dar resultados ligeiramente melhores do que MLF.

8. O USO DAS FASES NO REFINAMENTO POR MÁXIMA VEROSSIMILHANÇA

Quando se toma a integral ao redor do círculo de fase para tornar a distribuição de probabilidades dos fatores de estrutura numa gaussiana 2D numa distribuição de Rice, assume-se implicitamente que não se tem nenhuma outra informação das fases além do fator de estrutura calculado, isto é, que todas os possíveis valores de fases têm probabilidade igual. O efeito é que o método da verossimilhança depende apenas da amplitude de F_C e não da sua fase. A figura 9-10 a seguir ilustra isso.

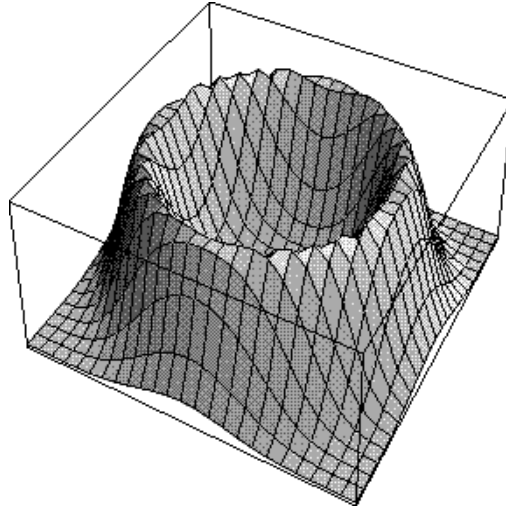


Figura 9-10: A função de verossimilhança contribui para uma dada reflexão como uma função de F_C : a parte real de F_C é representada ao longo do eixo horizontal e a parte imaginária ao longo do eixo vertical. A depressão no centro indica a perda na verossimilhança.

Por outro lado, se existe informação experimental das fases, ela pode ser usada para ponderar a integral. A informação de fases experimentais pode ser expressa em termos dos coeficientes de Hendrickson-Lattman:

$$p(\alpha) \propto \exp[A \cos(\alpha) + B \sin(\alpha) + C \cos(2\alpha) + D \sin(2\alpha)] \quad [9.11]$$

E a integral sobre as fases pode ser feita como mostrado na seguinte equação:

$$p(|F|; F_C) = \int_0^{2\pi} p_{\text{experimentais}}(\alpha) p(|F|, \alpha; F_C) d\alpha \quad [9.12]$$

Agora a verossimilhança depende da fase de F_C . Na figura 9-11, uma curva de probabilidade de fases é mostrada à esquerda e, à direita está a verossimilhança como uma função de F_C , com suas partes real e imaginária. Quando as fases calculadas são reforçadas pela distribuição de fases experimentais, o ganho na verossimilhança é muito maior.

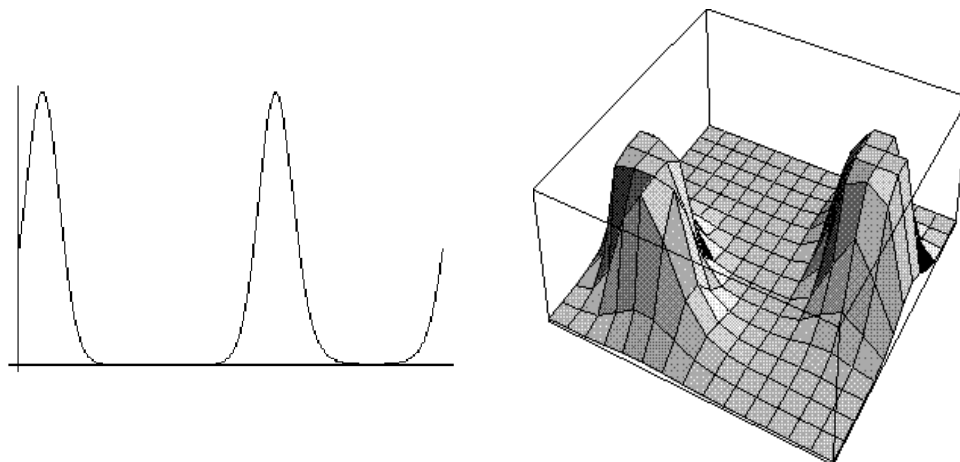


Figura 9-11: À esquerda é mostrada uma curva de probabilidades de fases e à direita, novamente a função de verossimilhança como uma função de F_c com suas partes real (horizontal) e imaginária (vertical) – agora a verossimilhança sofre um ganho indicado pelos picos no lugar da depressão.

O método da verossimilhança que usa a informação das fases experimentais é chamado MLHL de Maximum Likelihood com coeficiente de Hendrickson-Lattman e está disponível nos programas X-PLOR, CNS e REFMAC. É muito mais poderoso do que as opções que não utilizam as fases e deve ser usado sempre que alguma informação experimental de fases está disponível.

9. RESTRIÇÕES E CONSTRIÇÕES USADAS NO REFINAMENTO

Um grande problema na determinação de uma estrutura macromolecular é que a quantidade de parâmetros a serem determinados é muito maior do que a de observações na faixa de resolução na qual os cristais de proteína normalmente difratam. Se forem refinados apenas as posições e B-fatores de todos os átomos, o refinamento será pobre e o modelo atômico resultante será provavelmente muito deficiente.

Uma forma de se driblar isso é adicionando-se “observações” na forma de restrições ou reduzindo-se o número de parâmetros aplicando-se constrições ao modelo de alguma maneira. Restrições típicas incluem comprimentos e ângulos de ligações e distâncias de contatos de van der Waals. Existem também restrições para manter a planaridade dos grupos planares e a quiralidade dos centros quirais. As restrições são computadas como termos conhecidos durante o refinamento e são ponderadas de acordo com os desvios dos

valores ideais estimados a partir das estruturas de alta resolução encontradas nos bancos de dados.

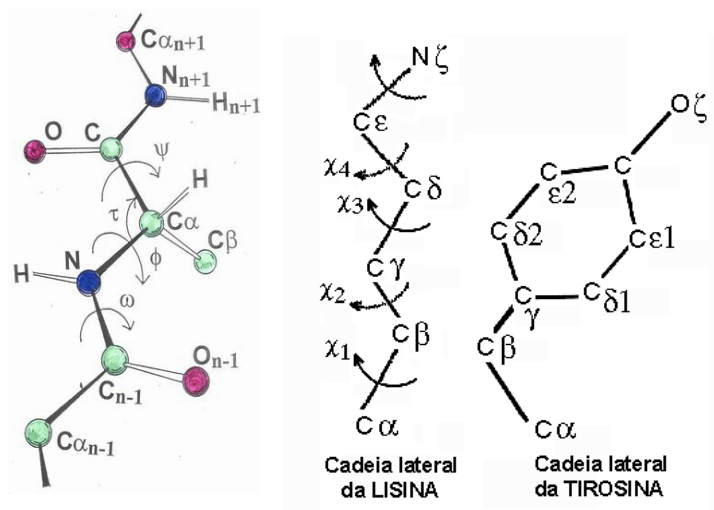


Figura 9-12: Alguns dos comprimentos e ângulos de ligações disponíveis nas bibliotecas usadas no refinamento. Esses e outros parâmetros que definem a estereoquímica e a quiralidade podem ser restringidos durante o refinamento.

As restrições são um pouco diferentes. Quando uma estrutura é restringida, na verdade a parametrização é mudada para reduzir o número total de parâmetros. Uma possibilidade é ajustar apenas os ângulos de torção ao invés de todas as coordenadas x,y,z individuais. Outro caso muito comum ocorre quando existem múltiplas cópias de uma molécula na unidade assimétrica do cristal. O modelo pode ser uma cópia da molécula replicada por rotações e translações para criar as outras cópias. Se, por exemplo, existem três cópias na unidade assimétrica, o número de parâmetros ajustáveis pode ser reduzido por um fator de três.

10. DINÂMICA MOLECULAR

Durante o processo de refinamento a função a ser minimizada vai convergindo para um mínimo e, se o modelo de partida não for muito diferente do modelo real, então o refinamento logo converge para a solução correta. Entretanto, se a discrepância entre os modelos for muito acentuada, o refinamento pode convergir para um mínimo local e ficar aprisionado ao invés de atingir o mínimo global ou verdadeiro mínimo.

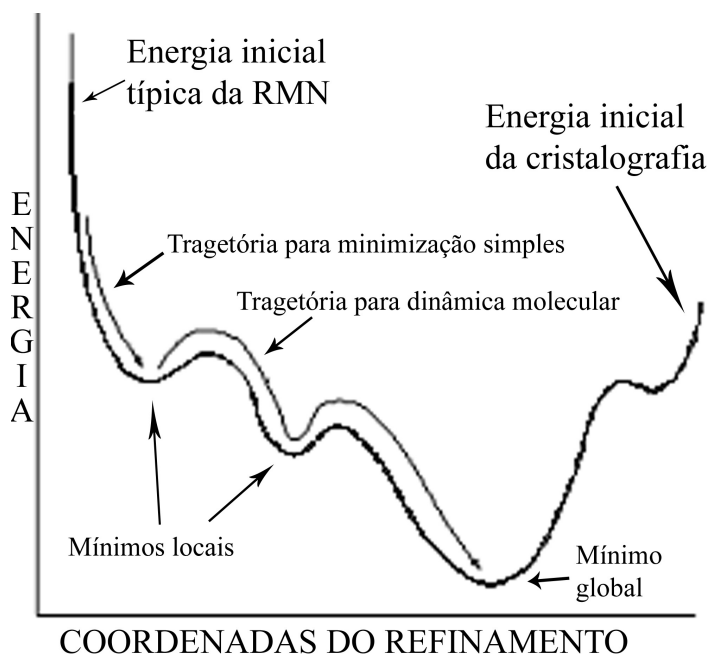


Figura 9-13: Mudanças na energia durante o refinamento.

Para evitar essa situação seria necessária uma técnica de refinamento que trabalhasse tanto no sentido de maximizar quanto de minimizar a função permitindo a transposição de barreiras de energia. Essa técnica foi introduzida por Brünger (BRÜNGER, 1987; BRÜNGER & NILGES, 1993) e consiste na dinâmica molecular na qual o comportamento dinâmico de um sistema de partículas é simulado. Essa simulação envolve um arranjo de estruturas que é energeticamente submetido à uma dada temperatura e pressão: a distribuição de energias das estruturas segue a lei de Boltzmann, a qual diz que o número de estruturas com uma energia potencial ϵ_{pot} é proporcional à $\exp[\epsilon_{\text{pot}}/kT]$ sendo k a constante de Boltzmann e T a temperatura absoluta. A energia potencial depende das posições relativas dos átomos e é calculada com base nas funções de energia potencial conhecidas, as quais são sempre aproximadamente acuradas. Os termos da energia potencial levam em conta as energias das ligações e dos ângulos de ligação (a aproximação harmônica é usada): (DRENTH, 1994)

$$E_{\text{ligação}} = \frac{1}{2} K_{\text{ligação}} (b - b_0)^2 \quad [9.13]$$

$$E_{\text{ângulo-de-ligação}} = \frac{1}{2} K_{\tau} (\tau - \tau_0)^2 \quad [9.14]$$

Sendo b_0 a distância correspondente ao mínimo de energia e b a distância real entre os átomos e τ_0 o ângulo de ligação correspondente ao mínimo de energia e τ o ângulo de ligação real entre eles.

Da mesma maneira, o termo da energia de torção é expresso por:

$$E_{torção} = \frac{1}{2} K_{\xi} (\xi - \xi_0)^2 \quad [9.15]$$

E a energia para um ângulo diedro é:

$$E_{diedro} = K_{\theta} [1 + \cos(m\theta + \delta)] \quad [9.16]$$

θ é o ângulo de rotação e δ o ângulo de fase determinados no ponto zero da rotação, m é a frequência de rotação (3 para a ligação CC no etano, por exemplo). Essa é uma representação bem simples da energia do diedro e para a energia de interação de van der Waals um termo de atração e um termo de repulsão devem ser levados em conta:

$$E_{van-der-Waals} = (A \times r^{-12}) + (B \times r^{-6}) \quad [9.17]$$

Também se adiciona um termo eletrostático:

$$E_{eletrostática} = \sum \frac{q_i \times q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \quad [9.18]$$

ϵ_0 é a constante de permissividade ($4\pi\epsilon_0 = 1.11264 \times 10^{-10} \text{C}^2 \text{N}^{-1} \text{m}^{-2}$), ϵ é a constante dielétrica local, a qual tem um valor estimado entre 1 e 80 dependendo da posição dos átomos i e j com distância interatômica r_{ij} .

Os cálculos da dinâmica molecular numa molécula começam assinalando-se as velocidades dos átomos derivadas de uma distribuição de Maxwellian em uma temperatura apropriada. No tempo $t=0$ os átomos estão numa configuração inicial que tem uma energia potencial E_{pot} para a molécula inteira. Em cada átomo i na posição r_i está atuando uma força que é a derivada da energia potencial: $\text{força}(i) = -\partial E_{\text{pot}} / \partial r_i$. Usando a mecânica newtoniana, $\text{força}(i) = m_i (d^2 r_i / dt^2)$, a aceleração $d^2 r_i / dt^2$ para o átomo i com massa m_i pode ser calculada e então aplicada. Após um curto intervalo de tempo Δt , na faixa de femtosegundos ($1 \text{fseg} = 10^{-15} \text{seg}$), o processo é repetido com os átomos numa nova posição. Se o número de passos é suficiente (1000-10000), o mínimo de E_{pot} é encontrado e a

informação sobre o comportamento dinâmico dos átomos na molécula é obtido. Se a temperatura do sistema é elevada a um valor alto, os átomos terão uma velocidade alta e uma energia cinética alta e podem transpor altas barreiras de energia. A idéia básica da dinâmica molecular é elevar a temperatura suficiente para os átomos transporem barreiras de energia e então resfriar lentamente o sistema para se aproximar do mínimo global de energia. Num método alternativo a temperatura é mantida constante mas todos os termos de energia potencial são diminuídos por um fator de escala global c , com $0 \leq c \leq 1$. Isto é baseado na lei de Boltzman. Um aumento em T com E_{pot} constante tem o mesmo efeito de diminuir E_{pot} com T constante.

Um exemplo típico de barreira de energia ocorre no ‘flipping’ de um peptídeo plano. É impossível para as outras técnicas de refinamento transpor essa barreira, mas ela pode ser facilmente transposta com dinâmica molecular.

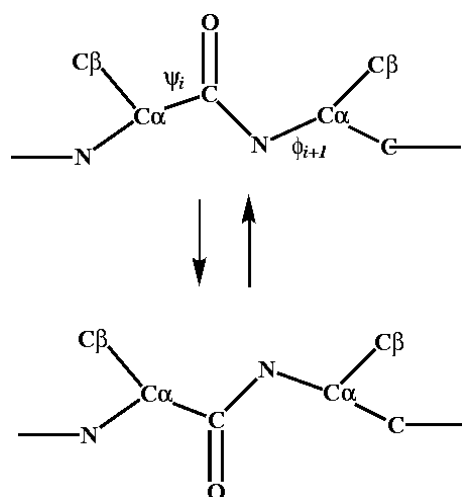


Figura 9-14: A orientação de um peptídeo plano está intimamente ligada à localização no diagrama de Ramachandran dos dois resíduos ligados a ele. O ‘flipping’ no peptídeo plano entre os resíduos i e $i+1$ muda o ângulo ψ do resíduo i e o ângulo ϕ do resíduo $i+1$ por 150 - 180° . Se o resíduo $i+1$ tem um valor de ϕ negativo, um ‘flip’ errado do peptídeo quase sempre resulta em que o resíduo caia numa região desfavorável do diagrama.

Na aplicação da dinâmica molecular ao refinamento cristalográfico, os fatores de estrutura calculados do sistema são restringidos para os fatores de estrutura com valores marcados, adicionando-se o termo cristalográfico de discrepância:

$$E_x = k_x \times \sum_{hkl} [|F_{obs}(hkl)| - k |F_{calc}(hkl)|]^2 \quad [9.19]$$

como um termo de pseudoenergia para a energia potencial do sistema. k_x é um fator de ponderação e k é o fator de escala entre $|F_{obs}(hkl)|$ e $|F_{calc}(hkl)|$ com $|F_{calc}(hkl)|$ calculado para o presente modelo.

A energia total ($E_x + E_{pot}$) é minimizada durante o refinamento. No princípio, F_{calc} deve ser calculado após cada intervalo de tempo Δt . O mesmo é verdadeiro para as derivadas de $|F_{calc}|$ com respeito aos parâmetros atômicos. Essas derivadas são requeridas para calcular a força. Entretanto, os parâmetros atômicos mudam muito pouco durante Δt e, por isso, é suficiente calcular F_{calc} e suas derivadas apenas após uma mudança significativa nas coordenadas ser observada, por exemplo, 1/10 da resolução.

Uma vantagem do refinamento por dinâmica molecular é o grande raio de convergência, o qual pode ser de muitos Ângstrons. Em outras palavras, dinâmica molecular é capaz de mover grupos de átomos por muitos Ângstrons com relação ao modelo inicial, correspondendo ao mínimo de energia potencial, sem qualquer intervenção manual. Erros grandes no modelo inicial podem ser corrigidos. Isso acelera o refinamento apreciavelmente. Apesar do método demandar muito tempo computacional, com os modernos computadores de alta velocidade isso não é mais um problema.

11. OS MAPAS DE DENSIDADE ELETRÔNICA

O refinamento em si não consegue corrigir os erros mais grosseiros nem construir uma nova estrutura. O papel principal do refinamento é produzir novos mapas de densidade eletrônica para análise e correção de erros no modelo. Os mapas gerados após o refinamento são tipicamente melhores que os mapas iniciais e vão sendo melhorados conforme o modelo fica melhor. Dois tipos de mapas são particularmente úteis:

Mapa ($2F_O - F_C$): mostra a melhor estimativa da densidade eletrônica para a estrutura atual. Esse é o mapa no qual o modelo é construído e ajustado. (Os mapas gerados pelo refinamento usando o método da máxima verossimilhança têm amplitude $(2mF_O - DF_C)$ sendo que m é a figura de mérito e D é relativo a σ_A .)

Mapa da diferença ($mF_o - DF_C$): mostra a melhor estimativa da diferença entre a verdadeira estrutura e o modelo atual. Idealmente, densidade positiva indica átomos que devem ser adicionados e densidade negativa que eles devem ser removidos (ou movidos e ajustados).

Durante a reconstrução, é útil exibir ambos os mapas, e os contornos positivos e negativos do mapa de diferença em diferentes cores. Em geral o mapa da diferença indica alguma coisa que está errada ou faltando no modelo e o outro mapa mostra o que fazer sobre isso. Nos primeiros ciclos de reconstrução, pode ser útil exibir o mapa experimental, pelo menos na regiões mais difíceis, pois esse mapa não é influenciado pelo modelo.

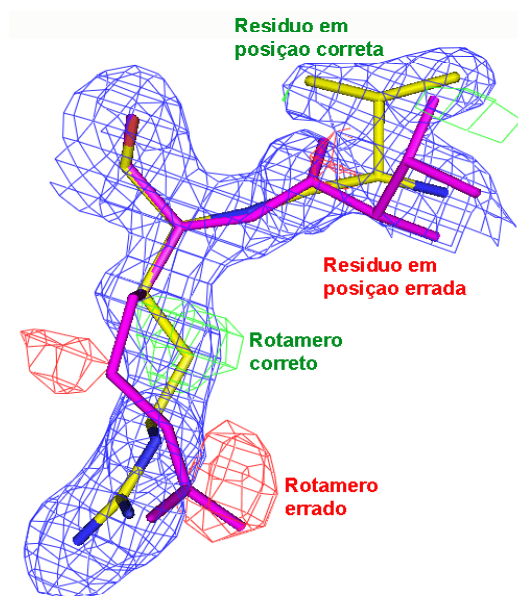


Figura 9-15: Mapa ($2mF_o - F_c$) em azul superposto com o mapa da diferença para uma região de um modelo mostrando picos de densidade positiva em vermelho e negativa em verde. São mostradas a arginina A24 e a leucina A23 do modelo de CTI em rosa antes do refinamento e em amarelo no modelo final.

Embora a maior parte de uma estrutura geralmente seja clara, é comum algumas partes serem pouco ordenadas e difíceis até mesmo impossíveis de modelar. Se nenhuma densidade está presente então essa parte não pode ser construída, embora algumas vezes a densidade apareça com fases melhoradas pelo melhoramento do modelo. O maior problema é a densidade que está presente mas não é interpretável. Ela pode representar conformações múltiplas que se sobrepõem ou a presença de ligantes.

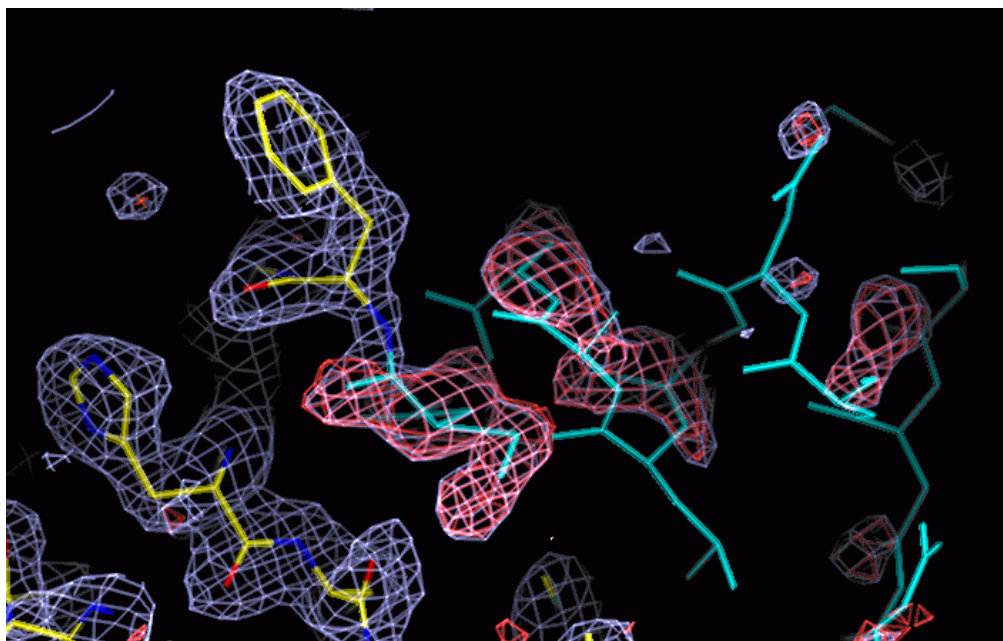


Figura 9-16: Boa densidade à esquerda e nenhuma densidade à direita. No meio densidade que é difícil de explicar com o modelo atual. Aqui a densidade positiva é mostrada em vermelho. A parte da molécula colorida em cian foi omitida do refinamento e do cálculo dos mapas (colocando a ocupância igual a zero) mas é modelada a partir de uma segunda molécula na unidade assimétrica para a qual essa região é mais bem ordenada.

12. MOLÉCULAS DE ÁGUA E OUTRAS MOLÉCULAS

As águas são uma importante parte da estrutura: uma molécula de água bem ordenada contribui mais para o espalhamento de raios-x do que uma parte da macromolécula pobremente ordenada. Isso é claramente visível nos mapas experimentais e, particularmente, nos mapas da diferença, pelo menos em médias e altas resoluções. A resoluções piores que 2,8 - 3 Å, águas não podem ser consistentemente colocadas e nesses casos deve-se usar o R_{free} como guia já que, de uma forma geral, a adição de águas melhora o modelo e o R_{factor} conseqüentemente. É uma boa idéia inspecionar cada molécula de água que for adicionada para evitar a inserção de águas em regiões de densidade que seria mais bem interpretada como outro tipo de ligante ou partes da macromolécula que ainda não foram construídas. As águas podem ser inseridas automaticamente em programas, como CNS ou ARP por exemplo, mas devem ser manualmente checadas. Alguns critérios podem ser usados como guias na inserção/inspeção das moléculas de água: as moléculas de água devem apresentar densidade eletrônica clara acima de 3σ no mapa ($2mF_o - DF_C$), deve

fazer pelo menos uma ligação de hidrogênio com a proteína ou com outra molécula de água.

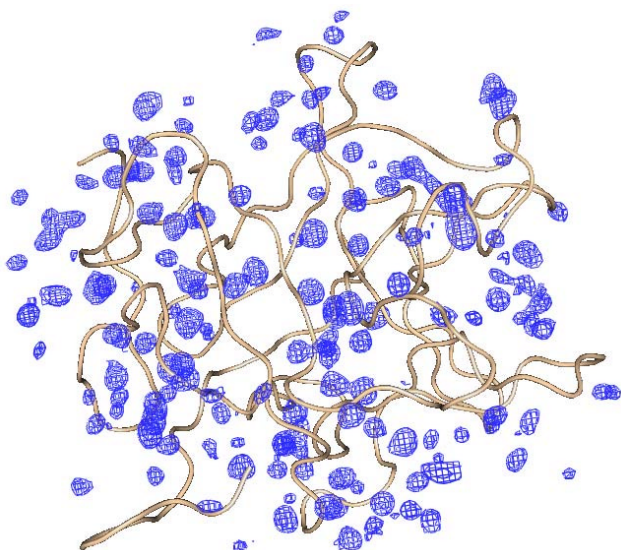


Figura 9-17: As águas têm um aspecto esférico no mapa de densidade positiva mostrado aqui em azul contornado a 1,5sigma. São mostradas as águas localizadas para o modelo de DrTI.

Outros ligantes esperados ou inesperados podem aparecer nos mapas e devem ser adicionados ao modelo quando eles podem ser interpretados. (É uma boa idéia verificar o que continha a solução de cristalização.)

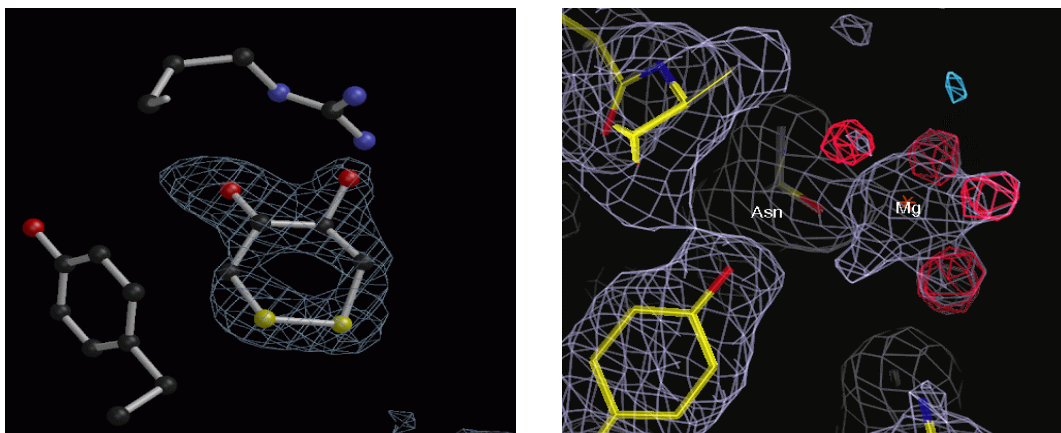


Figura 9-18: À esquerda, uma molécula de DTT oxidado. A resolução de 1,7Å permitiu o seu reconhecimento e modelagem. À direita, um cátion magnésio numa coordenação octaédrica ligada ao oxigênio da cadeia lateral de uma asparagina da proteína. Uma água tinha sido construída na posição do magnésio. O mapa (2FoFc) mostra apenas uma água e o mapa da diferença de densidade positiva mostra as outras quatro.

13. VALIDAÇÃO: MEDINDO A CONFIABILIDADE DO MODELO

Após o modelo estrutural da macromolécula ter sido refinado, ele ainda pode conter erros que tenham escapado durante a interpretação dos mapas de densidade eletrônica, particularmente nas regiões onde a densidade eletrônica é fraca. Alguns erros são óbvios e devem causar suspeitas imediatas, por exemplo: a presença de hélices de mão-esquerda quase sempre é fácil de perceber. A maioria dos programas de modelagem segue a regularização da geometria, mas não garante a boa qualidade global do modelo final. Como vimos, uma boa impressão qualitativa da confiabilidade do modelo estrutural pode ser obtida pela inspeção dos mapas de densidade: a conectividade da cadeia principal e das cadeias laterais, a distorção dos átomos de oxigênio das carbonilas em relação à cadeia principal e a colocação das cadeias laterais na densidade eletrônica.

Uma impressão mais quantitativa da confiabilidade da estrutura é obtida a partir de índices residuais cristalográficos como os R_{factor} e R_{free} . O R_{factor} (BRÄNDÉN & JONES, 1990) mede a concordância entre os fatores de estrutura observados e calculados e é calculado como a média da discrepância fracional entre eles.

$$R_{\text{factor}} = \frac{\sum_{hkl} || F_{\text{obs}} | - k | F_{\text{calc}} ||}{\sum_{hkl} | F_{\text{obs}} |} \quad [9.20]$$

sendo k o fator de escala. Entretanto o R_{factor} cristalográfico normal pode ter valores surpreendentemente baixos no refinamento de modelos estruturais que se mostram depois incorretos. Isso acontece porque o modelo pode ter sido ‘forçado’ a se ajustar aos dados experimentais durante o refinamento (overfitting), especialmente em resoluções moderadas. BRÜNGER (1992a) sugeriu resolver essa situação com a introdução do R_{free} , o qual não é influenciado pelo processo de refinamento. As reflexões são divididas num conjunto teste (T) de reflexões únicas e num conjunto de trabalho (W). O conjunto teste é uma seleção aleatória de 5-10% das reflexões observadas. O refinamento usa apenas o conjunto de trabalho e o R_{free} é calculado usando apenas o conjunto teste de observações:

$$R_{\text{free}} = \frac{\sum_{hkl \in T} || F_{\text{obs}} | - k | F_{\text{calc}} ||}{\sum_{hkl \in T} | F_{\text{obs}} |} \quad [9.21]$$

sendo que $hkl \subset T$ significa todas as reflexões pertencentes ao conjunto teste. Brünger mostrou que existe uma alta correlação entre o R_{free} e a honestidade do modelo.

Esses residuais cristalográficos devem ser acompanhados durante todo o refinamento e não apenas para o modelo final. Além disso, existem alguns outros parâmetros, como comprimentos de ligação, ângulos de ligação e fatores de temperatura, cujos valores também podem e devem ser monitorados durante o refinamento.

O fator de temperatura é um fator de correção que deve ser multiplicado pelo fator de espalhamento a fim de se levar em conta a vibração dos átomos ao redor da posição de equilíbrio provocada pela temperatura. No caso mais simples no qual os componentes de vibração são os mesmos em todas as direções, a vibração é chamada isotrópica e o fator de temperatura **B** é calculado pela equação 9.22:

$$B_{iso} = 8\pi^2 \times \overline{u^2} \quad [9.22]$$

Sendo $\overline{u^2}$ o quadrado da média do deslocamento provocado pela vibração atômica.

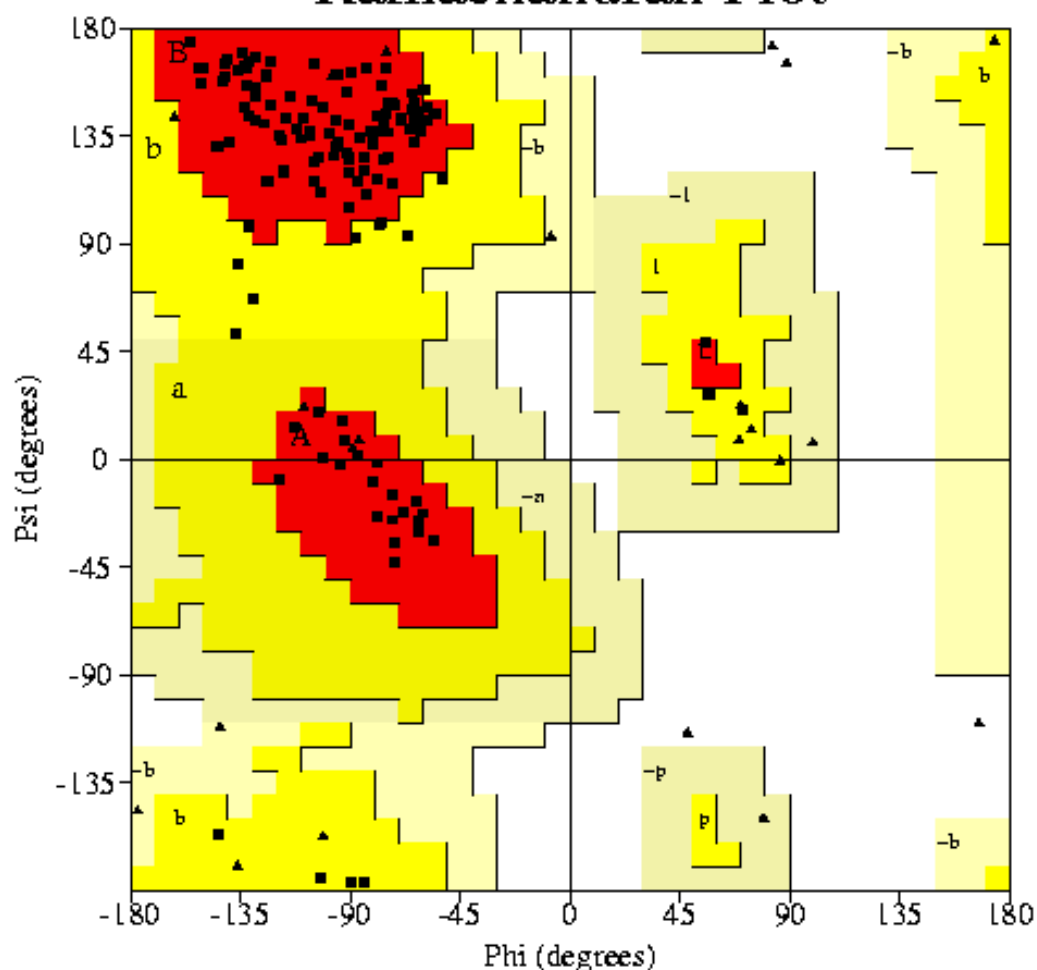
Os próprios programas de refinamento, como REFMAC (CCP4, 1994) ou CNS (BRÜNGER *et al.*, 1998), fornecem listas com os valores para esses parâmetros a cada ciclo. Entretanto, esses parâmetros checados pelos programas de refinamento são geralmente restringidos (restrições e constrições) durante o refinamento, de forma que nem sempre a obtenção de valores favoráveis garante a qualidade do modelo pois existe a possibilidade de um ajuste ‘forçado’ do modelo aos parâmetros aceitáveis imposto pelas restrições. Por outro lado, existem alguns atributos conformacionais, como ângulos de torção e volumes de empacotamento, que geralmente não são restringidos durante o refinamento e que têm sido usados no desenvolvimento de vários pacotes de programas de validação. Dentre esses estão os pacotes PROCHECK (LASKOWISK *et al.*, 1993), SQUID (OLDFIELD, 1992), WHATCHECK (HOOFT *et al.*, 1996d) e PROVE (PONTIUS *et al.*, 1996). O propósito dos três primeiros pacotes de programas é (1) verificar a sintaxe do arquivo para deposição nos bancos de dados como o PDB, (2) checar a consistência do modelo atômico com as bibliotecas atuais e verificar a presença de discrepâncias que devem ser investigadas, (3) detectar erros grosseiros na estrutura ou partes faltantes, (4) checar a presença de anomalias pontuais de estereoquímica e (5) verificar a qualidade

estereoquímica global. O pacote de programas PROVE avalia os desvios dos volumes atômicos padrões.

Uma ferramenta bastante usada para avaliar a estereoquímica do modelo é o chamado diagrama de RAMACHANDRAN (1963) que pode ser calculado pelo pacote de programas PROCHECK. Nesse diagrama os ângulos diedros ϕ e ψ para cada resíduo são plotados numa matriz quadrada que é dividida em regiões permitidas (resíduos com geometria favorável) e regiões proibidas. Para estruturas bem refinadas e com uma resolução razoável, aproximadamente todos os valores de ϕ/ψ devem estar localizados em regiões favoráveis. Devido à falta de cadeia lateral, os resíduos de glicina podem adotar uma ampla faixa de ângulos ϕ e ψ . Com exceção das glicinas, os resíduos fora das regiões favoráveis do diagrama devem ter sua densidade eletrônica cuidadosamente checada.

Outros aspectos que devem ser observados com cuidado são: se a estrutura mostra poucas pontes de hidrogênio satisfeitas e se existem ângulos diedros eclipsados ou ângulos ω incomuns em cadeias laterais. Isso é energeticamente bastante desfavorável. Deve-se prestar atenção também à resíduos ou partes de resíduos com B-fatores com valores muito altos, resíduos com carga livre no interior da molécula e/ou participando de contatos de van der Waals anormalmente próximos. Essa checagem do modelo deve ser cuidadosamente feita não somente no modelo final, mas também durante todo o processo de refinamento e assim, os erros mais grosseiros vão sendo eliminados. É muito difícil determinar quando o refinamento está terminado. Idealmente o objetivo é que nada seja observado no mapa de diferença, mas na prática procura-se não deixar picos ou regiões de densidade muito discrepantes no mapa de diferença. Na prática costuma-se refinar um modelo *ad tedium*.

Ramachandran Plot



Plot statistics

Residues in most favoured regions [A,B,L]	111	88.8%
Residues in additional allowed regions [a,b,l,p]	14	11.2%
Residues in generously allowed regions [-a,-b,-l,-p]	0	0.0%
Residues in disallowed regions	0	0.0%

Number of non-glycine and non-proline residues	125	100.0%
Number of end-residues (excl. C ₁ and Pro)	4	
Number of glycine residues (shown as triangles)	22	
Number of proline residues	12	

Total number of residues	163	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Figura 9.19: diagrama de Ramachandran obtido para o modelo final de CTI.

PARTE

III

RESULTADOS E DISCUSSÃO:
A HEXOQUINASE PII E A MUP
OS INIBIDORES CTI E DrTI

CAPÍTULO 10

A HEXOQUINASE PII E A MUP

1. INTRODUÇÃO

No início do doutorado, a fim de adquirir conhecimentos práticos das técnicas de cristalização e cristalografia de proteínas, trabalhou-se com a enzima hexoquinase PII extraída de *Saccharomyces cerevisiae* e com a proteína recombinante de urina de camundongo (MUP). Essas proteínas estavam sendo estudadas pela Dra. Paula Regina Kuser que desenvolvia seu pós-doutoramento no grupo de cristalografia de proteínas no Laboratório Nacional de Luz Síncrotron sob a orientação do professor Dr. Igor Polikarpov.

A enzima hexoquinase PII se encontrava em fase de refinamento e a proteína MUP estava no estágio inicial de cristalização. Assim, foi possível participar em várias etapas do processo de resolução de estruturas protéicas e adquirir conhecimentos valiosos, muito importantes para a posterior determinação das estruturas de ambos os inibidores tipo Kunitz extraídos de *Copaifera langsdorffii* (CTI) e de *Delonix régia* (DrTI).

Desses estudos envolvendo ambas proteínas foram gerados os trabalhos reproduzidos em anexo no final da tese.

CAPÍTULO 11

CTI: O INIBIDOR EXTRAÍDO DAS SEMENTES DE *Copaifera langsdorffii*

1. INTRODUÇÃO

CTI é o inibidor extraído das sementes da árvore *Copaifera langsdorffii*, conhecida vulgarmente como copacaíba ou óleo-de-copaiba (Caesalpinioideae, Leguminosae). Essa árvore é amplamente distribuída pelo Brasil sendo muito comum nas regiões do cerrado brasileiro e é muito usada pela população dessas regiões. A sua madeira é usada na confecção de casas, móveis e outros utensílios, o seu óleo é muito utilizado como combustível para barcos e pequenas embarcações (biodiesel). Além disso, o seu óleo também possui propriedades medicinais sendo usado como auxiliar no tratamento de gastrites e outros problemas estomacais, bem como para infecções (LORENZI, 1992).

As sementes foram coletadas e o inibidor foi extraído e purificado no Laboratório de Química de Proteínas do Departamento de Bioquímica da UNICAMP, sob a orientação do prof. Dr. Sérgio Marangoni, por José Antônio da Silva, quem nos enviou todas as amostras de inibidor puro necessárias aos experimentos de cristalização. Todo os procedimentos de extração e purificação do inibidor estão descritos em SILVA *et al.*, 2001.



Figura 11-1: A *Copaifera langsdorffii*.

2. CRISTALIZAÇÃO E COLETA DOS DADOS DE DIFRAÇÃO DE CTI

Inicialmente pensava-se que CTI era, na verdade, dois inibidores diferentes (SILVA et al., 2001). Com base nas seqüências N-terminais de cada um desses ‘inibidores’, um deles pertencia à família (STI) Kunitz de inibidores, embora com uma massa molecular de cerca da metade dos demais membros dessa família. Já o outro não apresentava homologia com qualquer proteína e sugeriu-se que ele poderia pertencer a uma nova subclasse de inibidores de baixa massa molecular da subfamília Caesalpinioideae.

Por isso, inicialmente foram feitas inúmeras tentativas para cristalizar cada um dos ‘inibidores’ separadamente mas, sem nenhum resultado positivo. Foi a partir de uma amostra ‘errada’ da proteína, que continha os dois ‘inibidores’ juntos sem purificação, que foram obtidos os primeiros cristais coletados.

Os cristais adequados à coleta dos dados de difração foram crescidos pelo método de difusão de vapor em gota pendurada a 291⁰K em 0,1M de tampão acetato de sódio com pH entre 4 e 4,5 usando PEG4000 20–25% como precipitante. As gotas consistiam de iguais volumes de proteína a 10mg/mL e solução do poço.

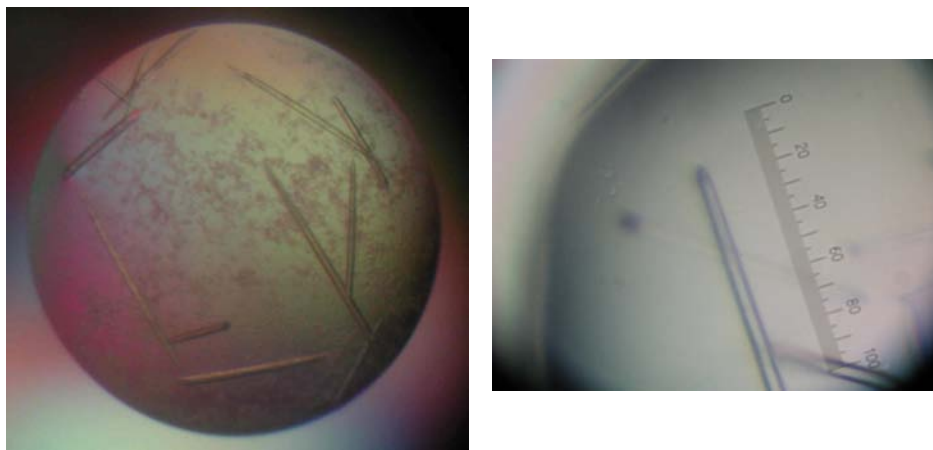


Figura 11-2: Gota com cristais tipo agulha obtidos para CTI. A ampliação à direita indica o tamanho dos cristais, a régua mede micrometros.

Tabela 11-1: Detalhes da coleta dos dados de difração para o primeiro cristal nativo de CTI. Valores estatísticos para a camada de maior resolução estão mostrados entre *parênteses*.

Grupo espacial	P4 ₃₍₁₎ 2 ₁ 2
Parâmetros da célula unitária (Å)	a=b=58,86; c=93,83
Comprimento de onda (Å)	1,535
Distância cristal-detector (mm)	100
Resolução (Å)	25,3–1,83 (1,87–1,83)
No. de observações	123604
No. De observações únicas	15024
$\langle I/\sigma(I) \rangle$	23,6 (3,3)
Mosaicidade (°)	0,57
Multiplicidade	8,2 (5,4)
Dados coletados (°)	357
Completeza (%)	99,7 (97,6)
* R _{merge} (%)	8,9 (50,8)

$$* R_{merge} = \frac{\sum_{hkl} |I - \langle I \rangle|}{\sum_{hkl} I}$$

Foram coletados dados para um cristal montado num loop de rayon, imerso por 30 segundos numa solução crio protetora (20% etileno glicol misturado no líquido mãe) e rapidamente congelado a 100⁰K no feixe de nitrogênio. Toda a coleta dos dados de difração

de raios-x foi feita na linha de Cristalografia de Proteínas (POLIKARPOV *et al.*, 1997) no Laboratório Nacional de Luz Síncrotron (Campinas – SP – Brasil), usando uma placa de imagem MAR345. Os dados foram auto-indexados e integrados com o programa DENZO (OTWINOWSKI & MINOR, 1993) e o escalonamento desses dados foi feito com o programa SCALEPACK (OTWINOWSKI & MINOR, 1993).

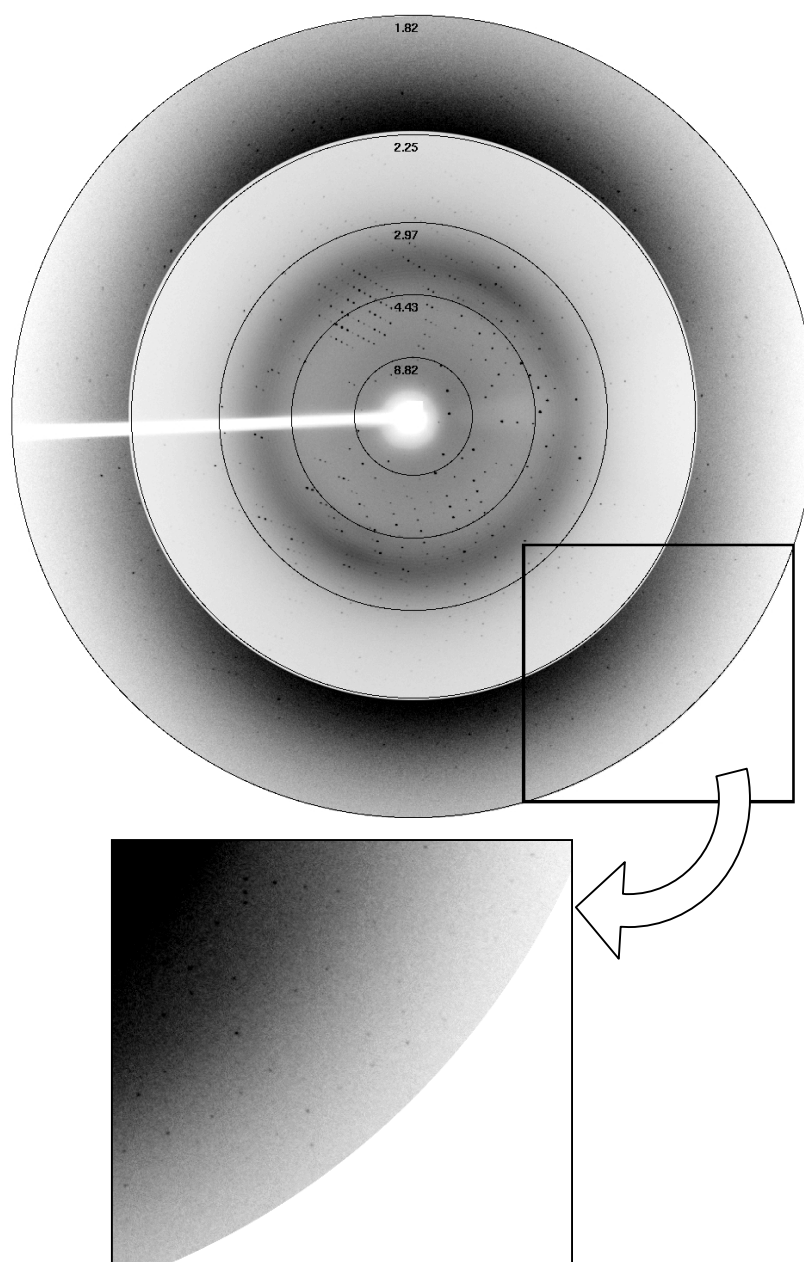


Figura 11-3: Padrão de difração do cristal. A área externa da figura de difração (entre os anéis de resolução 2,25Å e 1,82Å) é mostrada cerca de 4 vezes mais saturada que a parte interna da figura. A ampliação da parte externa da figura mostra reflexões até a resolução de 1.83Å.

3. TENTATIVAS DE RESOLUÇÃO DE CTI POR SUBSTITUIÇÃO MOLECULAR

Nesse ponto do trabalho, pensava-se que o inibidor cristalizado era TDII (SILVA et al., 2001) cujo N-terminal apresenta homologia com os inibidores da família (STI) Kunitz, mas cuja massa molecular é cerca de metade dos membros dessa família. Como a sequência de aminoácidos desse inibidor ainda estava sendo determinada, usando a sequência de aminoácidos do N-terminal foi feita uma busca nos bancos de dados por inibidores que apresentassem homologia com TDII e cuja estrutura 3D fosse conhecida a fim de se usar o método de substituição molecular para a resolução da estrutura de TDII.

Os inibidores extraídos das sementes de soja *Glycine max* (código PDB 1AVU) e de *Erythrina caffra* (código PDB 1TIE) foram usados como modelos de busca nas tentativas de se resolver a estrutura de TDII por substituição molecular. Entretanto, como TDII possui metade da massa molecular desses inibidores, os modelos foram cortados e apenas a metade que apresentou homologia com o N-terminal de TDII foi usada.

Como era de se esperar, o que logo ficará claro, todos os esforços para se resolver a estrutura de TDII dessa maneira foram frustrados. Devido a isso, decidiu-se tentar resolver a estrutura por substituição isomorfa e, usando-se a técnica de *quick-cryo soaking*, foram preparados e coletados mais dois cristais derivados, um usando césio e o outro iodo.

Os cristais usados foram preparados da mesma maneira que o primeiro já coletado. As diferenças estavam nas soluções crioprotetoras usadas e nos tempos de soaking de cada derivado. A coleta e o processamento dos dados também foram feitos seguindo-se o mesmo procedimento. A incorporação dos átomos pesados dentro do cristal foi inicialmente checada através da análise do sinal anômalo. Detalhes da preparação dos derivados bem como um resumo das informações estatísticas do processamento para os conjuntos de dados dos derivados são apresentados na tabela 11-2.

Tabela 11-2: Detalhes da preparação e informações estatísticas dos cristais derivados. Valores estatísticos para a camada de maior resolução estão mostrados entre parênteses.

	Derivado de Césio	Derivado de Iodo
Grupo espacial	P4 ₃₍₁₎ 2 ₁ 2	P4 ₃₍₁₎ 2 ₁ 2
Parâmetros da célula unitária (Å)	a=b=58,42; c=93,91	a=b=58,33; c=93,80
Comprimento de onda (Å)	1,535	1,535
Distância cristal-detector (mm)	110	110
Resolução (Å)	25,3–1,92 (1,96–1,92)	22,9–2,00 (2,05–2,00)
No. de observações	285176	312518
No. De observações únicas	12871	11449
$\langle I/\sigma(I) \rangle$	33,7 (13,9)	34,1 (13,1)
Mosaicidade (°)	0,42	0,53
Multiplicidade	12,1 (12,0)	15,0 (14,7)
Dados coletados (°)	292	360
Completeza (%)	99,5 (93,9)	99,8 (99,5)
* R _{merge} (%)	7,3 (17,0)	9,3 (23,8)
Solução crio-protetora	Solução mãe + 20% etileno glicol + 1,0M CsCl	Solução mãe + 20% etileno glicol + 0,5M NaI
Tempo de soaking (s)	300	180

$$* R_{merge} = \sum_{hkl} |I - \langle I \rangle| / \sum_{hkl} I$$

4. DETERMINAÇÃO DA ESTRUTURA CTI POR SUBSTITUIÇÃO ISOMORFA

4.1. DETERMINAÇÃO DA SUBESTRUTURA DOS ÁTOMOS PESADOS

Após a coleta dos dados de difração, a próxima etapa é determinar, para cada derivado, a posição ou a subestrutura dos átomos pesados. Para isso utiliza-se o sinal anômalo presente no conjunto de dados ou pode-se comparar cada derivado com o cristal nativo e usar as diferenças isomorfas para estimar a posição dos átomos pesados.

Foram usadas as diferenças anômalas para a determinação das posições dos átomos pesados. Para isso se analisam as quantidades $\Delta F^{\text{ano}}/\sigma(\Delta F^{\text{ano}})$ e $\Delta F^{\text{ano}}/F$ em função da resolução para cada derivado, o que está representado na figura 11-4. As diferenças anômalas normalizadas foram calculadas com o programa DREAR (BLESSING & SMITH, 1999).

Um dos fatores mais importantes nessa etapa é a alta redundância dos dados, o que permite estimar com maior confiança os erros para cada uma das reflexões e quantificar precisamente o valor das diferenças anômalas.

Essas diferenças anômalas foram usadas dentro do programa SHAKE-N-BAKE (WEEKS & MILLER, 1999) para a determinação da subestrutura dos átomos de césio e de iodo. Foram localizados 7 átomos de césio e 9 átomos de iodo sendo todos estes sítios com densidade eletrônica acima de 10 sigmas.

Na figura 11-5 é mostrada a sobreposição dos mapas da diferença anômala em torno do modelo tridimensional do inibidor, esses mapas mostram claramente os átomos de césio e de iodo. Nessa etapa é muito importante atentar-se para o fato de que a localização dos sítios dos átomos pesados em ambos os derivados deve ser feita tomando-se como referência uma mesma origem.

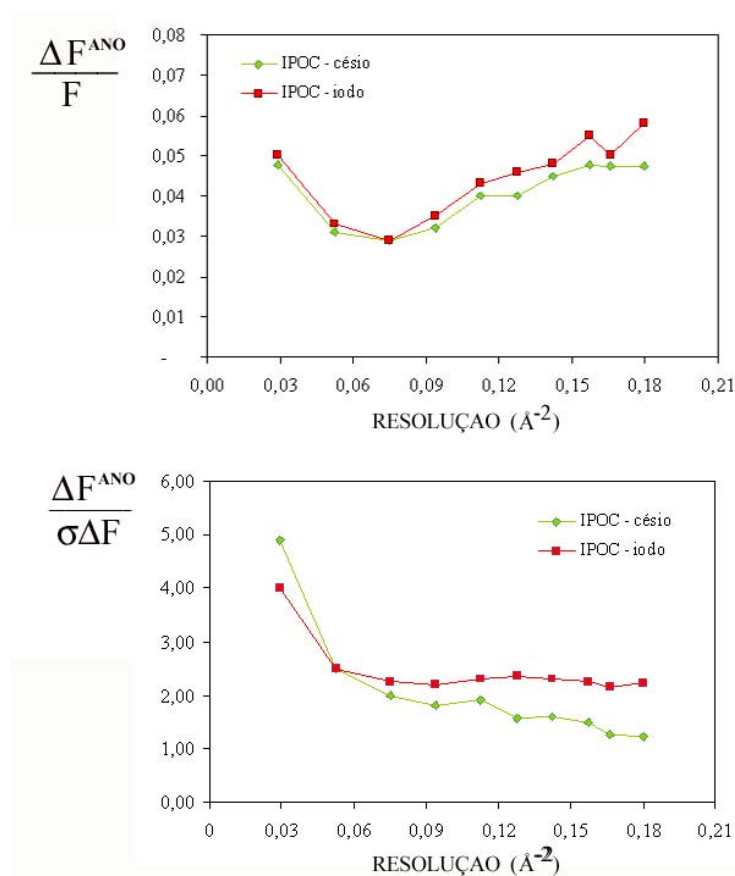


Figura 11-4: comportamento do sinal anômalo em função da resolução para cada um dos derivados.

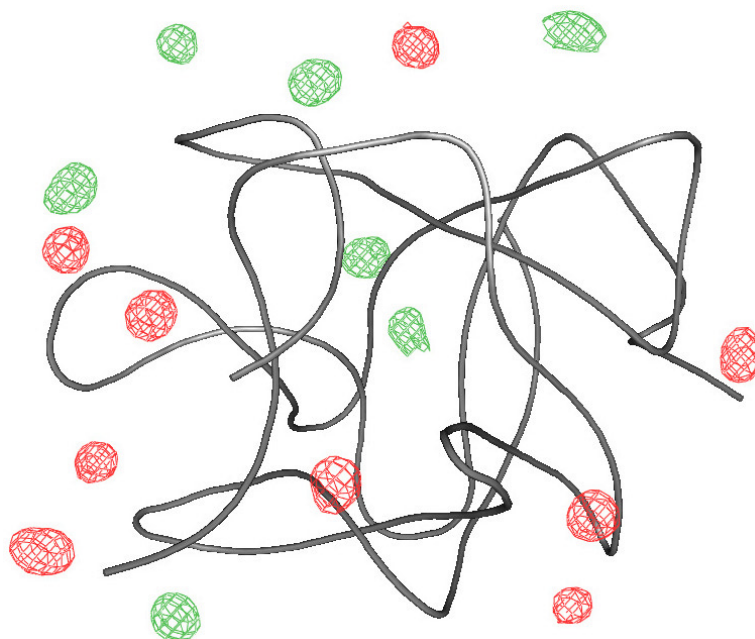


Figura 11-5: Sobreposição dos mapas de diferença anômala para os derivados mostrando os 7 sítios de césio (verde) e 9 sítios de iodo (vermelho) em torno de CTI, cujo traçado da cadeia principal é representado como um coil. Apenas picos superiores a 10 sigmas são mostrados.

Analisando o ambiente químico em torno desses átomos pesados, verifica-se que os cátions de cério interagem com as regiões negativamente carregadas presentes na superfície da proteína, principalmente com as cadeias laterais dos resíduos de ácido aspártico e glutâmico, com as moléculas de água e com o oxigênio do grupo carbonila da cadeia principal. Já os ânions de iodo estão localizados próximos aos resíduos de lisina e argininas superficiais, positivamente carregados (veja figura 11-6).

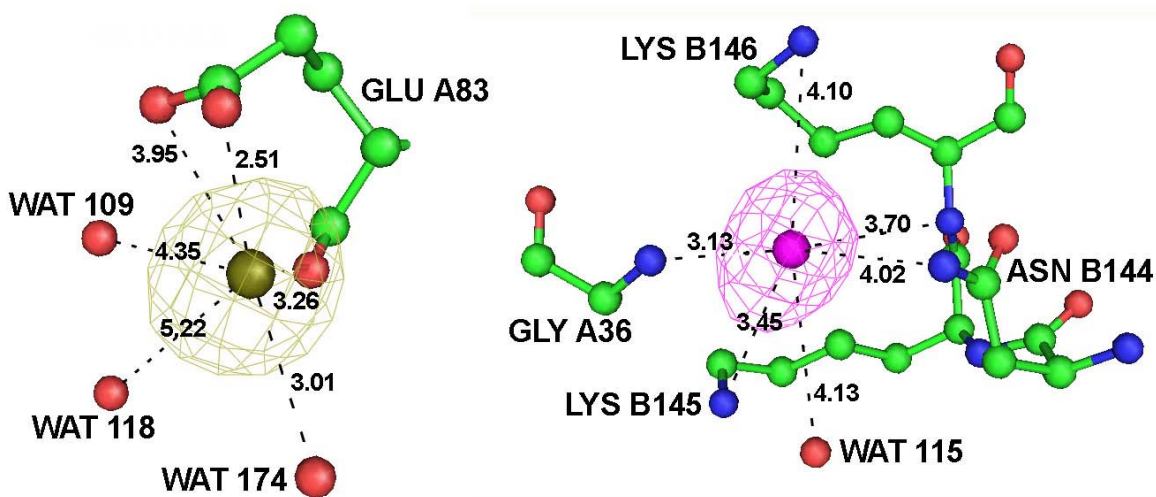


Figura 11-6: Ambiente químico de um dos cátions cério à esquerda com a sua correspondente densidade em verde e de um dos ânions de iodo à direita com a sua densidade em rosa. São mostrados os contatos que cada um deles faz com a proteína e com moléculas de água. Note que o cátion cério tende a fazer contatos com resíduos negativos e o ânion iodo com os resíduos positivos.

4.2. DETERMINAÇÃO DAS FASES DOS FATORES DE ESTRUTURA

As posições dos átomos pesados encontradas pelo programa SHAKE-N-BAKE foram inicialmente refinadas usando alguns dos programas do pacote CNS (BRÜNGER *et al.*, 1998), sendo usados mapas de diferença anômala e isomorfa no refinamento.

Essas posições refinadas foram usadas para a determinação das fases dos fatores de estrutura utilizando o programa SHARP (DE LA FORTELLE & BRICOGNE, 1997). Com essas fases foram construídos os primeiros mapas de densidade eletrônica. Para maior confiabilidade e para poder-se comparar os resultados, as fases dos fatores de estrutura para o cristal nativo foram obtidas usando SAD, SIRAS e MIRAS. As informações referentes ao processo de obtenção dos mapas de densidade eletrônica são mostradas na tabela 11-3.

Tabela 11-3: alguns parâmetros relevantes dos mapas de densidade eletrônica obtidos por SAD, SIRAS e MIRAS usando os dados do nativo e dos derivados em conjunto.

Método de faseamento	SAD	SIRAS	SIRAS	MIRAS
Conjuntos de dados	iodo	nativo+césio	nativo+iodo	nativo+césio+iodo
Faixa de resolução (Å)	22,9–	22,9–2,00	22,9–2,00	22,9–2,00
FOM antes de SOLOMON	0,39	0,48	0,53	0,63
FOM depois de	0,93	0,95	0,96	0,97
Coeficiente de correlação*	0,54	0,51	0,74	0,84

*Os coeficientes de correlação foram calculados usando um mapa ($2mF_{\text{obs}} - DF_{\text{calc}}$) obtido usando a estrutura final refinada como referência.

O mapa de densidade eletrônica obtido para o cristal nativo usando a técnica de MIRAS foi submetido à modificação de densidade eletrônica com o programa SOLOMON (ABRAHAMS & LESLIE, 1996). Essa modificação de densidade eletrônica é feita com o objetivo de melhorar a qualidade do mapa, a fim de melhorar a interpretação visual do mesmo.

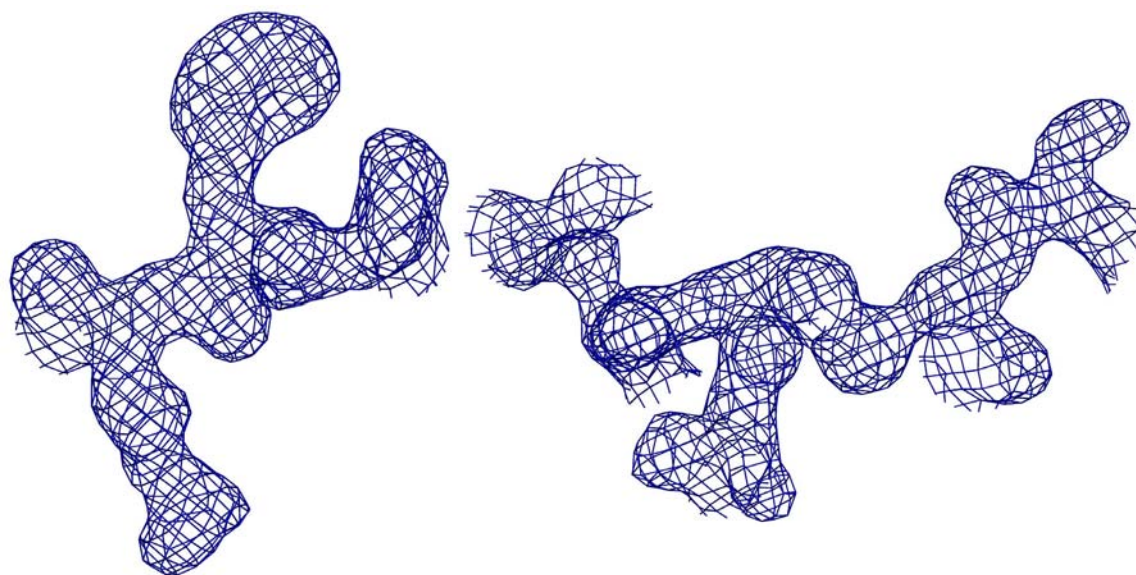


Figura 11-7: aqui são mostradas duas seções diferentes do mapa obtido após modificação de densidade eletrônica com o programa SOLOMON –note como a qualidade do mapa sugere claramente a continuidade da cadeia principal além de indicar as cadeias laterais, como no caso da fenilalanina e da lisina à esquerda. Como mostrado mais adiante (figura 11-9), à direita é mostrada a única ponte dissulfeto de CTI.

4.3. CONSTRUÇÃO DO MODELO 3D DE CTI

O novo mapa foi então usado para a construção automática do modelo usando o programa wARP (PERRAKIS et al., 1997). No programa wARP, primeiro são inseridos átomos de oxigênio em pontos de concentração de densidade eletrônica, estes vão sendo substituídos por átomos de carbono e nitrogênio e finalmente são reconhecidos padrões de ligação e esses átomos são conectados e substituídos por resíduos de glicina, alanina ou serina de acordo com o mapa de densidade eletrônica. A figura 11-8 mostra o comportamento dos parâmetros estatísticos R_{factor} e índice de conectividade bem como o aumento do número de resíduos e de cadeias traçadas automaticamente em função de cada ciclo de construção automática rodado pelo programa. Foi rodado um total de 20 ciclos de construção automática no modo 'warpNtrace' e, no último ciclo, cerca de 92% da estrutura foi traçada em 6 cadeias.

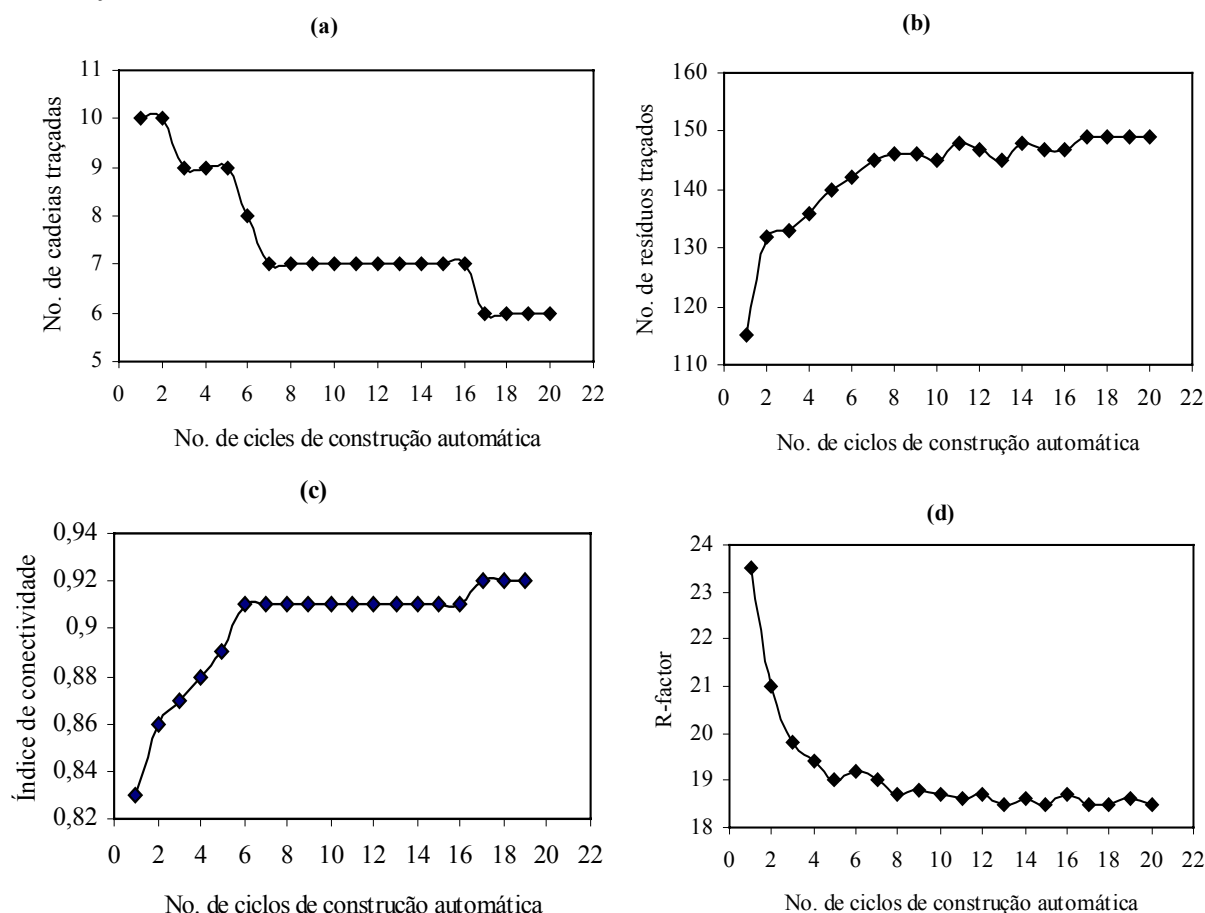


Figura 11-8: acompanhamento da construção automática do modelo de CTI: (a) número de cadeias traçadas, (b) número de resíduos traçados, (c) conectividade entre os resíduos e (d) R_{factor} em função do número de ciclos de construção automática rodados pelo programa.

4.4. AJUSTE MANUAL DO MODELO

Após a construção automática do modelo, alguns ajustes são necessários. Por exemplo, na figura 11-9, percebe-se claramente à esquerda que embora os resíduos tenham sido construídos como serinas pelo programa, pode-se ver a presença de vários átomos de oxigênio, sugerindo que esses resíduos são na verdade uma lisina seguida de uma fenilalanina. Já à direita, pode-se imaginar que os resíduos de serina bem próximos sejam na verdade duas meias cistinas e tenha-se a presença de uma ponte dissulfeto sugerida pela continuidade da cadeia mostrada pelo mapa de densidade eletrônica; isso pode e deve ser confirmado através da construção de um mapa de diferença anômala que mostre se existe a presença dos átomos de enxofre nessa região. Isso é possível pois o enxofre pode ser considerado um átomo pesado quando comparado ao carbono ou nitrogênio.

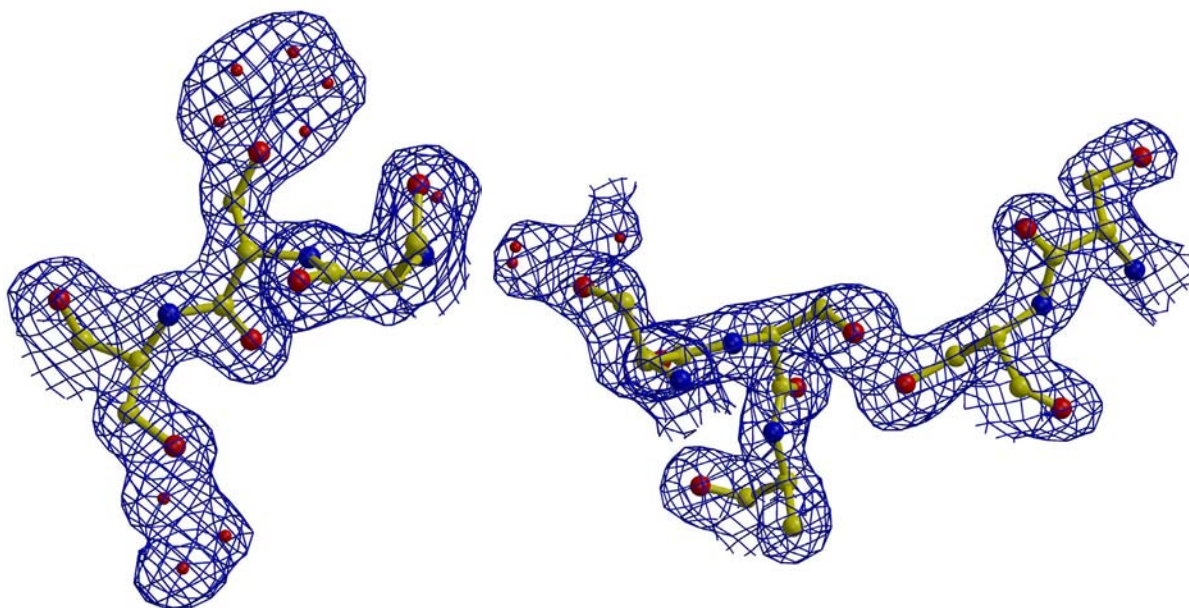


Figura 11-9: aqui são mostradas as mesmas seções do mapa após a construção automática com o programa wARP – primeiro são inseridos átomos de oxigênio em pontos de concentração de densidade eletrônica, estes vão sendo substituídos por átomos de carbono e nitrogênio e finalmente são reconhecidos padrões de ligação e esses átomos são conectados e substituídos por resíduos de glicina, alanina ou serina de acordo com o mapa de densidade eletrônica.

Assim, é necessário substituir-se os resíduos cuja densidade eletrônica seja diferente de glicina, alanina ou serina. Além disso, é necessária a construção de alguns loops, cuja presença é indicada pela interpretação dos contornos do mapa de densidade eletrônica e que

estariam conectando algumas cadeias (lembre-se que o programa foi capaz de construir 149 resíduos usando glicinas, alaninas e serinas, mas, eles estão distribuídos por 6 cadeias diferentes).

Esses ajustes, construções e definições de cadeias laterais, são feitos manualmente, usando-se programas gráficos como o O (JONES *et al.*, 1990), para a interpretação visual do mapa de densidade eletrônica.

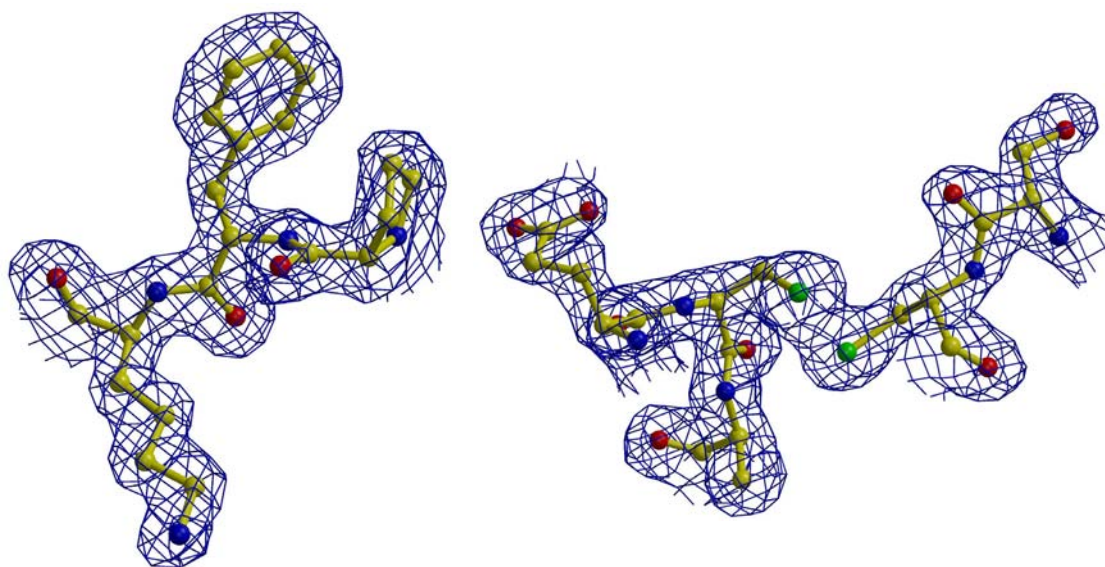


Figura 11-10: novamente são mostradas as mesmas seções do mapa com as alterações feitas manualmente.

Finalmente os resíduos foram substituídos por uma lisina e uma fenilalanina à esquerda e os oxigênios foram removidos e à direita, após a constatação da presença da ponte dissulfeto através do mapa de sinal anômalo, as duas serinas foram substituídas por meias cistinas.

A partir dos mapas de diferença anômala, foram confirmadas as presenças das metioninas A72 e A90 e a presença de apenas uma ponte dissulfeto entre os resíduos A40 e A84. Dessa forma, toda a seqüência de aminoácidos foi sendo substituída através da interpretação visual da densidade eletrônica sendo obtida uma composição '*adivinhada*' da seqüência de aminoácidos para CTI.

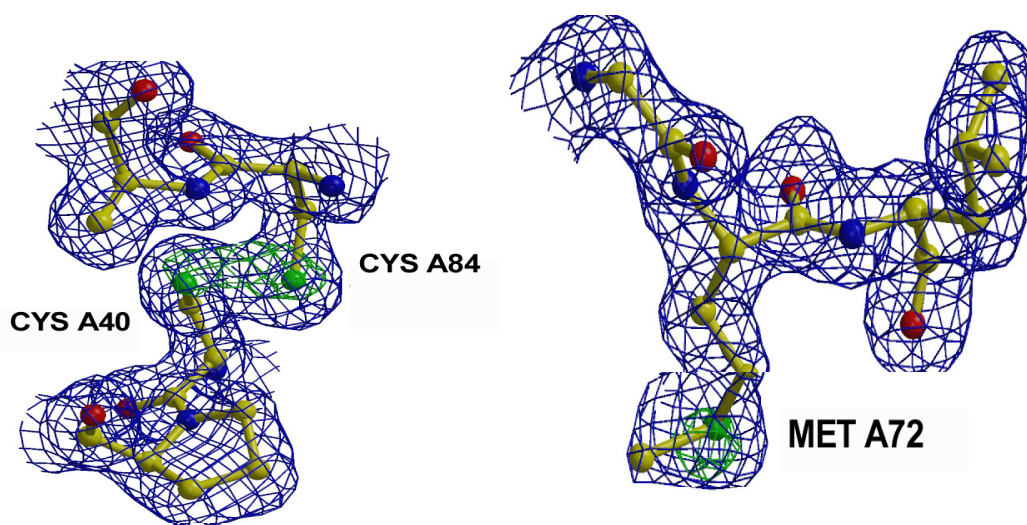


Figura 11-11: sobreposição dos mapas de densidade eletrônica (contorno em azul) e de diferença anômala a 3 sigmas (contorno em verde) confirmando a presença dos átomos de enxofre formando a única ponte dissulfeto de CTI e a presença de duas metioninas A72 e A90. Somente A72 é mostrada.

5. REFINAMENTO DO MODELO DE CTI

Esse modelo obtido para CTI foi então submetido a vários ciclos de refinamento usando o pacote de programas CNS (BRÜNGER *et al.*, 1998) empregando uma combinação de refinamento inicial de corpo rígido, seguido de refinamento com o método MLF intercalado com ciclos de dinâmica molecular. Os protocolos padrões dos programas para cada situação foram seguidos.

Cada script de refinamento foi intercalado com visualização e inspeção cuidadosa dos mapas de densidade eletrônica no programa O. Ajustes e correções necessárias foram sendo feitas manualmente.

Também foram averiguados os residuais cristalográficos e os parâmetros estereoquímicos do modelo a cada ciclo de refinamento sendo as discrepâncias cuidadosamente checadas. Para isso foram usados as saídas dos scripts de refinamento do CNS e os gráficos gerados com o conjunto de programas PROCHECK (LASKOWSKI *et al.*, 1993).

Finalmente foram inseridas 182 moléculas de água no modelo usando-se também scripts do CNS. Essas águas foram refinadas e atenciosamente checadas.

6. QUALIDADE DO MODELO DE CTI

As principais estatísticas do refinamento do modelo de CTI são fornecidas na tabela 11-4. Os valores finais dos residuais cristalográficos são 17,3% para R_{factor} e 20,3% para R_{free} .

A densidade é bastante clara, sendo contínua através de quase toda a estrutura protéica. Os ângulos diedros (ϕ, ψ) estão dentro dos limites do diagrama de Ramachandran para todos os resíduos e a estereoquímica do modelo é bastante boa. O modelo tem um desvio médio de 0,019Å nos comprimentos das ligações e de 1,6° nos ângulos das ligações.

Tabela 11-4: Dados estatísticos do refinamento que refletem a qualidade do modelo de CTI.

Parâmetros gerais	
No. de átomos de proteína	1449
No. de moléculas de água	183
No. de reflexões usadas no refinamento	13661
R_{factor} (%)	17.3
R_{free} (%)	20.3
Média dos B-fatores	
Cadeia principal (Å ²)	20.1
Cadeias laterais (Å ²)	22.4
Moléculas de água (Å ²)	39.4
Desvios médios da geometria ideal	
Comprimentos de ligação (Å)	0.019
Ângulos de ligação (graus)	1.6
Estatísticas do diagrama de Ramachandran	
Resíduos em regiões mais favoráveis	117 (70%)
Resíduos em regiões adicionais	13 (8%)
Resíduos em regiões desfavoráveis	0 (0%)

Um gráfico mostrando os valores dos fatores de temperatura é dado na figura 11-12, a qual também mostra a estrutura secundária de CTI e os seus índices de acessibilidade de acordo com o programa PROCHECK. Apenas as regiões de loop, as quais são naturalmente mais desordenadas, apresentam valores de B-fatores maiores do que o restante da proteína, porém sem valores altamente discrepantes. A densidade eletrônica nessas regiões foi checada e também é clara e contínua.

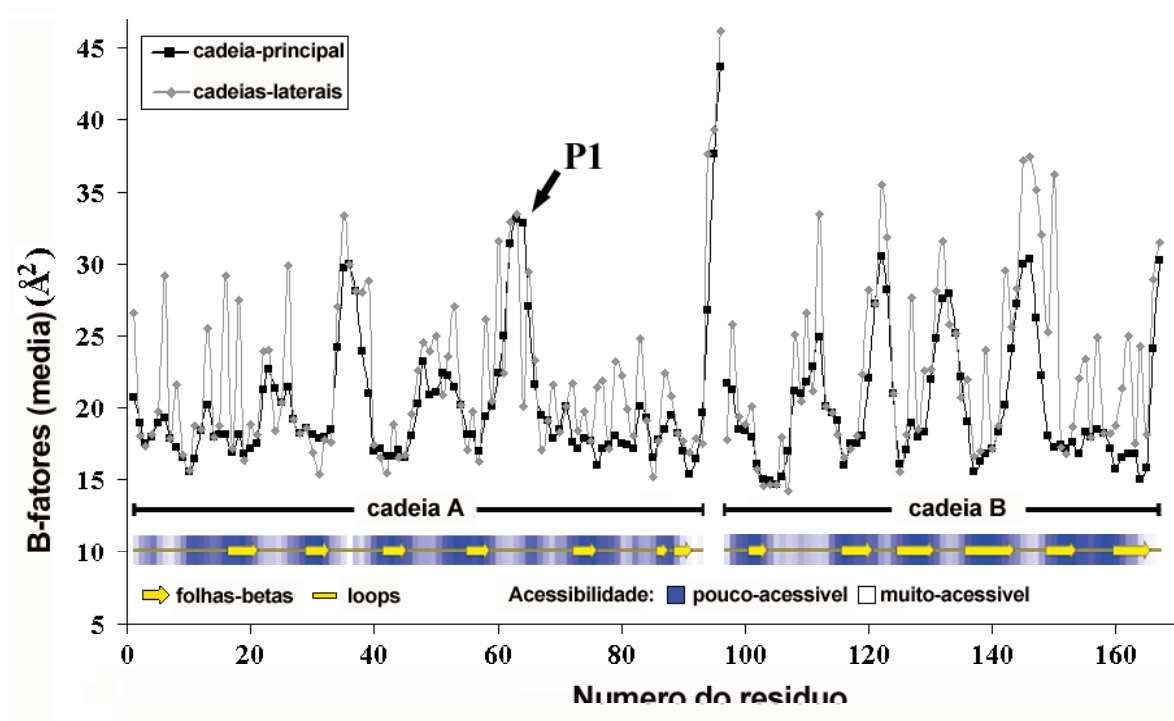


Figura 11-12: Gráfico dos B-fatores de CTI. O resíduo P1 (ARG64) do sítio ativo é indicado pela seta. As duas cadeias são indicadas. São mostrados também os elementos da estrutura secundária de CTI e os índices de acessibilidade. Como é de se esperar as regiões de loop são mais expostas e têm B-fatores e índices de acessibilidade mais altos.

7. O MODELO GLOBAL DE CTI

O modelo final de CTI é composto por 167 resíduos que formam duas cadeias polipeptídicas chamadas A (1-96) e B (97-167) e 182 moléculas de água bem ordenadas. CTI é uma proteína toda beta, formada por 12 fitas- β s antiparalelas com longos loops interconectando-as. Apresenta um arranjo estrutural conhecido como β -trefoil, que consiste de seis β -hairpins de duas fitas- β s cada, três dos quais formam uma estrutura na forma de barril- β e as outras três formam uma tampa triangular sobre o barril.

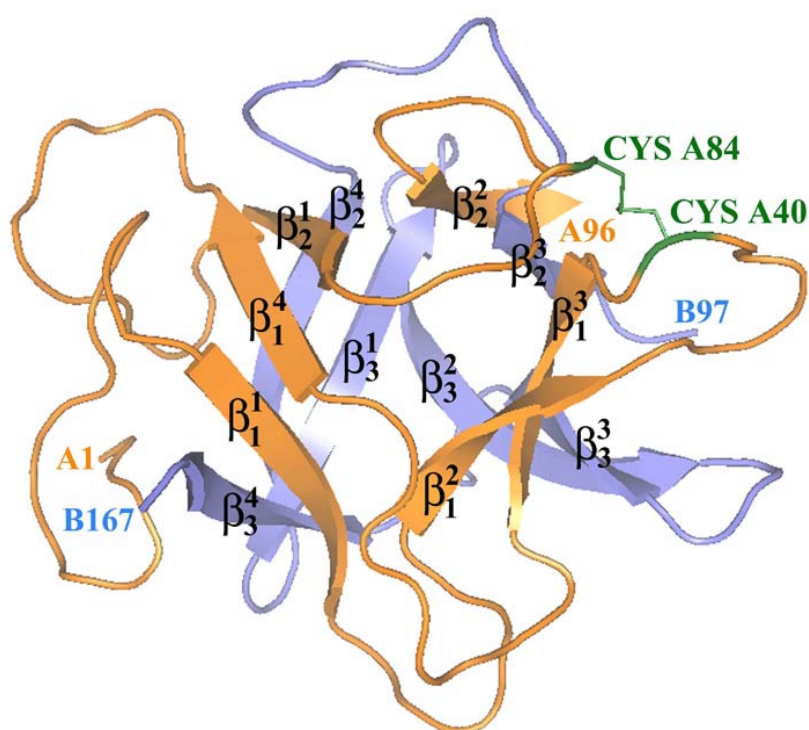


Figura 11-13: Estrutura 3D de CTI mostrando a cadeia A em laranja e a B em azul. O N- e C-terminais das cadeias são indicados e a ponte dissulfeto é mostrada em verde. As fitas- β s são numeradas de acordo com a subunidade do β -trefoil à qual elas pertencem.

A proteína pode ser subdividida em três partes aproximadamente iguais, relacionadas por um pseudo-eixo de ordem 3 aproximadamente paralelo ao eixo do barril do trefoil. Cada unidade de repetição consiste de cerca de 60 resíduos de aminoácidos, sendo que cerca de 30 formam 4 fitas- β s e os outros 30 formam os loops que conectam essas fitas- β s, estruturalmente organizados como $L_n^1\beta_n^1L_n^2\beta_n^2L_n^3\beta_n^3L_n^4\beta_n^4 = T_n$. As unidades de repetição foram nomeadas diferentemente da nomenclatura padrão estabelecida por McLachlan (1979), como T_1 , T_2 e T_3 ao invés de A, B e C, para maior clareza desde que as cadeias polipeptídicas de CTI já foram nomeadas de A e B. Quando os três subdomínios de CTI são superpostos, a concordância é particularmente boa nas regiões das fitas- β s mas não nas regiões de loop.

As três subunidades foram superpostas usando os 30 carbonos- α s topologicamente equivalentes que compõem as fitas- β s de cada subdomínio e o desvio médio entre as subunidades T_1 e T_2 foi 1,5Å, entre as subunidades T_1 e T_3 foi 1,9Å e entre as subunidades T_2 e T_3 foi 1,7Å. A superposição das subunidades estruturais que formam o β -trefoil é

mostrada na figura 11-14. Embora não haja qualquer similaridade sequencial entre essas subunidades, foi proposto que o arranjo β -trefoil evoluiu de um homotrímero progenitor o qual sofreu sucessivos eventos de duplicação gênica.

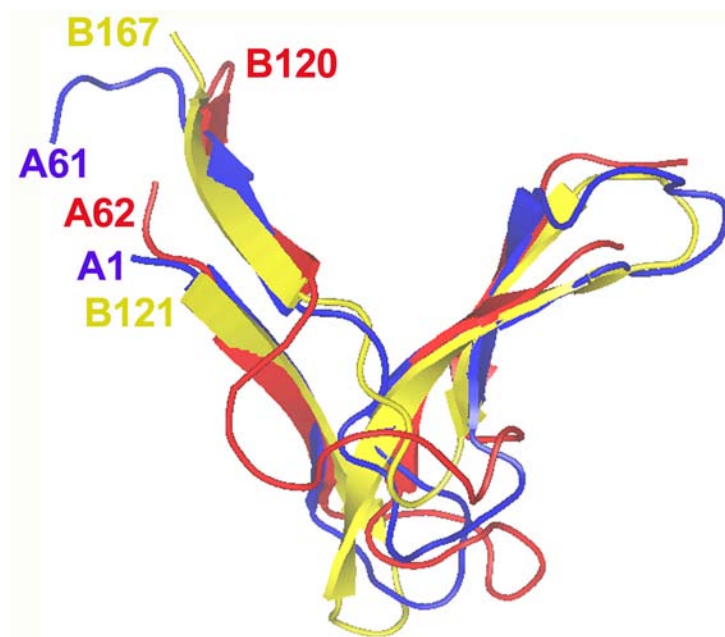


Figura 11-14: Superposição das três subunidades estruturais de repetição T_1 , T_2 e T_3 que formam o β -trefoil, mostradas em azul, vermelho e amarelo, respectivamente.

8. A ESTRUTURA SECUNDÁRIA DO MODELO DE CTI

Os elementos da estrutura secundária foram assinalados usando o programa PROMOTIF (HUTCHINSON & THORNTON, 1996), com base do padrão de pontes de hidrogênio e são descritos na tabela 11-5. As 12 fitas- β s de CTI são arranjadas em três folhas- β s. A folha- β 3 é composta de duas fitas- β s entre as cadeias A e B, formando uma interface de contato entre as porções N- e C-terminais das cadeias polipeptídicas A e B independentes; a folha- β 1 é composta por oito fitas- β s tanto da cadeia A quanto da B estando localizada na região enterrada na proteína e formando o barril- β ; e a folha- β 2 é composta por duas fitas- β s da cadeia A e está localizada entre os motivos T_1 e T_2 . Essas folhas- β s são estabilizadas por uma extensa rede de pontes de hidrogênio que mantém a rigidez e a estabilidade da estrutura, a despeito dessa ser composta por duas cadeias polipeptídicas independentes. Com base no padrão de pontes de hidrogênio ($O_i \dots N_{i+3}$), identificou-se que as regiões de loops na superfície da molécula não possuem uma estrutura

desorganizada mas sim predominantemente formada por β -turns. Existe um total de 16 β -turns na estrutura de CTI (tabela 11-5). Usando as distâncias entre os elementos da estrutura secundária envolvendo os loops e os ângulos ϕ e ψ de Ramachandran para definir a geometria e as conformações dos loops foram identificadas também 4 β -bulges e 2 γ -turns invertidas (tabela 11-5).

Tabela 11-5: Localização e classificação dos elementos da estrutura secundária de CTI.

LOOPS		FITAS β S		β -HAIRPINS		β -TURNS		BULGES	
NOME	LOCALIZAÇÃO	NOME	LOCALIZAÇÃO	LOCALIZAÇÃO	TIPO	LOCALIZAÇÃO	TIPO	LOCALIZAÇÃO	TIPO
L ₁ (T ₁)	A1–A16			L ₁ β ₁ L ₂ β ₂ (T ₁)	9:9	A4–A7	I		
		β ₁ (T ₁)	A17–A21			A11–A14	IV		
L ₂ (T ₁)	A22–A29					A23–A26	IV		
		β ₂ (T ₁)	A30–A33						
L ₃ (T ₁)	A34–A42			L ₃ β ₃ L ₄ β ₄ (T ₁)	14:14	A35–A38	II'	A33A42A43	AC
		β ₃ (T ₁)	A43–A46			A38–A41	IV	A47A28A29	AG
L ₄ (T ₁)	A47–A56					A49–A52	I		
		β ₄ (T ₁)	A57–A60						
L ₁ (T ₂)	A61–A74			L ₁ β ₁ L ₂ β ₂ (T ₂)	10:10	A69–A72	II		
		β ₁ (T ₂)	A75–A78						
L ₂ (T ₂)	A79–A88					A78–A81	VIII		
		β ₂ (T ₂)	A89–A94			A82–A85	I		
L ₃ (T ₂)	B95–B100			L ₃ β ₃ L ₄ β ₄ (T ₂)	2:4	A86–A89	VIb		
		β ₃ (T ₂)	B101–B103						
L ₄ (T ₂)	B104–B115					B106–B109	I		
		β ₄ (T ₂)	B116–B120			B110–B113	VIII		
L ₁ (T ₃)	B121–B124			L ₁ β ₁ L ₂ β ₂ (T ₃)	3:5	B121–B124	IV		
		β ₁ (T ₃)	B125–B130						
L ₂ (T ₃)	B131–B135					B131–B134	I	B131B134B135	AG
		β ₂ (T ₃)	B136–B143						
L ₃ (T ₃)	B144–B148			L ₃ β ₃ L ₄ β ₄ (T ₃)	3:5	B144–B147	I	B144B147B148	AG
		β ₃ (T ₃)	B149–B153						
L ₄ (T ₃)	B154–B159					B154–B157	II		
		β ₄ (T ₃)	B160–B165						

9. O ARRANJO β -TREFOIL

O arranjo conhecido como β -trefoil foi originalmente descrito para a estrutura de STI (MCLACHLAN, 1979) e pode ser observado num grande número de superfamílias de proteínas. Essas superfamílias incluem a superfamília das citoquinas, que inclui a família dos fatores de crescimento fibroblástico (ZHU *et al.*, 1991; ZHANG, *et al.*, 1991; ERIKSSON *et al.*, 1993; BELLOSTA *et al.*, 2001; PLOTNIKOV *et al.*, 2001) e a família das interleucinas (PRIESTLE *et al.*, 1988; GRAVES *et al.*, 1990; SCHREUDER *et al.*, 1997), a superfamília das lectinas ricina-B-like, que inclui a família das toxinas ricina-like (RUTENBER *et al.*, 1991), mais a xilanase 10A (NOTENBOOM *et al.*, 2002) e a porção N-terminal ou domínio rico em cisteínas do receptor de manose (LIU *et al.*, 2000), a superfamília das aglutininas, incluindo todas as aglutininas de plantas (TRANSUE *et al.*, 1997), a superfamília das proteínas ligadoras de actina como a hisactofilina e a fascina humana (HABAZETTL *et al.*, 1992; Ono *et al.*, 1997), e a superfamília das proteínas STI-like que inclui a família STI-Kunitz e as neurotoxinas tetânica (EMSLEY *et al.*, 2000) e botulínica (LACY *et al.*, 1998). Embora essas superfamílias de proteínas apresentem considerável homologia estrutural, elas têm muito pouca similaridade seqüencial, possuem funções distintas, diferentes localização celular, dividem pouca ou nenhuma similaridade seqüencial detectável e têm diferentes ligantes e diferentes modos de ligação a seus ligantes.

Alguns estudos filogenéticos sugerem que a causa da conservação estrutural do arranjo β -trefoil dentre essas proteínas não relacionadas, pode ser uma consequência de uma possível dupla duplicação do gene na superfamília dessas proteínas. Esses estudos fornecem evidência estatística suficiente de que as proteínas com o arranjo β -trefoil são homólogas e evoluíram de um ancestral comum por evolução divergente. (PONTING & RUSSELL, 2000)

10. A ESTRUTURA DAS MOLÉCULAS DE ÁGUA EM CTI

O conteúdo de solvente calculado usando o coeficiente de Matthews (MATTHEWS, 1968) é de 38,6% e um total de 182 moléculas de água foram inseridas no modelo de CTI. A média dos fatores de temperatura para as moléculas de água é 39,4 Å². De acordo com o programa CONTACT do pacote de programas CCP4 (1994) existem 4 moléculas de água

(W83, W120, W130, W161) com um B-fator médio de $35 \text{ \AA}^2$ que fazem 5 pontes de hidrogênio com a proteína, 20 moléculas de água que fazem 4 pontes de hidrogênio com a proteína e tem um B-fator médio de $37 \text{ \AA}^2$ e 29 moléculas de água com um B-fator médio de $38 \text{ \AA}^2$ fazendo 3 pontes de hidrogênio com a proteína. Um total de 154 moléculas de água com um B-fator médio de $39 \text{ \AA}^2$ fazem pelo menos 1 ponte de hidrogênio com a proteína e estão na primeira camada de hidratação, 7 moléculas de água com um B-fator médio de $45 \text{ \AA}^2$ fazem apenas pontes de hidrogênio com outras moléculas de água e constituem a segunda camada de hidratação.

Existem 34 moléculas de água com um B-fator de $31 \text{ \AA}^2$ enterradas em cavidades profundas da proteína, como obtido com o programa SURFACE do CCP4 (1994). Essas águas desempenham um importante papel na manutenção da estrutura global de CTI. Elas participam nas interações entre as fitas- β s antiparalelas de CTI, estando localizadas entre as fitas para manter o padrão de pontes de hidrogênio favorável das folhas- β s onde as fitas se encontram mais afastadas umas das outras. Em alguns casos as interações mediadas pelas moléculas do solvente envolvem fitas entre as cadeias A e B. É o caso por exemplo da água W2 que media a interação entre o O γ da Ser-A92 e o N da Asn-B106. Existem também algumas águas que interagem simultaneamente com diferentes fitas- β s, como por exemplo, a água W25 que interage com o N da Val-A44 que pertence à fita $\beta_3(T_1)$, com o O γ da Ser-B101 que pertence a fita $\beta_4(T_2)$ e com o grupo NH $_1$ da Arg-B150 que pertence à fita $\beta_3(T_3)$. Além disso, como se pode ver, cada fita pertence à uma subunidade estrutural diferente, isso forma uma intrincada rede de pontes de hidrogênio que reforça a importância dessas águas estruturalmente conservadas na manutenção da arquitetura global de CTI.

11. CTI É COMPOSTO DE DUAS CADEIAS POLIPEPTÍDICAS

Como já foi mencionado, CTI foi descrito inicialmente como dois inibidores diferentes (SILVA *et al.*, 2001). Entretanto, a partir da estrutura cristalográfica determinada ficou claro que ‘os diferentes inibidores’ são na verdade duas cadeias polipeptídicas não covalentemente ligadas do mesmo inibidor CTI. A figura 11-15 à esquerda mostra o mapa (2Fo-Fc) de densidade eletrônica para a região entre o C-terminal da cadeia A e o N-terminal da cadeia B onde se pode ver claramente a quebra entre elas.

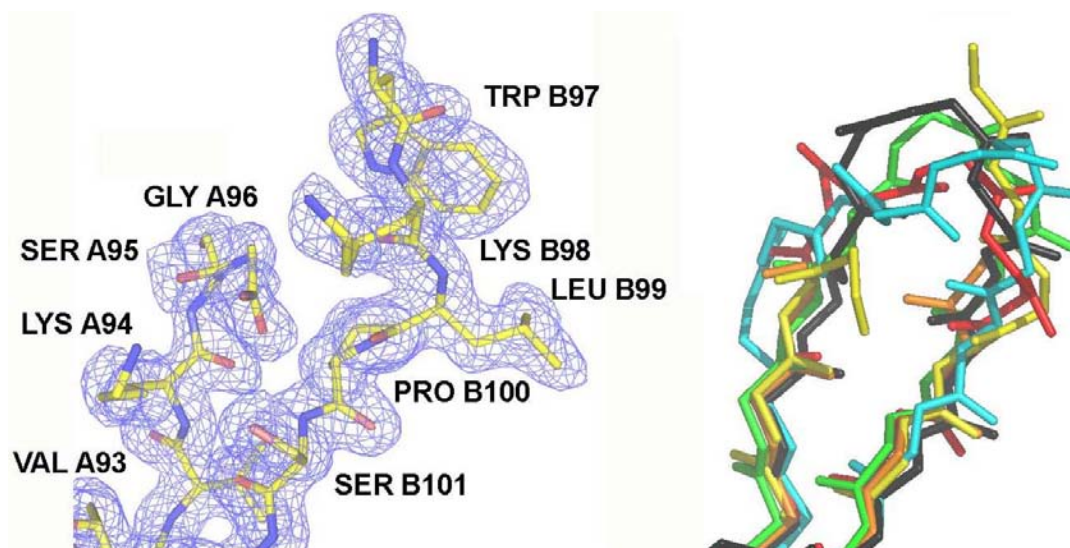


Figura 11-15: À esquerda é mostrada a região da quebra entre as cadeias A e B de CTI (amarelo) superposta com a respectiva porção do mapa (2Fo-Fc) de densidade eletrônica. À direita essa região é superposta com a região de loop equivalente dos membros da família Kunitz com STI em vermelho, ETI em laranja, WCI em verde, WBA em preto e BASI em cyan.

A existência de duas cadeias não covalentemente ligadas explica a confusão inicial causada pela presença de duas bandas com 9kDa e com 11kDa no gel de SDS-PAGE em condições desnaturantes e apenas uma banda em condições nativas. Explica também porque as primeiras tentativas de cristalização dos ‘inibidores’ separados não deram resultados. A amostra ‘errada’ que cristalizou rapidamente continha ambos os ‘inibidores’ ou ambas as cadeias de CTI, o que foi confirmado por SDS-PAGE (figura 11-16).

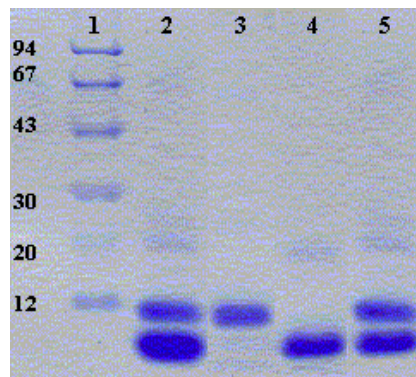


Figura 11-16: SDS-PAGE de CTI. Linha 1: marcadores de massa molecular fosforilase (94kDa), albumina (67kDa), ovoalbumina (43kDa), anidrase carbônica (30kDa), inibidor de tripsina do pâncreas (20kDa) e citocromo (12kDa). Linha 2: CTI reduzido com DTT 0,1M. Linhas 3 e 4: CTI reduzido com DTT 0,1M e purificado por HPLC usando coluna de fase reversa para separar os dois ‘inibidores diferentes’. Linha 5: CTI dos cristais dissolvidos.

A figura 11-15 à direita mostra uma superposição da região de quebra entre as cadeias com o loop equivalente de outros membros da família STI Kunitz de inibidores. Todos os inibidores são formados por uma cadeia polipeptídica única e, como esperado são contínuos nessa região mas, curiosamente, muitos resíduos no loop correspondente à essa região em ETI estão ausentes na estrutura final devido à falta de densidade eletrônica. Essa falta de densidade em ETI pode ser explicada por se tratar de uma região de loop mais flexível com uma certa desordem estrutural natural.

CTI é o primeiro inibidor do tipo STI Kunitz cuja estrutura consiste de duas cadeias polipeptídicas não covalentemente ligadas descrito.

12. CTI APRESENTA UMA ÚNICA PONTE DISSULFETO

Uma das características que distingue a família STI Kunitz é a presença de duas pontes dissulfeto, que são conservadas em quase todos os membros da família. Em contraste, CTI contém apenas uma ponte dissulfeto entre os resíduos A40 e A84. A existência de uma única ponte em CTI foi confirmada pelo mapa de diferença anômala calculado para a região onde as pontes estão presentes, ou deveriam estar presentes, comparando-se com os demais membros da família. Essas regiões são mostradas na figura 11-17.

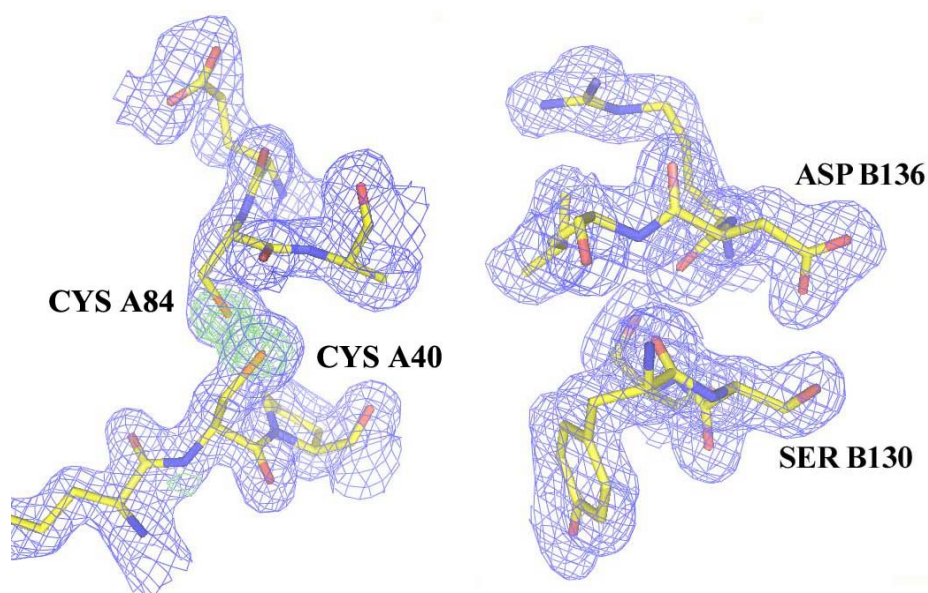


Figura 11-17: Porções do mapa (2Fo-Fc) e do mapa de diferença anômala superpostos com as regiões de CTI onde as pontes dissulfeto se encontram quando comparado aos demais membros da família STI Kunitz.

Apenas a primeira ponte, mostrada à esquerda, foi confirmada.

Uma busca por seqüências homólogas foi feita usando a seqüência de CTI no BLASTp (<http://www.ncbi.nlm.nih.gov>) sendo encontradas seqüências de vários inibidores com alta homologia. Um alinhamento entre eles revela que existe apenas um único membro na família STI Kunitz, WBA (MCCOY & KORTT, 1997), que não possui a primeira ponte dissulfeto presente em todas as outras seqüências comparadas. A ponte dissulfeto nessa posição une as subunidades T₂ e T₃ e está localizada entre o fim da última fita-β da subunidade T₂ e o começo da primeira fita-β da subunidade T₃. Essa ponte dissulfeto fixa o loop reativo dos inibidores, contribuindo para a manutenção da conformação canônica necessária para a atividade inibitória. WBA, o único membro da família que não possui essa ponte dissulfeto, não possui qualquer atividade inibitória sendo classificado como uma albumina. A alta conservação dessa ponte entre todos os membros da família STI Kunitz indica a importância do seu papel estrutural.

Entretanto a segunda ponte dissulfeto, que está ausente em CTI, apresenta um aspecto interessante pois, embora ela seja conservada na maioria dos inibidores comparados, ela está ausente em vários deles e existem duas pontes dissulfeto na região correspondente de alguns inibidores que apresentam três pontes ao invés de duas. Isso é uma indicação de que a ponte nessa posição não é de fundamental importância para a estabilidade estrutural ou atividade desses inibidores. Além disso, existem estudos que demonstram que STI retém a sua atividade inibitória após a redução da segunda ponte dissulfeto mas fica inativo no estado completamente reduzido (LEHLE *et al.*, 1994). Isso reforça a idéia de que a primeira ponte é essencial para a atividade dos inibidores embora a segunda não seja.

13. O LOOP REATIVO DE CTI

O loop com o sítio reativo de CTI está localizado na cadeia A e é formado pelos resíduos (Phe59–Lys60–Ala61–Ser62–Pro63–Arg64–Ser65–Lys66–Tyr67–Ile68–Ser69) desde a posição P6 até a P5' com o resíduo Arg64 ocupando a posição P1.

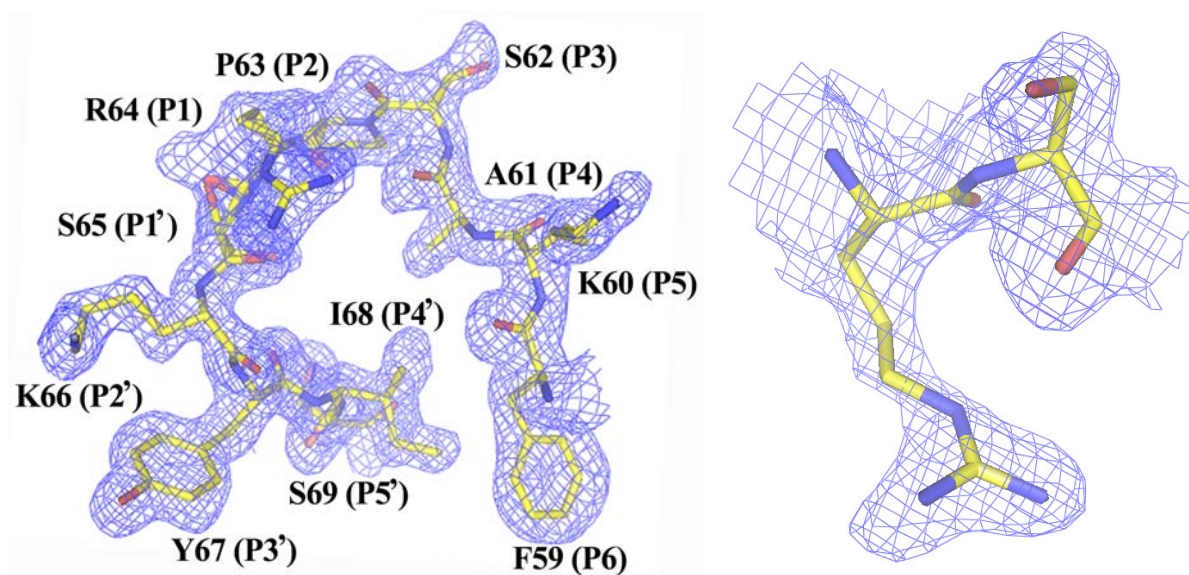


Figura 11-18: à esquerda está o mapa (2Fo-Fc) de densidade eletrônica mostrando o loop reativo de CTI e à direita é mostrado um zoom da ligação scissile entre P1-P1'.

Quando é feito um alinhamento entre os loops reativos dos membros da família Kunitz, verifica-se que assim como STI (SONG & SUH, 1998), ETI (ONESTI *ET AL.*, 1990) e WCI (RAVICHANDRAN *et al.*, 1999), CTI apresenta a conformação canônica para o loop reativo e todos os resíduos são perfeitamente superpostos. WBA (MCCOY & KORTT, 1997) apresenta uma inserção de 4 resíduos no loop reativo entre as posições P2 e P1 que torna o loop mais protuberante e como resultado WBA não mostra qualquer atividade inibitória. BASI (VALLEE *et al.*, 1998) tem a inserção de 3 resíduos, um entre P4 e P3, um entre P3 e P2 e um entre P1 e P1' e, conseqüentemente, o seu loop reativo também diverge dos outros membros da família. Embora estruturalmente BASI pertença à família STI Kunitz, o seu loop reativo não possui a conformação canônica característica encontrada para os inibidores e a sua atividade inibitória é bem diferente, uma vez que BASI pode inibir simultaneamente α -amilase e serinoproteinases da família da subtilisina (VALLEE *et al.*, 1998).

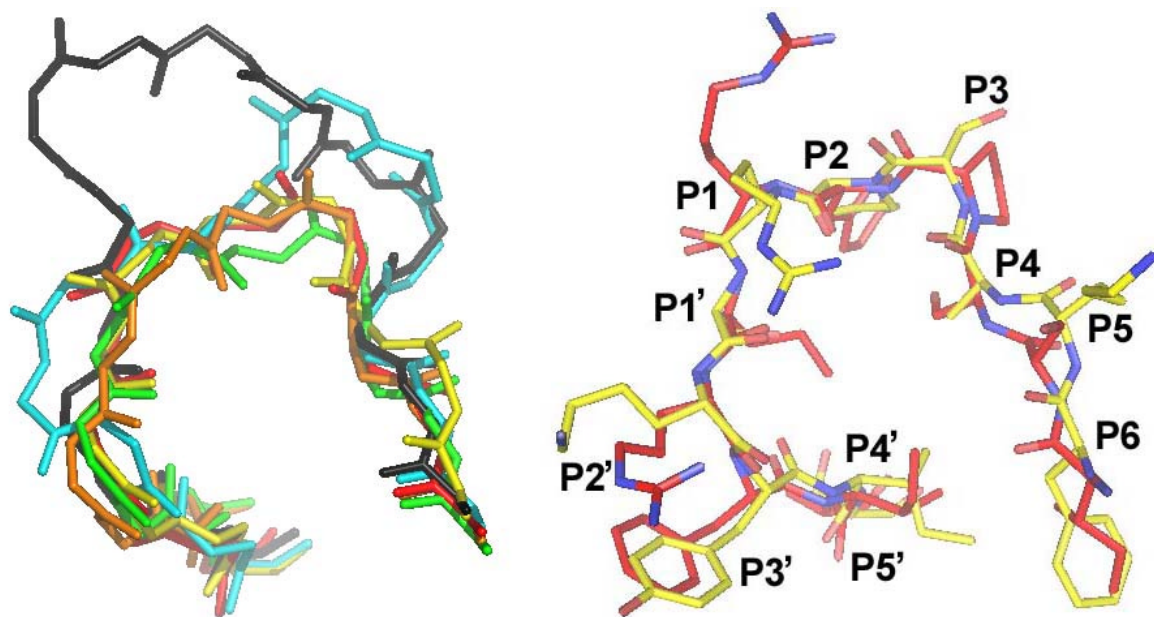


Figura 11-19: Comparação estrutural entre os loops reativos dos membros da família STI Kunitz à esquerda mostrando as cadeias principais de CTI (amarelo), STI (vermelho), ETI (laranja), WCI (verde), WBA (preto) e BASI (cyan). Na mesma orientação à direita são mostradas também as cadeias laterais de CTI (amarelo) e STI (vermelho).

14. COMPARAÇÃO DE CTI COM OUTROS MEMBROS DA FAMÍLIA KUNITZ

Como era de se esperar, CTI mostra o mesmo arranjo encontrado nos outros membros da família STI Kunitz, com as maiores concordâncias nas regiões das fitas- β s e as maiores discrepâncias nas regiões de loop, como se pode ver na figura 11-20. A tabela 11-7 mostra algumas informações a respeito desses membros alinhados, como a porcentagem de identidade sequencial entre eles e os respectivos desvios quadráticos médios (rms) obtidos dos alinhamentos feitos no programa O e o número de carbonos- α s topologicamente equivalentes que foram superpostos, também mostra a posição das pontes dissulfeto em cada um deles.

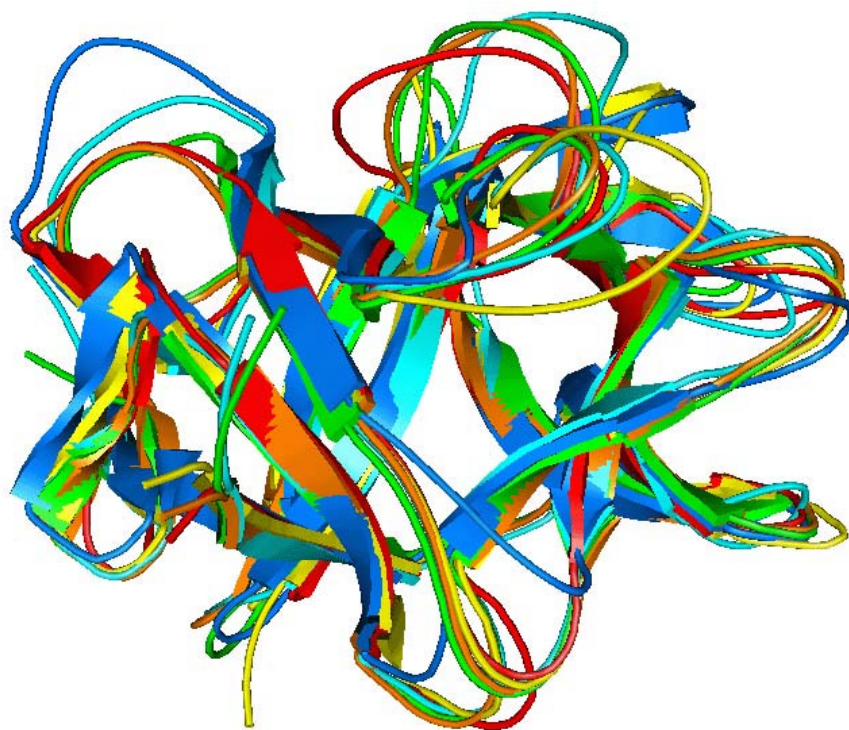


Figura 11-20: Alinhamento estrutural entre os membros da família STI Kunitz: CTI (vermelho), STI (amarelo), ETI (laranja), WCI (verde), WBA (azul) e BASI (cyan).

Tabela 11-7: Informações do alinhamento entre os membros da família STI Kunitz.

Proteína	Identidade seqüencial (%)	No. total de resíduos	No. de Cα superpostos	r.m.s. (Å)	Localização das pontes dissulfeto
CTI		163			A40–A84
STI	21	181	147	1.15	39–86 136–145
ETI	31	172	153	1.57	36–83 132–139
WCI	23	183	156	1.51	41–85 135–144
WBA	19	175	137	1.69	135–141
BASI	19	181	145	1.42	C43–C90 C144–C148
DrTI	25	185	146	1,57	44–89 139–147

15. DISCUSSÃO DA RELAÇÃO ENTRE A ESTRUTURA E A FUNÇÃO DE CTI

Similarmente aos inibidores da família STI Kunitz que adotam a conformação canônica para o loop reativo, CTI tem uma significativa atividade inibitória contra tripsina ($K_i=1,2\text{nM}$). Para compreender a base estrutural para esse fato, a estrutura de CTI foi comparada com as estruturas de STI livre (código PDB 1AVU) e STI complexado com tripsina porcina (código PDB 1AVW). O primeiro aspecto que chama a atenção é que nos inibidores que não estão complexados, o loop reativo não apresenta uma estrutura secundária muito bem definida, veja a seta na figura 11-21. Já no caso do complexo STI-Tripsina, as interações entre os sítios ativos da tripsina e do inibidor resultam numa mudança drástica na conformação formando uma fita- β no loop reativo de STI que parece com outra fita- β da tripsina formando uma folha- β . A formação da folha- β indica um ganho na estabilidade estrutural que implica num ganho energético, mostrando que a formação do complexo é um evento favorável.

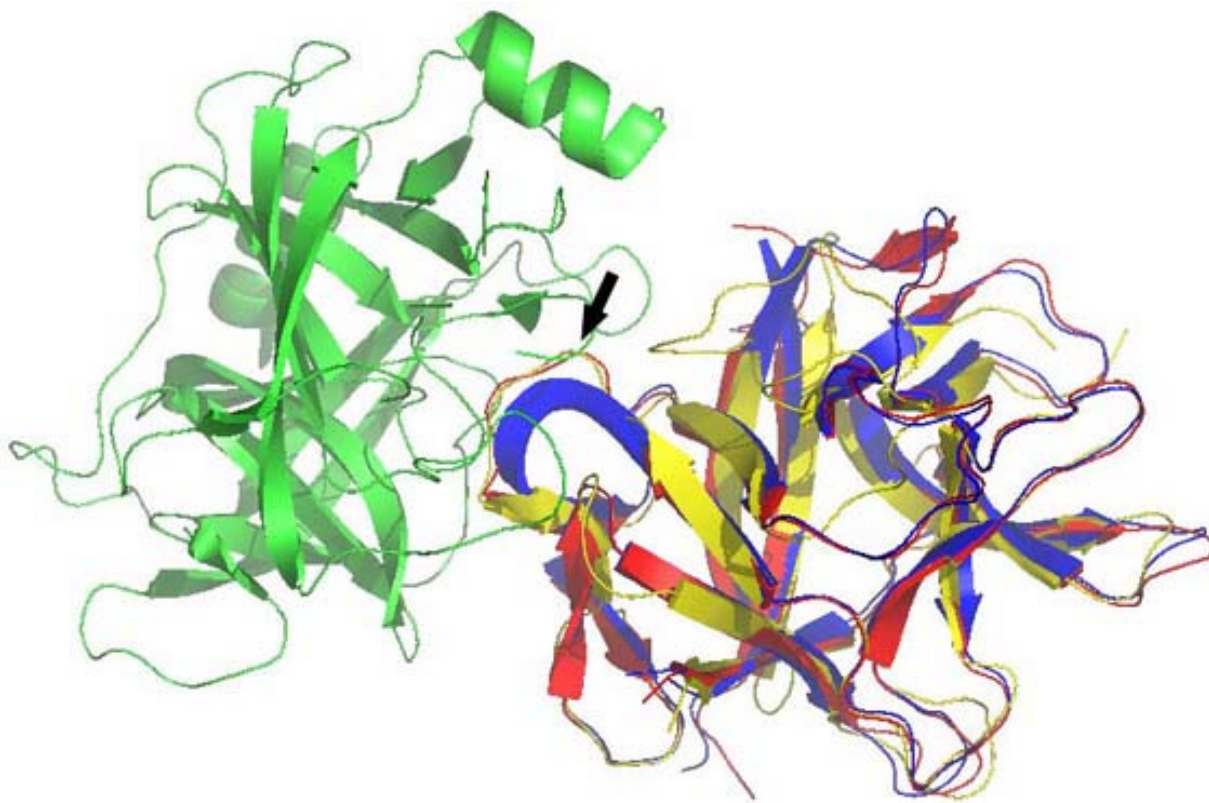


Figura 11-21: Superposição de CTI (amarelo) com STI livre (azul) e STI complexado (azul) com tripsina (verde). A seta indica o loop reativo dos inibidores, que forma uma fita- β na formação do complexo com tripsina.

As interações de STI com tripsina quando o complexo é formado não se restringem apenas aos sítios reativos. É formada uma interface de contatos entre inibidor-enzima, que envolve várias partes de ambas as moléculas. A superposição da estrutura de CTI sobre a estrutura de STI complexado mostra que a maioria dos contatos inibidor-tripsina seria preservada no suposto complexo entre CTI-Tripsina. A figura 11-22 mostra cada um dos resíduos de STI que fazem contato com tripsina superpostos com os respectivos resíduos de CTI mostrando os supostos contatos que esses fariam com a tripsina. É evidente que na formação do complexo CTI-Tripsina, alguns dos resíduos do inibidor, particularmente a ArgA64 na posição P1, sofrerão mudanças conformacionais para que haja uma melhor acomodação das interações. Por isso, são observados alguns contatos estericamente impedidos por serem muito próximos e outros em que os resíduos estão muito distantes para que haja uma interação efetiva. Seria necessário analisar a estrutura real do complexo CTI-Tripsina para se ter certeza definitiva da superfície de contatos entre eles mas as suposições inferidas aqui não devem ser muito diferentes dos contatos reais.

As razões estruturais para a alta atividade inibitória de CTI contra tripsina residem no fato de que a conformação canônica do loop reativo é mantida, a despeito da presença de duas cadeias polipeptídicas não covalentemente ligadas e da falta de uma das pontes dissulfeto características da família STI Kunitz. Apesar das diferenças estruturais, a densa rede de pontes de hidrogênio observada em CTI mantém a integridade e a estabilidade estrutural da molécula.

O complexo CTI-Tripsina foi preparado e, apesar das inúmeras tentativas de cristalização desse complexo, nenhum resultado promissor foi alcançado até o momento.

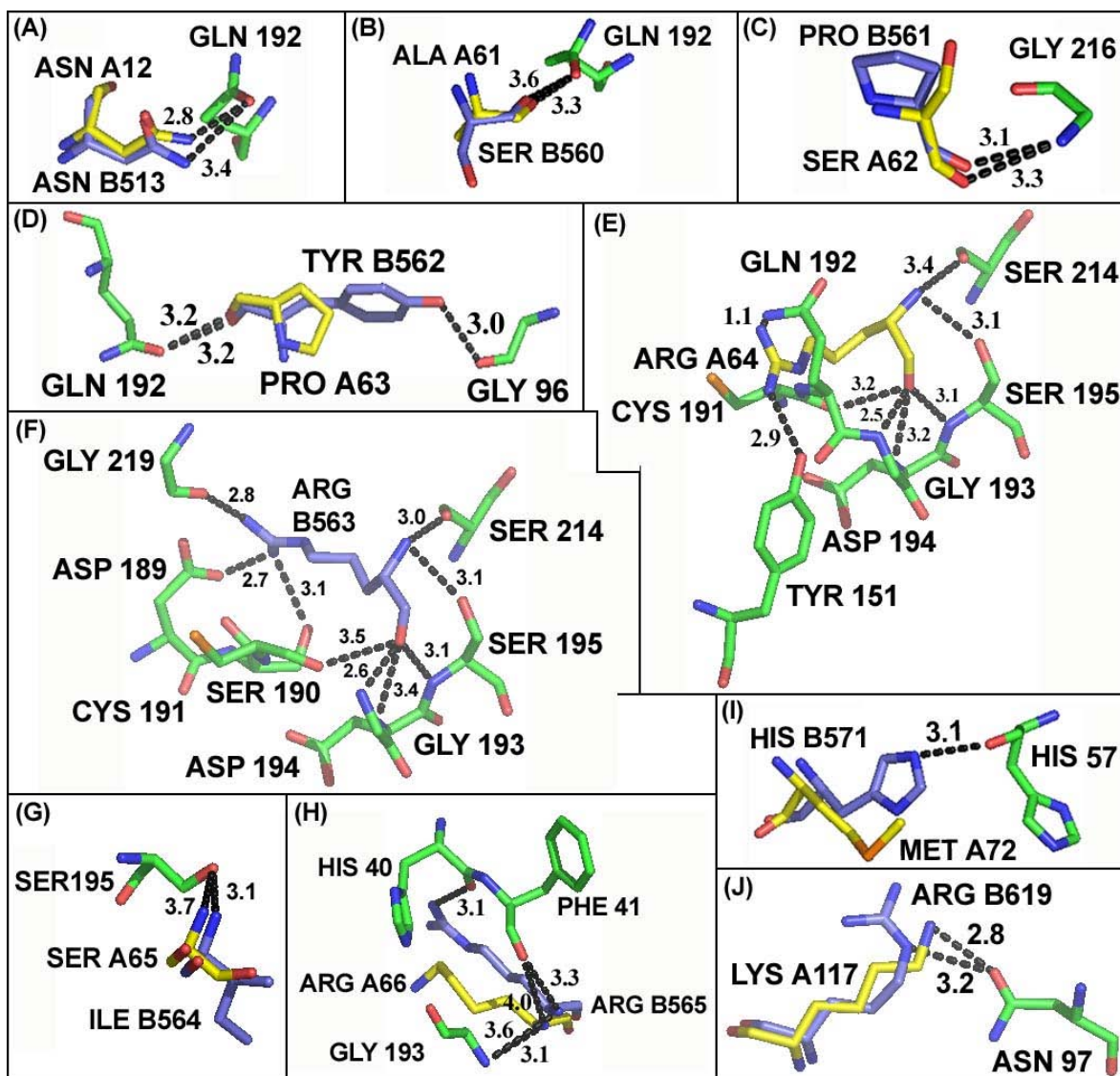


Figura 11-22: Os resíduos de STI (azul) que fazem contato com tripsina (verde) no complexo STITripsina superpostos com CTI (amarelo) mostrando os supostos contatos que esse faria com a tripsina. As distâncias são mostradas entre parênteses em Å. Esses contatos são detalhados na tabela 11-8. (A) (STI)asn13 e (CTI)asn12 com (trip)gln192; (B) (STI)ser60 e (CTI)ala61 com (trip)gln192; (C) (STI)pro61 e (CTI)ser62 com (trip)gly216; (D) (STI)tyr62 e (CTI)pro63 com (trip)gln192/gly96; (E) (CTI)arg64 com (trip)tyr151/cys191/gln192/gly193/asp194/ser195/ser214; (F) (STI)arg63 com (trip)asp189/ser190/gly193/asp194/ser195/ser214/gly219; (G) (STI)ile64 e (CTI)ser65 com (trip)ser195; (H) (STI)arg65 e (CTI)lys66 com (trip)his40/phe41; (I) (STI)his71 e (CTI)met72 com (trip)his57 e (J) (STI)arg119 e (CTI)lys117 com (trip)asn97.

Tabela 11-8: Lista de contatos feitos entre o inibidor e a enzima no complexo STITripsina e no complexo hipotético entre CTITripsina. Um script do CNS foi usado para analisar os contatos e a distância limite de 3,5Å foi considerada. Veja a correlação com a figura 11-22.

Átomos em STI	Distância (Å)	Átomos em	Distância (Å)	Átomos em CTI
		Tripsina		
ARG B565 NH1	3.1	HIS 40 O		
ARG B565 N	3.3	PHE 41 O	4.0	LYS A66 N
HIS B571 NE2	3.1	HIS 57 O		
TYR B562 OH	3.0	GLY 96 O		
ARG B619 NE	3.2	ASN 97 OD1	2.8	LYS A117 NZ
		TYR 151 OH	2.9	ARG A64 NH2
ARG B563 NH2	2.7	ASP 189 OD1		
ARG B563 NH2	3.1	SER 190 OG		
ARG B563 O	3.5	CYS 191 O	3.2	ARG A64 O
ASN B513 ND2	3.4	GLN 192 OE1	2.8	ASN A12 ND2
SER B560 O	3.6	GLN 192 OE1	3.3	ALA A61 O
TYR B562 O	3.2	GLN 192 OE1	3.2	PRO A63 O
		GLN 192 NE2	1.1	ARG A64 NH1
ARG B563 O	2.6	GLY 193 N	2.5	ARG A64 O
ARG B565 N	3.6	GLY 193 N	3.1	LYS A66 N
ARG B563 O	3.4	ASP 194 N	3.2	ARG A64 O
ARG B563 N	3.1	SER 195 OG	3.1	ARG A64 N
ILE B564 N	3.1	SER 195 OG	3.7	SER A65 N
		SER 195 N	3.1	ARG A64 O
ARG B563 N	3.0	SER 214 O	3.4	ARG A64 N
PRO B561 O	3.1	GLY 216 N	3.3	SER A62 O
ARG B563 NH1	2.8	GLY 219 O		

CAPÍTULO 12

DrTI: O INIBIDOR EXTRAÍDO DAS SEMENTES DE *Delonix regia*

1. INTRODUÇÃO

Aqui é discutida a solução da estrutura cristalográfica de outro novo inibidor também pertencente à família STI-Kunitz. Trata-se do inibidor extraído das sementes de *Delonix regia* (DrTI), o qual possui massa molecular de 22kDa e duas pontes dissulfeto. Os estudos bioquímicos prévios estabeleceram que DrTI é um inibidor efetivo de tripsina e de calicreína do plasma humano (HPK do inglês *Human Plasma Kallikrein*) com valores de K_i de 21,9nM e 5,3nM, respectivamente, mas não é ativo contra quimotripsina ou calicreína de tecido (PANDO *et al.*, 2001). Como veremos, alguns aspectos particulares da estrutura de DrTI mostram as bases moleculares para esse comportamento.

A *Delonix regia*, conhecida comumente como flamboyant, flor do paraíso ou pau rosa, é uma árvore de médio a grande porte de copa bem ampla, que pertence à família Leguminosae-Caesalpinoideae. Originária de Madagascar, pode ser encontrada por todo o Brasil, principalmente nas cidades em parques e praças, pois é bastante apreciada para ornamentação devido às suas belas flores vermelhas – daí o nome flamboyant do francês *flamejante*. Os frutos são vagens grandes, com aproximadamente 50cm, que demoram em torno de um ano para atingir a maturação. (LORENZI, 1992)



Figura 12-1: Árvore de *Delonix regia*.

2. CRISTALIZAÇÃO DE DrTI

As sementes de *Delonix regia* foram coletadas e o inibidor foi extraído e purificado por Silvana Cristina Pando, sob orientação do prof. Dr. Sérgio Marangoni, no Laboratório de Química de Proteínas do Departamento de Bioquímica do Instituto de Biologia da UNICAMP (PANDO *et al.*, 2001). Quantidades suficientes para cristalização nos foram cedidas.

O inibidor liofilizado foi dissolvido em tampão fosfato de sódio 25mM pH 8,0 para uma concentração de 10mg/mL. Os cristais adequados para a coleta dos dados de difração foram crescidos pelo método de difusão de vapor com a gota pendurada a 291K numa solução de fosfato de sódio 100mM pH 8,5 usando 250mM de sulfato de amônio e 25% de PEG8000 como precipitantes (POLIKARPOV *et al.*, 1999). As gotas foram formadas com volumes iguais (10 μ L) de solução de proteína e solução do poço. Cristais com dimensões aproximadas de (0,10 $\times$ 0,10 $\times$ 0,30) mm³ cresceram após cerca de três semanas. Para a coleta dos dados de difração, um cristal foi pescado da gota usando um loop de rayon e mergulhado por 30 segundos numa solução crio-protetora contendo 25% de etileno glicol misturado com solução do poço. O cristal foi então repescado sendo rapidamente congelado num feixe de nitrogênio a 100K.

3. COLETA E PROCESSAMENTO DOS DADOS DE DIFRAÇÃO PARA DrTI

Um conjunto de dados de raios-x foi coletado na linha de Cristalografia de Proteínas (POLIKARPOV *et al.*, 1997) no Laboratório Nacional de Luz Síncrotron (Campinas – SP – Brasil), usando uma placa de imagem 345mm MAR Research como detector e o comprimento de onda da radiação síncrotron foi ajustado para 1,38Å. Imagens com intervalo angular de 1° foram coletadas com tempos de exposição variando entre 180 e 480 segundos/imagem. O cristal de DrTI coletado difratou até 1,75Å de resolução.

Os dados foram auto-indexados e integrados com o programa DENZO (OTWINOWSKI & MINOR, 1993). O escalonamento e fusão dos dados foram feitos com o programa SCALEPACK (OTWINOWSKI & MINOR, 1993). As estatísticas da coleta e processamento dos dados são resumidas na tabela 12-1.

Tabela 12-1: resumo das informações estatísticas do processamento dos dados coletados. Valores estatísticos para a camada de maior resolução estão mostrados entre parênteses.

Grupo espacial	P2 ₁ 2 ₁ 2 ₁
Distância cristal ao detector (mm)	130
Dimensões da cela unitária (Å)	a = 32,13, b = 67,25, c = 72,06
Resolução (Å)	30.0 - 1.75 (1,89-1,75)
No. de reflexões	75603
No. de reflexões únicas	16616
<I/σ(I)>	9,0 (2,4)
Completeza (%)	82 (98.0)
R _{merge} * (%)	6,1 (29,2)
Mosaicidade (°)	0,6
Intervalo de dados coletados	194°
Redundância	4,6 (3,2)
Solução crio-protetora	Solução mãe com 25 % etileno glicol

$$* R_{\text{merge}} = \frac{\sum_{\text{hkl}} |I - \langle I \rangle|}{\sum_{\text{hkl}} I}$$

4. APLICAÇÃO DA SUBSTITUIÇÃO MOLECULAR A DrTI

Então os dados de difração foram utilizados na resolução da estrutura tridimensional do inibidor DrTI por substituição molecular. Primeiramente usando-se a estrutura do inibidor extraído de *Erythrina caffra* (ONESTI *et al.*, 1990; código 1TIE no PDB) o qual apresenta a maior identidade (35%) com DrTI. Mas, apesar da solução obtida ser clara, o modelo inicial assim obtido, apresentou uma grande dificuldade no refinamento, possivelmente porque o modelo de *Erythrina* não possui alguns loops, que por apresentarem falta de densidade eletrônica, foram omitidos da estrutura final. Justamente essas regiões é que apresentaram os maiores problemas durante o refinamento.

Por isso, decidiu-se aplicar novamente o método da substituição molecular empregando-se o modelo do inibidor extraído das sementes de *Copaífera langsdorffii* (CTI) cuja estrutura foi previamente determinada por substituição isomorfa e possui todos os loops muito bem definidos com exceção do loop referente à quebra entre as cadeias.

A estrutura de CTI foi usada como modelo de busca no programa AMORE (NAVAZA, 1994) do pacote CCP4. As soluções obtidas estão resumidas na tabela 12-2.

A solução correta é aquela que apresenta o maior coeficiente de correlação. Além disso, essa solução também deve apresentar um valor para o coeficiente de correlação distante das demais. Pelos dados percebe-se que a solução 1 atende esses requisitos e, por isso, foi refinada e aplicada ao modelo de CTI para a obtenção de um modelo inicial para DrTI.

Tabela 12-2: soluções obtidas para a substituição molecular usando o programa AMORE do pacote CCP4 ordenadas por coeficiente de correlação. São mostradas as soluções das funções de rotação, translação, os coeficientes de correlação, o R_{factor} e também a contagem de picos no mapa de Patterson.

SOLUÇÃO	α (°)	β (°)	γ (°)	Tx (Å)	Ty (Å)	Tz (Å)	Corr_F (%)	R_{factor} (%)	Corr_I (%)	Pkcount	Dmin
1	111.72	80.03	170.32	0.0142	0.0343	0.0217	53.9	53.4	45.1	1	33.3
2	111.72	80.03	170.32	0.0111	0.0351	0.2833	50.8	54.5	40.7	5	34.0
3	111.72	80.03	170.32	0.0166	0.2836	0.0216	50.2	54.9	39.6	2	17.0
4	111.72	80.03	170.32	0.0120	0.0358	0.1320	49.9	54.7	39.3	10	37.7
5	111.72	80.03	170.32	0.1941	0.2770	0.3556	49.8	54.9	37.0	7	26.5
6	111.72	80.03	170.32	0.1822	0.0323	0.0234	49.5	55.2	38.5	4	33.6
7	111.72	80.03	170.32	0.1070	0.0347	0.0215	49.4	55.6	38.6	8	33.3
8	111.72	80.03	170.32	0.0123	0.2424	0.0199	49.1	55.3	36.8	6	16.3
9	111.72	80.03	170.32	0.1827	0.2815	0.0234	49.0	55.5	37.8	15	17.0
10	111.72	80.03	170.32	0.4331	0.1520	0.3053	49.0	55.2	36.5	9	34.8
11	111.72	80.03	170.32	0.4409	0.1033	0.3040	48.3	55.6	36.1	11	34.7
12	111.72	80.03	170.32	0.1801	0.2798	0.2365	48.1	55.6	36.8	12	35.6
13	111.72	80.03	170.32	0.3971	0.2578	0.4715	47.9	55.9	34.4	13	16.6
14	111.72	80.03	170.32	0.4367	0.4816	0.3027	47.6	55.6	35.3	14	34.7
15	111.72	80.03	170.32	0.3393	0.2177	0.0718	47.1	56.4	35.3	3	19.6

5. REFINAMENTO DO MODELO DE DrTI

O refinamento do modelo inicial de DrTI foi feito usando o programa REFMAC5 do pacote de programas CCP4 (1994). Inicialmente o modelo foi submetido ao refinamento de corpo rígido. Foram calculados os mapas de diferença de densidade eletrônica e, através de análise visual com o programa O, a sequência de aminoácidos foi mudada de acordo com aquela determinada para DrTI.

Os valores de R_{factor} e R_{free} no começo do refinamento foram 30% e 39%, respectivamente. Foram rodados muitos ciclos de refinamento posicional e restrito com B-fatores isotrópicos, usando o programa REFMAC5, intercalados com dinâmica molecular usando o script padrão do CNS (BRÜNGER *et al.*, 1998). Esses ciclos de refinamento automático foram alternados com intervenção manual no programa O, usando os mapas ($2m|F_{obs}| - D|F_{calc}|$, ϕ_{calc}) e ($m|F_{obs}| - D|F_{calc}|$, ϕ_{calc}) para analisar o e corrigir o modelo.

Tabela 12-3: Resumo das estatísticas do refinamento de DrTI.

No. de átomos de proteína	1417
No. de moléculas de água	183
R _{factor} (%)	21.5
*R _{free} (%)	25.3
Média dos B-fatores	
Cadeia principal (Å ²)	32.5
Cadeias laterais (Å ²)	34.8
Moléculas de água (Å ²)	45.4
Todos os átomos	34.9
Desvios quadráticos da geometria ideal	
Comprimento das ligações (Å ²)	0.0050
Ângulos das ligações (graus)	1.2
Estatísticas do diagrama de Ramachandran	
Resíduos em regiões mais favoráveis	134 (88.2%)
Resíduos em regiões adicionalmente favoráveis	13 (8.6%)
Resíduos em regiões generosamente favoráveis	5 (3.3%)
Resíduos em regiões desfavoráveis	0 (0%)

* O R_{free} foi calculado com um subconjunto de 5% de reflexões aleatoriamente selecionadas.

O progresso do refinamento foi monitorado seguindo os parâmetros estatísticos: figura-de-mérito, R_{factor} e R_{free}. As moléculas de água foram sendo gradualmente localizadas durante o refinamento. Elas foram adicionadas de acordo com o critério de que uma molécula de água deve fazer pelo menos uma ponte de hidrogênio estereoquimicamente razoável, deve ter densidade bem definida acima de 1σ no mapa (2m|F_{obs}|-D|F_{calc}|) e deve ter densidade acima de 3σ no mapa (m|F_{obs}|-D|F_{calc}|).

Como a densidade observada para o loop que contém o sítio ativo não estava em concordância com a sequência determinada para essa região, particularmente para o resíduo A69 o qual foi inicialmente determinado como uma serina, DrTI foi re-sequenciado. Os resíduos em discordância foram substituídos de acordo com a nova sequência e, na verdade, o resíduo A69 é uma lisina (ver figura 12-3 com o loop reativo).

Um resumo das estatísticas do refinamento é dado na tabela 12-3.

6. QUALIDADE DO MODELO DETERMINADO PARA DrTI

O modelo final é composto de 185 resíduos de aminoácidos e 183 moléculas de água bem ordenadas. Como se pode ver pela tabela 12-3, o modelo de DrTI é razoavelmente bem determinado. A densidade eletrônica é clara para a maioria dos resíduos a $2.5-3\sigma$. Entretanto, a densidade eletrônica para os resíduos 180-185 na porção C-terminal de DrTI, é um pouco fraca e fragmentada.

Os desvios dos valores ideais para os comprimentos e ângulos das ligações são pequenos e não existem resíduos em regiões desfavoráveis do diagrama de Ramachandran. Cerca de 70% dos resíduos estão em regiões enterradas da molécula formando as folhas- β s de DrTI e os outros 30% estão em regiões acessíveis formando os loops. As regiões de loop, bem como as porções N- e C-terminais apresentam valores para os B-fatores mais altos do que as partes mais bem ordenadas da molécula.

7. ARQUITETURA GLOBAL DO MODELO DETERMINADO PARA DrTI

Como era de se esperar, DrTI apresenta o mesmo arranjo já descrito para CTI. Sua estrutura consiste de 12 fitas- β s antiparalelas conectadas por longos loops divididas em 6 β -harpins com duas fitas- β s cada. Três dos β -hairpins formam um barril- β enquanto os três restantes formam uma tampa triangular em cima desse barril, formando o conhecido arranjo tipo β -trefoil.

Para DrTI as fitas- β s e os loops foram nomeados seguindo-se a nomenclatura padrão de MCLACHLAN, sendo marcadas com A, B e C de acordo com a qual das três subunidades elas pertencem e com um número de 1 a 4 indicando a sequência da cadeia a seguir. Assim o arranjo estrutural começa na subunidade de A1 a A4, segue pela subunidade de B1 a B4 e termina na subunidade de C1 a C4 como mostrado na figura 12-2 à esquerda.

A superposição estrutural das subunidades, mostrada na figura 12-2 à direita, resultou num desvio quadrático médio de $1,77\text{\AA}$ entre as subunidades A e B, $1,53\text{\AA}$ entre A e C e $1,59\text{\AA}$ entre B e C.

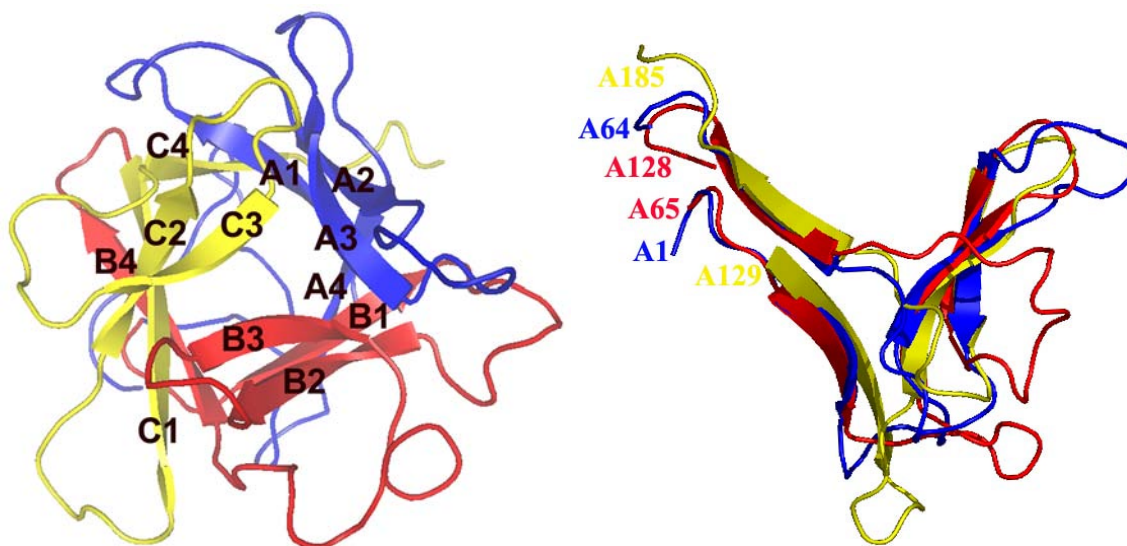


Figura 12-2: À esquerda é mostrado a arquitetura global da molécula de DrTI com as três subunidades estruturais A, B e C mostradas em azul, vermelho e amarelo respectivamente. À direita é mostrado um alinhamento estrutural dessas subunidades.

8. OS ELEMENTOS DA ESTRUTURA SECUNDÁRIA DE DrTI

Os elementos da estrutura secundária de DrTI foram calculados usando o programa PROMOTIF (HUTCHINSON & THORNTON, 1996) e estão descritos na tabela 12-4. DrTI tem duas pontes dissulfeto, entre os resíduos 44 e 89 e entre os resíduos 139 e 147, muito bem definidas no mapa de densidade eletrônica. A primeira ponte dissulfeto liga o loop que conecta as fitas- β s A2 e A3 com o loop que conecta as fitas- β s B1 e B2, e a segunda ponte une o fim da fita- β C1 e o começo da fita- β C2.

As doze fitas- β s da estrutura de DrTI são divididas em 4 folhas- β s antiparalelas. As numerosas pontes de hidrogênio que formam e estabilizam as folhas- β s em conjunto com as duas pontes dissulfeto são responsáveis pela rigidez e estabilidade da estrutura de DrTI.

Tabela 12-4: Estrutura secundária de DrTI.

Fitas- β s		Loops		β -turns	
				Localização	Tipo
		N-term	1–18	8–11	I
				15–18	II
A1	19–24				
		A1–A2	25–33	25–28	I
A2	34–37				
		A2–A3	38–46	39–42	II
A3	47–50				
		A3–A4	51–63	53–56	I
				54–57	VIII
A4	61–64				
		A4–B1	65–79	65–68	I
				74–77	IV
B1	80–83				
		B1–B2	84–94	83–86	VIII
				87–90	I
B2	95–98				
		B2–B3	99–104	99–102	IV
B3	105–108				
		B3–B4	109–121	110–113	I
				111–114	I
B4	122–128				
		B4–C1	129–133	130–133	IV
C1	134–140				
		C1–C2	141–147	140–143	I
				141–144	IV
				143–146	IV
C2	148–153				
		C2–C3	152–158	154–157	IV
				155–158	IV
C3	159–162				
		C3–C4	163–171	166–169	IV
C4	172–175				
		C-term	176–180	179–182	IV

9. O LOOP REATIVO DE DrTI

A comparação estrutural de DrTI com outros inibidores da família STI Kunitz define o seu loop reativo desde o resíduo Ser65(P4) até o resíduo Ile73(P4'). A figura 12-3 no topo mostra o mapa de densidade eletrônica (2Fo-Fc) final para essa região de DrTI.

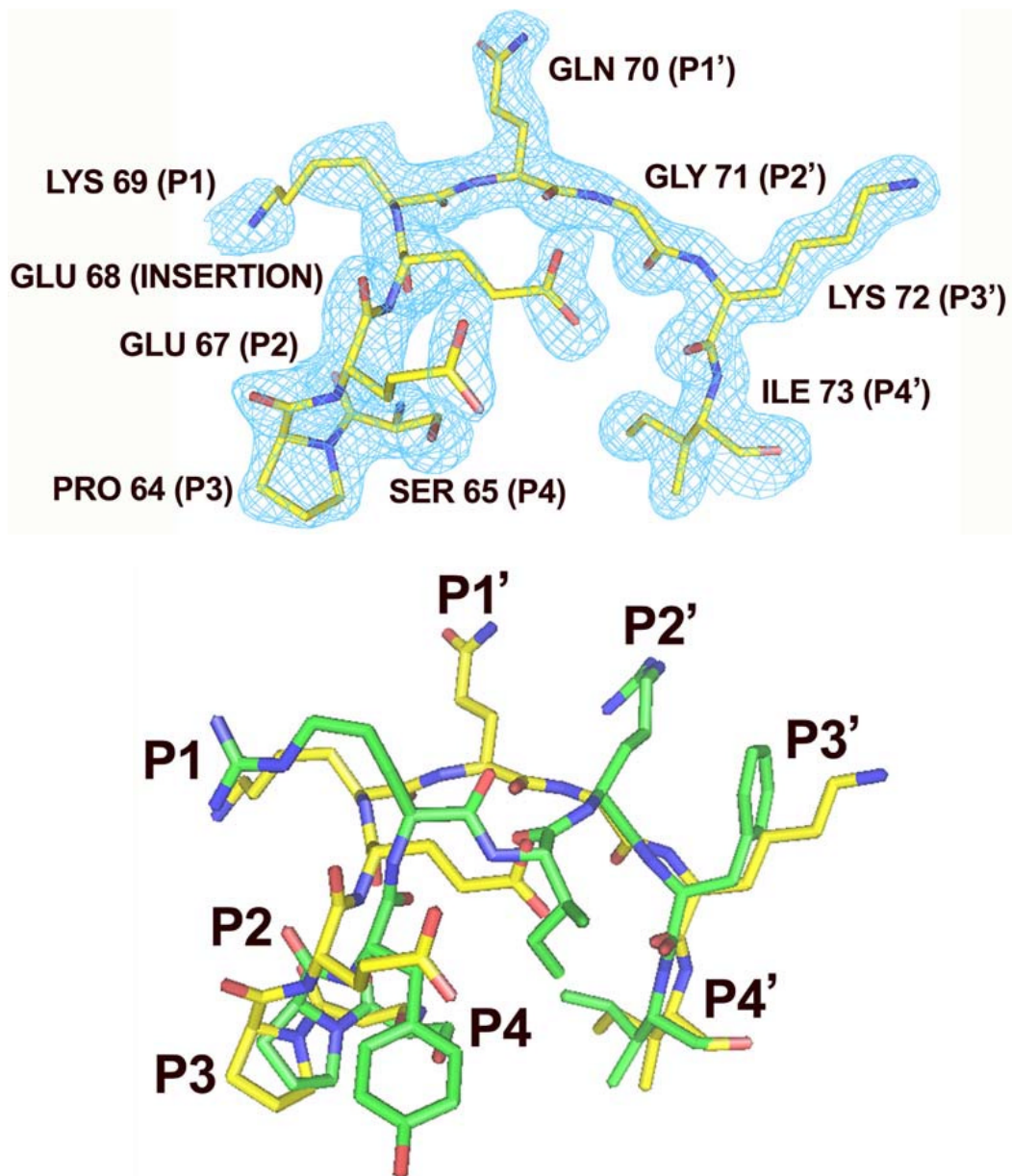


Figura 12-3: Densidade eletrônica para a região do loop reativo de DrTI mostrada no topo e a superposição com a mesma região de STI em verde embaixo. Os resíduos foram numerados de acordo com STI.

A superposição entre o loop reativo de DrTI com a mesma região de STI (SONG & SUH, 1998) (figura 12-3 embaixo) mostra que, embora DrTI tenha um resíduo positivo,

Lys69, na posição P1, ele apresenta a inserção de um resíduo, Glu68, entre os resíduos P1 e P2 (seguindo a numeração de STI) o que faz com que o seu loop seja ligeiramente distorcido da conformação canônica. Essa deformação pode explicar a atividade inibitória reduzida de DrTI contra tripsina o que é expresso pelo valor de K_i ($=21,9\text{nM}$) que é ligeiramente maior do que o valor de K_i ($\sim 1\text{nM}$) para os inibidores canônicos como STI, ETI, WCI ou CTI.

Além disso, DrTI também apresenta dois resíduos negativos, Glu67 e Glu68, no loop ativo que podem apontar a razão da especificidade de DrTI contra caliceína do plasma humano (HPK) com um $K_i=5\text{nM}$.

10. INTERAÇÕES DE DrTI COM TRIPSINA E COM HPK

As caliceínas são uma subfamília de serinoproteínases com um alto grau de especificidade e expressão diversa em vários tecidos e fluidos biológicos, as quais são divididas em dois grupos: as caliceínas de plasma (EC 3.4.21.34) e de tecido (EC 3.4.21.35), significativamente diferentes em massa molecular, especificidade, características imunológicas, estrutura gênica e tipo de kinina liberada. A caliceína do plasma é expressa apenas no fígado e está envolvida na coagulação sanguínea, fibrinólise, regulação da pressão sanguínea e reações inflamatórias. As caliceínas de tecido são um grande grupo de enzimas que estão envolvidas no processamento pós-translacional de polipeptídeos, como kininogênio por exemplo, e na liberação de peptídeos biologicamente ativos, kinina por exemplo (BHOOLA *et al.*, 1992; CLEMENTS, 1997). Embora tanto a sequência da caliceína do plasma humano (HPK) quanto a de caliceína de tecido (TK) sejam conhecidas, apenas a estrutura de TK foi determinada até o momento.

Na tentativa de encontrar o fundamento estrutural de porque DrTI é um inibidor de tripsina e um inibidor efetivo de HPK mas não apresenta qualquer atividade inibitória contra quimotripsina ou TK, foram estudadas as supostas interações de DrTI com tripsina e caliceína.

Para isso, primeiro a estrutura de DrTI foi superposta com a estrutura de STI complexado com tripsina (código PDB 1AVW) e os supostos contatos entre DrTI e tripsina foram analisados. Esses contatos envolvem não apenas resíduos do sítio ativo mas também de outras partes das moléculas formando uma interface entre DrTI e tripsina. A grande

maioria dos contatos é preservada, o resíduo positivo Lys69 na posição P1 faz a maioria dos contatos com a tripsina, assim como a ArgB564 em STI. Isso pode explicar porque apesar de distorcer levemente a conformação do loop reativo de DrTI, a inserção de um resíduo, Glu68, não abole completamente a sua atividade contra a tripsina diminuindo apenas um pouco a sua eficiência inibitória.

A partir desse alinhamento estrutural, que gerou o complexo hipotético entre DrTI-Tripsina, as estruturas de quimotripsina (BLEVINS & TULINSKY, 1985; código PDB 5CHA) e de TK (CHEN & BODE, 1983; código PDB 2KAI) foram superpostas em cima da tripsina. Uma diferença interessante entre TK e tripsina é a presença de inserções em vários pontos nas regiões que supostamente fariam contatos com o inibidor em TK. Por exemplo, as inserções de Ala-Asp-Gly-Lys entre os resíduos Ser97(TK) e Asp98(TK) e de uma Pro entre Thr219(TK) e Cys220(TK). Essas inserções modificam a forma molecular de TK e, por isso, irão diminuir a superfície de complementaridade entre DrTI e TK. Inserções em regiões de contato também estão presentes na estrutura de quimotripsina.

Finalmente, como a estrutura 3D de HPK ainda é desconhecida, a sua sequência (CHUNG *et al.*, 1986; código swiss-prot P03952) foi alinhada com as outras enzimas seguindo-se o alinhamento estrutural prévio. Apenas a parte da sequência de HPK que apresenta homologia com serinoproteinases (resíduos 360-646) foi usada no alinhamento (figura 12-4).

A análise desse alinhamento mostra que a maioria dos resíduos nas regiões de contato são conservados ou conservativamente substituídos entre tripsina e HPK, o que não acontece para TK. Por exemplo, o resíduo Ser37(tripsina), o qual participa em supostos contatos com Lys5 e Lys72 de DrTI, é substituído por Ala422(HPK), o que não causaria sérias perturbações na interface de contato com o inibidor, mas é substituído por Tyr37(TK), o que com certeza causaria problemas de acomodação dos contatos devido à grande cadeia lateral da tirosina. Dentre todas as substituições, existe uma que chama muita atenção: a substituição de Gln192(tripsina) por Lys583(HPK) e por Met192(TK). O resíduo Gln192 da tripsina está supostamente envolvido em contatos com os resíduos Ser65, Glu68, Lys69 e Gln70 do loop reativo de DrTI. A substituição de Gln192(tripsina) por Met192(TK) troca o grupo amida lateral por um grupo metila que é muito mais inerte, o que elimina esses contatos. Mas a substituição de Gln192(tripsina) por Lys583(HPK) pode ser rapidamente acomodada abrindo a possibilidade de novas interações eletrostáticas tanto

com o resíduo Glu68 quanto com Glu67 do loop ativo de DrTI. Isso poderia esclarecer a base estrutural para a eficiência de DrTI na inibição de HPK, embora ele seja inativo contra TK. Além disso, HPK não apresenta as inserções em regiões de contato como TK ou quimotripsina, um outro aspecto que favorece os contatos com o inibidor.

TRYPSIN	-----IVGGYTCAANSIPYQVSLN-----S--GS	39
TK	-----IIGGRECEKNSHPQVAIYH---Y--SS	39
HPK	AYGTQGSSGYSLRLCNTGDNVCTTKTSTRIVGGTSSWGEPWQVSLQVKLTA--QR	424
CHYMO	-----RIVNGEEAVPGSWPQVSLQ----DKTGF	41
		+ +
TRYPSIN	HF ⁺ CGGSLINSQWVVSAAH ⁺ CYKSR-----IQVRLGEHNIDVLEGNEQFINAAKIITHPN	93
TK	FOCGGVLVNPKNVLTAAHCKNDN-----YEVWLGRHNLFENENTAQFFGVTADFPHPG	93
HPK	HL ⁺ CGGSLIGHQVVLTAAH ⁺ CFDGLPLQDVWRIYSGILNLSDITKDTPFPSQIKEIIHQN	482
CHYMO	HF ⁺ CGGSLINENWVVTAAH ⁺ CGVTT---SDVVVAGEFDQGSSEKIQKLKIAKVFKNSK	96
	++ ***+	
TRYPSIN	FNGN---TLDNDIMLIKLS ⁺ SPATLNSRVATVSLPR--SCAAAGTECLISGWGNTKSS	147
TK	FNLSADGKDYS ⁺ DLMLRLQSPAKITDAVKVLELPT--QEPELGSTCEASGWGSIEPG	147
HPK	YKVS---EGNH ⁺ DIALIKLQAPLNYTEFQKPICLPSKGDSTSTIYTN ⁺ CWVTGWGFSKEK	536
CHYMO	YNSL---TINNDITLLKLSTAASFSQTVSAVCLPSASDDFAAGTTCVTTGWGLTRYT	152
	++ + *	+
TRYPSIN	--GSSYP ⁺ SL ⁺ LQCLKAPVLS ⁺ DSSCKSSYPG-QITGNMICVGFLEGGK ⁺ DSCQGD ⁺ SGGPVV	200
TK	PDDF ⁺ EP ⁺ DEIQCVQLTLLQNTFCADAH ⁺ PD-KVTESMLCAGYLP ⁺ GGK ⁺ DT ⁺ CMGD ⁺ SGGPLI	200
HPK	--GE-IQNILQKVNIPLVTNEECQKRYQDYK ⁺ ITQRMVCAGYKEGGK ⁺ DACKGD ⁺ SGGPLV	591
CHYMO	--NANT ⁺ PDRLQQASLPLLSNTNCKYWG ⁺ T-KIKDAMICAGAS--GVSS ⁺ CMGD ⁺ SGGPLV	203
	+	++++*
TRYPSIN	CN----G----QLQGI ⁺ VS ⁺ WGYG-CAQKNKPGVYTKVCNYVNWIIQQTIAAN-----	245
TK	CN----G--MW--QGITS ⁺ WGHT ⁺ PCGSANKPSIYTKLIFYLDWIDDTITENP-----	246
HPK	CK----HNGMWRLVGITS ⁺ WGE ⁺ G-CARREQPGVYTKVAEYMDWILEKTQSSDGKAQM ⁺ QSPA	646
CHYMO	CKKNGAW---TLVGIV ⁺ SWGSSIT ⁺ CTST-STPGVYARVTALVNWVQQTLAAN-----	252
	***+ *	

Figura 12-4: Alinhamento estrutural entre as seqüências de tripsina (código PDB 1AVW), quimotripsina (código PDB 5CHA) e caliceína de tecido (código PDB 2KAI). Usando esse primeiro alinhamento, a seqüência de caliceína do plasma humano (código swiss-prot P03952) cuja estrutura ainda não é conhecida, foi alinhada. Apenas a parte de HPK que apresenta homologia com serinoproteinasas (resíduos 360646) foi usada no alinhamento. Os resíduos que estão envolvidos em supostos contatos com DrTI estão nas caixas pretas. Os * indicam os resíduos conservados e + os que não são.

11. DISCUSSÃO DA RELAÇÃO ENTRE ESTRUTURA E FUNÇÃO DE DrTI

Existe um intrincado número de fatores que determinam a especificidade e a eficiência dos inibidores de proteinases. A composição e a conformação dos seus loop reativos são claramente importantes, mas vários outros aspectos da estrutura molecular devem ser considerados. Isso porque não é apenas o loop reativo do inibidor que interage

com a enzima, mas existem outras importantes regiões de contato entre eles. Nesse sentido, a molécula inteira deve ser considerada.

A arquitetura global, por exemplo, é essencial para a estabilidade, rigidez e eficiência do inibidor. O loop reativo dos inibidores da família STI Kunitz está localizado entre as subunidades A e B e a clivagem da ligação peptídica *scissile* teoricamente poderia provocar a separação da subunidade A das subunidades B e C. Entretanto as subunidades A, B e C são unidas por uma extensa rede de pontes de hidrogênio intramoleculares. E existe substancial evidência de que esta rede de pontes de hidrogênio, a qual não existe num substrato comum vulnerável à proteólise, permanece intacta no inibidor intacto (SHAW *et al.*, 1995). Esses numerosos contatos estabilizam a nova porção N-terminal formada e a mantêm numa orientação ótima para um ataque nucleofílico ao acil-enzima o que resulta na religação. Além disso, o posicionamento do grupo amina do inibidor também impede estericamente que a molécula de água entre até uma proximidade necessária para a ativação nucleofílica da base de histidina da enzima, o que é necessário para que a clivagem se concretize.

A composição do sítio ativo determina a especificidade do inibidor. O aminoácido na posição P1, o qual se liga ao sítio ativo da enzima, é de particular importância. Além da especificidade, a conformação do loop também determina a eficiência do inibidor e quanto mais próximo à conformação canônica o loop estiver, mais efetivo será o inibidor, pelo menos para os inibidores que seguem o mecanismo padrão (LASKOWSKY Jr *et al.*, 2000).

A falta de um resíduo hidrofóbico na posição P1 de DrTI, o qual é indispensável para a inibição de quimotripsina, pode explicar porque DrTI não tem qualquer atividade contra essa enzima. A atividade de DrTI contra tripsina pode ser compreendida levando-se em conta que a presença de um resíduo positivo na posição P1 (=Lys69) é essencial. Entretanto, quando comparado com outros inibidores de tripsina canônicos ($K_i \sim 1\text{nM}$), a atividade de DrTI contra tripsina é fraca ($K_i = 21,9\text{nM}$) e isso pode ser explicado pelo fato de que a conformação do seu loop reativo é ligeiramente distorcida devido a inserção de um resíduo (Glu68) entre os resíduos P1 e P2.

É a combinação de dois fatores cruciais que ajuda a explicar porque DrTI é um inibidor efetivo de HPK com $K_i = 5\text{nM}$ a despeito de não apresentar qualquer efeito inibitório sobre TK. O primeiro é a presença de dois ácidos glutâmicos no loop reativo de

DrTI, Glu67 e Glu68. A análise das regiões que estão envolvidas nas supostas interações entre o inibidor e a enzima, demonstra que existe uma substituição da Gln192 na molécula de tripsina pela Lys583 em HPK. Essa substituição, em adição às supostas interações com os resíduos Ser65, Glu68, Lys69 e Gln70 do loop reativo de DrTI, poderia abrir a possibilidade de novas interações eletrostáticas com as cadeias laterais de Glu67 e Glu68. Já na estrutura de TK, a Gln192(tripsina) é substituída pela Met192, a qual por ser mais inerte elimina a possibilidade de contatos.

O outro fator crucial é a presença de inserções na estrutura de TK nas áreas que fazem contatos com o inibidor. Essas inserções modificam a superfície molecular de TK dificultando as interações do inibidor com a enzima. O mesmo não ocorre com HPK que mostra uma seqüência bastante próxima à da tripsina, o que indica que a superfície de contato entre DrTI e HPK é presumivelmente preservada.

Embora nossos estudos não possam absolutamente assegurar que não existam outras razões estruturais, eles providenciam uma base para estudos de mutações sítio-dirigidas com o objetivo de verificar se as possíveis explicações para as diferenças observadas para a atividade inibitória de DrTI.

Os complexos de DrTI com tripsina porcina e com calicreína do plasma foram preparados e muitos experimentos de cristalização foram feitos, embora nenhum deles tenha resultado em cristais adequados à coleta de dados.

CONCLUSÕES

A determinação das estruturas cristalográficas dos dois inibidores possibilitou um contato abrangente com a cristalografia de proteínas. Primeiro com a resolução *ab initio* da estrutura de CTI usando MIRAS e depois com a resolução da estrutura de DrTI por substituição molecular usando a estrutura de CTI como modelo de busca.

Em ambos os casos a estrutura cristalográfica foi importante para esclarecer os resultados bioquímicos encontrados para cada um deles. No caso de CTI, a determinação da sua estrutura cristalográfica foi fundamental para esclarecer que a presença de dois inibidores diferentes, como foi primeiramente publicado, tratava-se na verdade de um único inibidor com duas cadeias polipeptídicas não covalentemente ligadas. Além disso, CTI é o primeiro inibidor tipo STI Kunitz com duas cadeias e uma única ponte dissulfeto completamente ativo a ter a sua estrutura elucidada. E é a análise da sua estrutura cristalográfica que ajuda a entender porque, apesar dessas discrepâncias estruturais, CTI é um inibidor efetivo de tripsina. Como os inibidores tipo Kunitz padrões (STI, ETI e WCI), CTI também apresenta a conformação canônica para o seu loop reativo, o que faz com que a maioria dos contatos que seriam feitos com a tripsina na formação do complexo enzima-inibidor sejam preservados. Esses contatos são preservados não apenas nas regiões dos sítios ativos da tripsina e de CTI mas, existe a formação efetiva de uma interface de contato entre as moléculas, que envolve várias partes das moléculas e que não é afetada pela quebra entre as cadeias polipeptídicas ou pela falta de uma das pontes dissulfeto das duas que são características dos inibidores padrões.

No caso de DrTI, a sua estrutura cristalográfica mostra algumas peculiaridades que ajudam a explicar a sua atividade inibitória diferenciada. A atividade inibitória de DrTI contra tripsina é mais fraca com relação aos inibidores tipo Kunitz padrões, entretanto DrTI é um inibidor efetivo de caliceína do plasma humano embora não apresente qualquer atividade contra caliceína de tecido. Embora a sua estrutura global seja comum, com uma cadeia polipeptídica e com duas pontes dissulfeto nas posições esperadas, a composição do seu loop reativo é diferente com a presença de dois ácidos glutâmicos, A67 e A68, e a

conformação do seu loop é distorcida em relação à conformação canônica devido à inserção de um resíduo, justamente Glu-A68, entre os resíduos nas posições P1 e P2 do loop reativo (seguindo a numeração de STI). Na verdade existe uma combinação de fatores que ajuda a entender o comportamento de DrTI. A distorção do seu loop reativo explica o enfraquecimento da sua atividade inibitória sobre tripsina em relação aos inibidores padrões cujo loop reativo apresenta a conformação canônica. Quanto mais próxima da conformação canônica tanto mais efetivo é o inibidor. Isso funciona para tripsina, mas para explicar a sua atividade contra caliceína do plasma humano (HPK), a despeito da falta de atividade contra caliceína de tecido (TK), é preciso analisar os possíveis contatos que DrTI faria com essas enzimas nos complexos enzima-inibidor. Analisando essas regiões de contato, verifica-se que existe uma substituição do resíduo Gln192 em tripsina por Lys583 em HPK que, além dos contatos com os resíduos Ser65, Glu68, Lys69 e Gln70 do loop reativo de DrTI, abre a possibilidade de novas interações eletrostáticas com Glu67 e intensifica os contatos com Glu68. O resíduo Gln192 em tripsina é substituído por Met192 em TK que não possibilita tais interações. Além disso, a estrutura de TK mostra que existem algumas inserções nas regiões que fariam contatos com o inibidor. Essas inserções modificam a superfície molecular de TK impedindo a formação de uma interface de contatos efetiva. Essas inserções não estão presentes em HPK, cuja sequência é bem próxima à da tripsina, o que indica que a superfície de interação entre DrTI e HPK é presumivelmente preservada.

É a partir desse cruzamento interdisciplinar de resultados, numa busca por caminhos que ajudem a explicar o comportamento bioquímico em nível atômico, que a relação estrutura-função fornece informações científicas valiosas. Essas informações podem ser usadas para estabelecer o mecanismo de ação/interação de proteínas, para planejar estratégias para o desenvolvimento de fármacos específicos, para aplicações biotecnológicas como o desenvolvimento/aperfeiçoamento de métodos laboratoriais, etc.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABRAHAMAS, J. P., LESLIE, A. G. W. *Acta Cryst.* D52, p.30-42, 1996.
- AGERBERTH, B., SÖDERLING-BARROS, J., JÖRNVALL, H., CHEN, S.-W., ÖSTENSEN, C.-G., EFENDIC, S., & MUTT, V. Isolation and characterization of a 60 residue intestinal peptide structurally related to the pancreatic secretory type of trypsin inhibitor: Influence on insulin secretion. *Proc. Nat. Acad. Sci.*, 86, p.8590-4, 1989.
- APOSTOLUK, W. & OTLEWSKI, J. Variability of canonical loop conformation in serine proteinase inhibitors and other proteins. *Proteins* 32, p.45974, 1998.
- BERGFORDS, T. M. *Protein Crystallization: techniques, strategies and tips*. International University Line, La Jolla, California, 1999.
- BHoola, K. D., FIGUEROA, C. D., WORTHY, K. Bioregulation of kinins: kallikreins, kininogens and kininanes, *Pharmacol Rev* 44, p.1-80, 1992.
- BIJVOET, M. *Nature* 173, p.88891, 1954.
- BILLINGS, P. C., MORROW, A. R. RYAN, C. A. & KENNEDY, A. R. *Carcinogenesis*, 10, p.68791, 1989.
- BILLINGS, P. C., NEWBERNE, P. M. & KENNEDY, A. R. *Carcinogenesis*, 11(7), p.10836, 1990.
- BLANCO-LABRA, A., CHAGOLLA-LOPEZ, A., GALLARDO, M. & RODRIGUEZ, V., Further Characterization of the 12 kDa Protease I α -amylase Inhibitor Present in Maize Seeds. *Food & Nutrition Press*, 3, p.27-41, 1995.
- BLESSING, R. H. & SMITH, G. D. Difference structure factor normalization for heavy-atom or anomalous-scattering substructure determinations. *J. Appl. Crystallog.*, 32, p.664–70, 1999.
- BODE, W. & HUBER, R., *Natural Protein Proteinase Inhibitors and Their Interaction with Proteinases*. *Eur. J. Biochem.*, 204, p.433-41, 1992.
- BELLOSTA, P., IWAHORI, A., PLOTNIKOV, A. N., ELISEENKOVA, A. V., BASILICO, C. & MOHAMMADI, M. Identification of receptor and heparin binding sites in fibroblast growth factor 4 by structure-based mutagenesis. *Mol. Cell. Biol.*, 21(17), p.5946–57, 2001.

BIRK, Y. & APPLEBAUM, Y., The Bowman-Birk Inhibitor: Trypsin and Chymotrypsin Inhibitor from Soybeans. *Int. J. Pept. Protein Res.*, 25, p.113-31, 1985.

BLEVINS, R. A., TULINSKY, A. The refinement and the structure of the dimer of alpha-chymotrypsin at 1.67 angstroms resolution, *J. Biol. Chem.* 260, p.4264, 1985.

BLUNDELL, T.L. & JOHNSON, L.N. *Protein Crystallography*. Academic Press, New York, 1976.

BRÄNDÉN, C. I. & JONES, T. A. Between objective and subjective. *Nature*, 343, p.6879, 1990.

BRINKMANN, T., SCHÄFERS, J., GÜRTLER, L., KIDO, H., NIWA, Y., KATUNUMA, N., AND TSCHESCHE, H. Inhibition of tryptase TL2 from human T4+ lymphocytes and inhibition of HIV-1 replication in H9 cells by recombinant aprotinin and bikunin homologues, *J. Prot. Chem.*, 16, p.651-60, 1997.

BRUENING, G., SANDERSON, J. L. *J. Cell. Biochem. (Suppl.)* 8B, p.1418, 1984.

BRÜNGER, A. T., KURIYAN, J. & KARPLUS, M. Crystallographic R Factor Refinement by Molecular Dynamics. *Science* 235, p.45860, 1987.

BRÜNGER, A.T.; KRUKOWSKI, A.; ERICKSON, J.W. Slow-cooling protocols for crystallographic refinement by simulated annealing, *Acta Cryst.* A46, p.585-93, 1990.

BRÜNGER, A.T. "X-PLOR Version 3.1. A System for X-ray Crystallography and NMR", Yale University Press, 1992.

BRÜNGER, A. T. Free R-value a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355, p.4725, 1992.

BRÜNGER, A. T. & NILGES, M., Computational Challenges for Macromolecular Structure Determination by X-ray Crystallography and Solution NMR Spectroscopy. *Quarterly Rev. Biophys.* 26, p.49, 1993.

BRÜNGER, A.T., ADAMS, P.D., CLORE, G.M., DELANO, W.L., GROS, P., GROSSE-KUNSTLEVE, R.W., JIANG, J., KUSZEWSKI, J., NILGES, M., PANNU, N.S., READ, R.J., RICE, L.M., SIMONSON, T., WARREN, G.L. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Cryst.* D54, p.905–21, 1998.

CHEN, Z., BODE, W. Refined 2.5 angstroms x-ray crystal structure of the complex formed by porcine kallikrein A and the bovine pancreatic trypsin inhibitor., *J. Mol. Biol.* 164, p.283, 1983.

CHRISTELLER, J. T., LANIG, W. A., MARKICK, N.P. & BURGESS, E. P. J. Midgut Proteases Activities in 12 Phytophagous Lepidoteran Larvae: Dietary and Protease Inhibitors Interaction. *Insect. Biochem. Mol. Biol.*, 22, p.735-46, 1992.

CHUNG, C. H., IVES, H. E., ALMEIDA, S., GOLBERG, A. L. *J. Biol. Chem.* 258, p.110328, 1983.

CHUNG, D. W., FUJIKAWA, K., MCMULLEN, B. A., DAVIE, E. W. Human plasma prekallikrein, a zymogene to a serine protease that contains four tandem repeats, *Biochem.* 25, p.2410-7, 1986.

CLEMENTS, J. A. The molecular biology of the kallikreins and their roles in inflammation, in: S. G. Farmer (Ed.), *The Kinin System*, Academic Press, San Diego, CA, vol. 15, p.71-97, 1997.

Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst.*, D50, p.760-3, 1994.

CORONEL, C. E., SAN AGUSTIN, J., LARDY, H. A. Purification and structure of caltrin-like proteins from seminal vesicle of the guinea pig. *J Biol Chem.* 265(12), p.68549, 1990.

COWTAN, K Modified phased translation functions and their application to molecular fragment location. *Acta Cryst.* D54, p.750-6, 1998.

CROWTHER, R.A. *The Molecular Replacement Method*; Rossmann, M.G., Ed.; Gordon and Breach: New York; p.174-8, 1972.

DAUTER, Z., DAUTER, M. & RAJASHANKAR, K. R. Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Cryst.* D56, p.232-7, 2000.

DAVIES D.R., *Annu Rev Biophys Biophys Chem.*19, p.180-215, 1990.

DEAR, T. N., RAMSHAW, I. A., KEFFORD, R. F. Differential expression of a novel gene, WDNM1, in nonmetastatic rat mammary adenocarcinoma cells. *Cancer Res.* 48(18), p.52039, 1988.

DE LA FORTELLE, E., BRICOGNE, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.*, 276, p.472-94, 1997.

DENZER, A.J., SCHULTHESS T, FAUSER, C., SCHUMACHER, B., KAMMERER, R.A., ENGEL, J., AND RUEGG, M.A. Electron microscopic structure of agrin and mapping of its binding site in laminin-1. *EMBO J.*, 17, p.335-43, 1998.

- DODSON, E. J. The role of validation in macromolecular crystallography. *Acta Cryst.* D54, p.110918, 1998.
- DRENTH, J., Principles of Protein X-Ray Crystallography. Springer-Verlag Berlin Heidelberg, New York, NY, EUA, 1994.
- DUCRUIX, A. E GIEGÉ, R. Crystallization of Nucleic Acids and Proteins. A practical approach. Second Edition, Oxford University Press, Oxford, New York, 1999.
- EMSLEY, P., FOTINOU, I., FAIRWEATHER, N. F., CHARLES, I. G., WATTS, C., HEWITT, E., ISAACS, N. W. The Structures of the H_C Fragment of Tetanus Toxin with Carbohydrate Subunit Complexes Provide Insight into Ganglioside Binding, *J. Biol. Chem.* 275(12), p.8889–94, 2000.
- ERIKSSON, A. E., COUSENS, L. S. & MATTHEWS, B. W. Refinement of the structure of human basic fibroblast growth factor at 1.6-Å resolution and analysis of presumed heparin binding sites by selenate substitution. *Prot. Sci.*, 2(8), p.1274–84, 1993.
- FINKENSTADT, W. R., LASKOWSKY, M. Jr. *J. Biol. Chem.* 240, p.962, 1965.
- FOLKERS, P. J., CLORE, G. M., DRISCOLL, P. C., DODT, J., KOHLER, S., GRONENBORN, A. M. Solution structure of recombinant hirudin and the Lys-47=Glu mutant: a nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing study. *Biochemistry* 28(6), p.260117, 1989.
- FRENKEL, K. CHRZAN, K. RYAN, C. A. WIESNER, R. & TROLL, W. *Carcinogenesis*, 8, p.1207-12, 1987.
- GATEHOUSE, A. M. R., *Antinutritional Proteins in Plants. Developments in Food Proteins*, B.J.F. ed., Elsevier, 3, p.245-93, 1984.
- GIACOVAZZO, C. et al., *Fundamentals of Crystallography*. International Union of Crystallography, Oxford University Press, Oxford, EUA, 1992.
- GRASBERGER, B. L., CLORE, G. M., GRONENBORN, A. M. High-resolution structure of *Ascaris* trypsin inhibitor in solution: direct evidence for a pH-induced conformational transition in the reactive site. *Structure* 2(7), p.66978, 1994.
- GRAVES, B. J., HATADA, M. H., HENDRICKSON, W. A., MILLER, J. K., MADISON, V. S. & SATOW, Y. Structure of Interleukin 1 α at 2.4Å resolution. *Biochemistry*, 29(11), p.2679–89, 1990.
- GREENSPAN, D. S. The carboxy terminal half of type VII collagen including the NC-2 domain and intron/exon organization of the corresponding region of the COL7A1 gene. *Hum. Mol. Genet.* 2, p.273-8, 1993.

GUSS, J. M., MERRITT, E. A., PHIZACKERLEY, R. P., HEDMAN, B., MURATA, M. HODGSON, K. O. & FREEMAN, H. C. Phase determination by multiple-wavelength X-ray diffraction: crystal structure of a basic blue copper protein from cucumbers. *Science* 241, p.806-11, 1989.

HABAZETTL, J., GONDOL, D., WILTSHECK, R., OTLEWSKI, J., SCHLEICHER, M. & HOLAK, T. A. Structure of hisactophilin is similar to interleukin-1b and fibroblast growth factor. *Nature*, 359(6398), p.855–8, 1992.

HALLIDAY, D., RESNICK, R. & WALKER, J., *Fundamentos de Física*, 4a ed. LTC Editora SA, Rio de Janeiro, RJ, Brasil, 1993.

HANNAPEL, D. J. Nucleotide and deduced amino acid sequence of the 22kDa cathepsin D inhibitor protein of potato (*Solanum tuberosum* L.). *Plant Physiol.* 101, p.7034, 1993.

HENDRICKSON, W. A. & TEETER, M. M. *Nature* 290, p.107-13, 1981.

HENDRICKSON, W. A. Determination of Macromolecular Structures from Anomalous Diffraction of Synchrotron Radiation. *Science* 254, p. 518, 1991.

HENNIGHAUSEN, L. G., SIPPEL, A. E. Mouse whey acidic protein is a novel member of the family of 'four-disulfide core' proteins. *Nucleic Acids Res.* 10(8), p.267784, 1982.

HILDMANN, T., EBNETH, M., PENA-CORTES, H., SANCHEZ-SERRANO, J. J., WILLMITZER, L., AND PRAT, S. General roles of abscisic and jasmonic acids in gene activation as a result of mechanical wounding. *The Plant Cell*, 4, p.1157-70, 1992.

HIRONO, S., AKAGAWA, H., MITSUI, Y., IITAKA, Y. Crystal structure at 2.6 Å resolution of the complex of subtilisin BPN' with streptomyces subtilisin inhibitor. *J. Mol. Biol.* 178(2), p.389-414, 1984.

HOGG, P.J., JIMENEZ, B. M., AND CHESTERMAN, C. N. Identification of possible inhibitory reactive centers in thrombospondin 1 that may bind cathepsin G and neutrophil elastase. *Biochemistry*, 33, p.6531-7, 1994.

HOHENESTER, E., MAURER, P., AND TIMPL, R. Crystal structure of a pair of follistatin-like and EF-hand calcium binding domains in BM-40. *EMBO J.*, 16, p.3778-86, 1997.

HOOFT, R. W. W., SANDER, C., VRIEND, G. Errors in protein structures. *Nature*, 381, p.272, 1996.

HUBBARD, S. J., CAMPBELL, S. F. & THORNTON, J. M. Molecular recognition: Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. *J. Mol. Biol.*, 220, p.507–30, 1991.

- HUBBARD, S. J., EISENMENGER, F. & THORNTON, J. M. Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. *Prot. Sci.*, 3(5), p.757–68, 1994.
- HUBBARD, S. J. The structural aspects of limited proteolysis of native proteins. *Biochim. Biophys. Acta*, 1382, p.191–206, 1998.
- HUTCHINSON, E. G. & THORNTON, J. M. PROMOTIF - a program to identify and analyze structural motifs in proteins. *Prot. Sci.*, 5(2), p.212–20, 1996.
- HUYNH, Q. K., BORGMAYER, J. R. & ZOBEL, J. F., Isolation and Characterization of a 22 kDa Protein with Antifungal Properties from Maize Seeds. *Biochemical and Biophysical Research Communications* 182, p.1, 1992.
- ITOH, H., IDE, H., ISHIKAWA, N., AND NAWA, Y. Mast cell protease inhibitor, trypstatin, is a fragment of inter- α -trypsin inhibitor light chain. *J. Biol. Chem.*, 269, p.3818-22, 1994.
- JACKSON, R. M. Comparison of proteinprotein interactions in serine protease inhibitor and antibodyantigen complexes: implications for the protein docking problem. *Protein Sci.* 8, p.60313, 1999.
- JONES, T. A., BERGDOLL, M., KJELDGAARD, M. Crystallographic and Modeling Methods in Molecular Design. (BUGG, C., EALICK, S., eds), Springer–Verlag Press, p.189–95, 1990.
- KENNEDY, A. R., Prevention of Carcinogenesis by Protease Inhibitors. *Cancer Research*, 54, p.19995-20055, 1994.
- KENNEDY, A. R. *Pharmacol. Ther.*, 78(3), p.167-209, 1998.
- KLEYWEGT, G.J. & JONES, T.A. From First Map to Final Model. BAILEY, S., HUBBARD, R. & WALLER, D.A. Eds., SERC Daresbury Laboratory, Warrington, U.K., p.59-66, 1994.
- KOHFELDT, E., GÖHRING, W., MAYER, U. ZWECKSTETTER, M., HOLAK, T.A., CHU, M.-L., AND TIMPL, R. Conversion of the Kunitz type module of collagen VI into a highly active trypsin inhibitor by site-directed mutagenesis. *Eur. J. Biochem.*, 238, p.333-40, 1996.
- KONDO, K., TODA, H., AND NARITA, K. Amino acid sequence of α -bungarotoxin from *Bungarus multicinctus* venom. The amino acid substitutions in the B chains. *J. Biochem.*, 91, p.1519-30, 1982.

KORTT, A.A., STRIKE, P.M., AND DE JERSEY, J. Amino acid sequence of a crystalline seed albumin (winged bean albumin-1) from *Psophocarpus tetragonolobus* (L.) DC. *Eur. J. Biochem.*, 181, p.403-8, 1989.

KRIZAJ, I., DROBNIC-KOSOROK, M., BRZIN, J., JERALA, R., AND TURK, V. The primary structure of cysteine proteinases from potato. *FEBS Letters*, 333, p.15-20, 1993.

KRYSTEK, S., STOUCH, T. & NOVOTNY, J. Affinity and specificity of serine endopeptidase-protein inhibitor interactions: empirical free energy calculations based on X-ray crystallographic structures. *J. Mol. Biol.*, 234, p.661–79, 1993.

KUMAZAKI, T., KAJIWARA, K., KOJIMA, S., MIURA, K., AND ISHII, S. Interaction of *Streptomyces subtilisin* inhibitor (SSI) with *Streptomyces griseus* metallo-endopeptidase II (SGMP II). *J. Biochem.*, 114, p.570-5, 1993.

KUNITZ, M., NORTHROP, J. H. *J. Gen. Physiol.*, 19, p.991–1007, 1936.

KUNITZ, M. *Science*, 101, p.668–9, 1945.

LACY, D. B., TEPP, W., COHEN, A. C., DASGUPTA, B. R., STEVENS, R. C. Crystal structure of botulinum neurotoxin type A and implications for toxicity, *Nat. Struct. Biol.* 5, p.898–902, 1998.

LAPATTO, R., KRENGEL, U., SCHREUDER, H. A., ARKEMA, A., DE BOER, B., KALK, K. H., HOL, W. G., GROOTENHUIS, P. D., MULDER, J. W., DIJKEMA, R., THEUNISSEN, H. J., DIJKSTRA, B. W. X-ray structure of antistasin at 1.9 Å resolution and its modeled complex with blood coagulation factor Xa. *EMBO J.* 16(17), p.5151-61, 1997.

LASKOWSKI, R. A., MACARTHUR, M. W., MOSS, D. S., & THORNTON, J. M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* 28, p.338-40, 1993.

LASKOWSKY Jr, M. & KATO, I. Protein inhibitors of proteinases. *Annu. Rev. Biochem.*, 49, p.593–626, 1980.

LASKOWSKY Jr., M., QASIM, M. A. & STEPHEN, M. L. Protein–Protein recognition (KLEANTHOS, C., HAMES, B. D. & GLOVER, D. M., eds), Oxford University Press, New York, p.228–79, 2000.

LAWRENCE, P. K., KOUNDAL, K. R. Plant protease inhibitors in control of phytophagous insects. *Electronic J. of Biotechnology* 5(1), p.93-109, 2002.

LEGOUIS R, HARDELIN JP, LEVILLIERS J, CLAVERIE JM, COMPAIN S, WUNDERLE V, MILLASSEAU P, LE PASLIER D, COHEN D, CATERINA D. The

candidate gene for the X-linked Kallmann syndrome encodes a protein related to adhesion molecules. *Cell*, 67(2), p.42335, 1991.

LEHLE, K., WRBA, A. & JAENICKE, R. Erythrina trypsin inhibitor retains its native structure and function after reducing its disulfide groups. *J. Mol. Biol.*, 239, p.276–84, 1994.

LIANG, Z., SOTTRUP-JENSEN, L., ASPAN, A., HALL, M., SODERHALL, K. Pacifastin, a novel 155-kDa heterodimeric proteinase inhibitor containing a unique transferrin chain. *Proc Natl Acad Sci USA* 94(13), p.66827, 1997.

LINDQVIST, A., AND AKERSTRÖM, B. Bovine α 1-microglobulin / bikunin. Isolation and characterization of liver cDNA and urinary α 1-microglobulin. *Biochim. Biophys. Acta*, 1306, p.98-106, 1996.

LIU, Y., CHIRINO, A. J., MISULOVIN, Z., LETEUX, C., FEIZI, T., NUSSENZWEIG, M. C. & BJORKMAN, P. Crystal structure of the cysteine-rich domain of mannose receptor complexed with a sulfated carbohydrate ligand. *J. Exptl. Med.*, 191(7), p.1105–15, 2000.

LORENZI, H. Árvores brasileiras: Manual de identificação e Cultivo de plantas arbóreas nativas do Brasil, pp. 36, Platarum Ed., Nova Odessa, SP, Brazil, 1992.

MACEDO, M. L. R., ANDRADE, L. B. S., MORAES, R. A. & XAVIER-FILHO, J. Vicilins Variants and the Resistance of Cowpea (*Vigna unguiculata*) Seeds to the Cowpea Weevil (*Callosobruchus maculatus*). *Comp. Biochem. Physiol.*, 105, p.89-94, 1994.

MANEN, J.-F., SIMON, P., VANSLOOTEN, J.-C., ØSTERÅS, M., FRUTIGER, S. & HUGHES, G.J. A nodulin specifically expressed in senescent nodules of winged bean is a protease inhibitor. *The Plant Cell*, 3, p.259-70, 1991.

MARES, M., MELOUN, B., PAVLIK, M., KOSTKA, V., AND BAUDYS, M. Primary structure of cathepsin D inhibitor from potatoes and its structure relationship to soybean trypsin inhibitor family. *FEBS Letters*, 251, p.94-8, 1989.

MARSHALL, D.L. & HARAVEY, A.L. Protease inhibitor homologues of dendrotoxin do not bind to dendrotoxin acceptors on synaptosomal membranes or facilitate neuromuscular transmission. *Biol. Chem. Hoppe-Seyler*, 373, p.707-14, 1992.

MATTHEWS, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.*, 33, 491–497.

MATZUK, M.M., LU, N., VOGEL, H., SELLHEYER, K., ROOP, D.R., AND BRADLEY, A. Multiple defects and perinatal death in mice deficient in follistatin. *Nature*, 374, p.360-3, 1995.

MCCOY, A. J. & KORTT, A. A. The 1.8 Å crystal structure of winged bean albumin 1, the major albumin from *Psophocarpus tetragonolobus* (L.) DC. *J. Mol. Biol.*, 269, p.881–91, 1997.

MCGRATH, M. E., GILLMOR, S. A., FLETTERICK, R. J. Ecotin lessons on survival in a protease-filled world. *Protein Sci.* 4, p.1418, 1995.

MCLACHLAN, A. D. Three-fold structural pattern in the soybean trypsin inhibitor (Kunitz). *J. Mol. Biol.*, 133, p.557–63, 1979.

MCPHERSON, A., *Crystallization of Biological Macromolecules*. Cold Spring Harbor Laboratory Press, New York, 1994.

MCREE, D. *Practical Protein Crystallography*, Academic Press, Inc.: San Diego, p.42, 1993.

MOSS, G.W., MARSHALL, J., MORABITO, M., HOWE, J.R., AND MOCZYDLOWSKI, E. An evolutionarily conserved binding site for serine proteinase inhibitors in large conductance calcium-activated potassium channels. *Biochemistry*, 35, p.16024-35, 1996.

MUNDY, J., HEJGAARD, J., AND SVENDSEN I. Characterization of a bifunctional wheat inhibitor of endogenous α -amylase and subtilisin. *FEBS Letters* 167, p.210, 1984.

NAGEM, R. A. P., DAUTER, Z. & POLIKARPOV, I. Protein crystal structure solution by fast incorporation of negatively and positively charged anomalous scatterers. *Acta Cryst. D*57, p. 996–1002, 2001.

NAVAZA, J. AMORE: an automated procedure for molecular replacement, *Acta Crystallog. A*50, p.157–63, 1994.

NEURATH, H. Proteolytic processing and physiological regulation. *Trends Biochem. Sci.* 14, p.268–71, 1989.

NG, K. K., PETERSEN, J. F., CHERNEY, M. M., GAREN, C., ZALATORIS, J. J., RAO-NAIK, C., DUNN, B. M., MARTZEN, M. R., PEANASKY, R. J., JAMES, M. N. Structural basis for the inhibition of porcine pepsin by *Ascaris* pepsin inhibitor-3. *Nat. Struct. Biol.* 7(8), p.653-7, 2000.

NOTENBOOM, V., BORASTON, A. B., WILLIAMS, S. J., KILBURN, D. G. & ROSE, D. R. High resolution crystal structures of the lectin-like xylan binding domain from

Streptomyces lividans xylanase 10A with bound substrates reveal a novel mode of xylan binding. *Biochemistry*, 41(13), p.4246–54, 2002.

NOVOTNY, J. & BRUCCOLERI, R. E. Correlation among sites of limited proteolysis, enzyme accessibility and segmental mobility. *FEBS Letters*, 211, p.185–9, 1987.

ODANI, S., KOIDE, T. & ONO, T., Wheat Germ Trypsin Inhibitors: Isolation and Structural Characterization of a Single-headed and Double-headed Inhibitor of the Bowman-Birk Type. *J. Biochem.*, 100, p.975-83, 1986.

OHTSUBO, K.-I. & RICHARDSON, M. The amino acid sequence of a 20 kDa bifunctional subtilisin / α -amylase inhibitor from bran of rice (*Oryza sativa* L.) seeds. *FEBS Letters*, 309, p.68-72, 1992.

OLDFIELD, T. J. SQUID: A program for the analysis and display of data from crystallography and molecular dynamics. *J. Mol. Graphics* 10, p.24752, 1992.

ONESTI, S., BRICK, P. & BLOW, D. M. Crystal structure of a Kunitz-type trypsin inhibitor from *Erythrina caffra* seeds. *J. Mol. Biol.*, 217, p.153–76, 1990.

ONO, S., YAMAKITA, Y., YAMASHIRO, S., MATSUDAIRA, P. T., GNARRA, J. R., OBINATA, T. & MATSUMURA, F. Identification of an actin binding region and a protein kinase C phosphorylation site on human fascin. *J. Biol. Chem.*, 272(4), p.2527–33, 1997.

OTWINOWSKI, Z. Proceedings of the CCP4 Study Weekend: Data Collection and Processing (SAWYER, L., ISAACS, N. & BAILEY, S., eds), SERC Daresbury Laboratory, Daresbury, England, p.56–62, 1993.

PANDO, S. C., OLIVA, M. L. V., SAMPAIO, C. A. M., DI CIERO, L., NOVELLO, J. C., MARANGONI, S. Primary sequence determination of a Kunitz inhibitor isolated from *Delonix regia* seeds, *Phytochem* 57, p.625–31, 2001.

PENG, J. H., BLACK, L. L. *Phytopathology* 66, p.95863, 1976.

PERRAKIS, A., SIXMA, T. K., WILSON, K. S., LAMZIN, V. S. wARP: improvement and extension of crystallographic phases by weighted averaging of multiple refined dummy atomic models. *Acta Cryst.*, D53, p.448–55, 1997.

PLOTNIKOV, A. N., ELISEENKOVA, A. V., IBRAHIMI, O. A., SHRIVER, Z., SASISEKHARAN, R., LEMMON, M. A. & MOHAMMADI, M. Crystal structure of fibroblast growth factor 9 reveals regions implicated in dimerization and autoinhibition. *J. Biol. Chem.*, 276(6), p.4322–9, 2001.

POLIKARPOV, I., GOLUBEV, A. M., PERLES, L. A., PANDO, S. C., NOVELLO, J. C., MARANGONI, S., 1999. *Acta Crystallog.* D55, 1611–1613.

POLIKARPOV, I., OLIVA, G., CASTELLANO, E. E., GARRATT, R., ARRUDA, P., LEITE, A. & CRAIEVICH, A. Protein crystallography station at LNLS, The Brazilian National Synchrotron Light Source. Nucl. Instrum. Methods A405, p.159–64, 1998.

PONTING, C. P. & RUSSELL, R. B. Identification of distant homologues of FGFs suggests a common ancestor for all beta-trefoil proteins. J. Mol. Biol., 302, p.1041–7, 2000.

PONTIUS, J., RICHELLE, J. & WODAK, S. Deviations from standard atomic volumes as a quality measure for protein crystal structures. J. Mol. Biol. 264, p.12136, 1996.

PRAKASH, B., SELVARAJ, S., MURTHY, M. R. N., SREERAMA, Y. N., RAO, D. R. & GOWDA, L. R., Analysis of Aminoacid Sequences of Plant Bowman-Birk Inhibitors. Biochem. Molec. Biol., 22, p.41-7, 1996.

PRIESTLE, J. P., SCHÄR, H. P. & GRÜTTER, M. Crystal structure of the cytokine interleukin-1 β . EMBO J., 7(2), p.339–43, 1988.

RADISKY, E. S. & KOSHLAND Jr., D. E. A clogged gutter mechanism for protease inhibitors. Proc. Natl Acad. Sci. USA, 99(16), p.10316–21, 2002.

RAMACHANDRAN, G. N., RAMAKRISHNAN, C. & SASISEKHARAN, V. Stereochemistry of polypeptide chain configurations. J. Mol. Biol. 7, p.95-9, 1963.

RAVICHANDRAN, S., SEN, U., CHAKRABARTI, C. & DATTA GUPTA, J. K. Cryocrystallography of a Kunitz-type serine protease inhibitor: The 90 K structure of a winged bean chymotrypsin inhibitor (WCI) at 2.13 Å resolution. Acta Cryst. D55, p.1814–21, 1999.

RAWLINGS, N.D. & BARRETT, A.J. Biochem. J., 290, p.205-18, 1993.

RICHARDSON, M. Seed Storage Proteins: The Enzyme Inhibitors. Methods in Plant Biochemistry, Amino Acids, Proteins and Nucleic Acids. Rogers, J. ed., v.5, 1991.

RITONJA, A., KRIZAJ, I., MESKO, P, KOPITAR, M., LUCOVNIK, P., OTRUKELJ, B., PUNGERCAR, J., BUTTLE, D.J., BARRETT, A.J., AND TURK, V. The amino acid sequence of a novel inhibitor of cathepsin D from potato. FEBS Letters, 267, p.13-15, 1990.

ROSSMANN, M. G. Acta Cryst. 14, p.3838, 1961.

ROULD, M.A. Methods in Enzymology - Macromolecular Crystallography Part A. C.W. CARTER & R.M. SWEET ed. Academic Press, San Diego, v. 276. Screening for Heavy-Atom Derivatives and Obtaining Accurate Isomorphous Differences, p.461-72, 1997.

RUTENBER, E., KATZIN, B. J., ERNST, S., COLLINS, E. J., MLSNA, D., READY, M. P. & ROBERTUS, J. D. Crystallographic Refinement of Ricin to 2.5 Å. *Proteins*, 10(3), p.240–50, 1991.

RYAN, C. A., Proteinase Inhibitors. *Biochemistry of Plants*, 6, p.351-70, 1981.

SALI, A. & BLUNDELL, T.L. *J.Mol.Biol.* 212, p.403-28, 1990.

SCHECHTER, I., BERGER, A. On the size of the active site in proteases *Biochem. Biophys. Res. Com.*, 27, p.157-62, 1967.

SCHREUDER, H., TARDIF, C., TRUMP-KALLMEYER, S., SOFFIENTINI, A., SARUBBI, E., AKESON, A., BOWLIN, T., YANOFSKY, S. & BARRETT, R. W. A new cytokine–receptor binding mode revealed by the crystal structure of the IL-1 receptor with an antagonist. *Nature*, 386(6621), p.194–200, 1997.

SHAW, G. L., DAVIS, B., KEELER, J. & FERSHT, A. R. Backbone Dynamics of Chymotrypsin Inhibitor 2. Effect of Breaking the Active-Site Bond and Its Implications for the Mechanism of Inhibition of Serine Proteases. *Biochemistry*, 34(7), p.2225–33, 1995.

SHELDRICK, G.M. Crystallographic Computing. MORAS, D. PODJARNY, A. D. & THIERRY, J. C. eds. I.U.Cr. & O.U.P.: Oxford, UK, p.14557, 1992.

SHERWOOD, D., "Crystals, X-rays and Proteins", Longman: London, 1976.

SKARZYNSKI, T. Crystal structure of dendrotoxin from the green mamba venom and its comparison with the structure of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.*, 224, p.671-83, 1992.

SAKAY, M., ODANI, S., IKENAKA, T. Importance of the carboxylterminal four amino acid residues in the inhibitory activity of *Streptomyces subtilisin* inhibitor (with a revision of its carboxylterminal sequence). *L. Biochem.* 87(3), p.8918, 1980.

SCHWEITZ, H., HEURTEAUX, C., BOIS, P., MOINIER, D., ROMEY, G., AND LAZDUNSKI, M. Calcicludine, a venom peptide of the Kunitz type protease inhibitor family, is a potent blocker of high-threshold Ca²⁺ channels with a high affinity for L-type channels in cerebellar granule neurons. *Proc. Nat. Acad. Sci.*, 91, p.878-82, 1994.

SILVA, J. A., MACEDO, M. L. R., NOVELLO, J. C. & MARANGONI, S. Biochemical characterization and N-terminal sequences of two new trypsin inhibitors from *Copaifera langsdorffii* seeds. *J. Prot. Chem.*, 20(1), p.1–7, 2001.

SONG, H. K. & SUH, S. W. Kunitz-type soybean trypsin inhibitor revisited: Refined structure of its complex with porcine trypsin reveals an insight into the interaction between

a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator. *J. Mol. Biol.*, 275, p.347–63, 1998.

TAMIR, S., KADNER, S., KATZ, J. & FINLAY, T. H. *Endocrinology*, 127(3), p.1319-28, 1990.

TANWILSON, A. L., RIGHTMIRE, B. R., WILSON, K. A. *Plant Physiol.* 70, p.4937, 1982.

THEERASILP, S., HITOTSUYA, H., NAKAJO, S., NAKAYA, K., NAKAMURA, Y., AND KURIHARA, Y. Complete amino acid sequence and structure characterization of the taste-modifying protein, miraculin. *J. Biol. Chem.*, 264, p.6655-9, 1989.

TRANSUE, T. R., SMITH, A. K., MO, H., GOLDSTEIN, I. J. & SAPER, M. A. Structure of benzyl T-antigen disaccharide bound to *Amaranthus caudatus* agglutinin. *Nat. Struct. Biol.*, 4(10), p.779–83, 1997.

TSUNOGAE, Y., TANAKA, I., YAMANE, T., KIKKAWA, J., ASHIDA, T., ISHIKAWA, C., WATANABE, K., NAKAMURA, S. & TAKAHASHI, K., Structure of the Trypsin-binding Domain of Bowman-Birk Type Protease Inhibitor and its Interaction with Trypsin. *J. Biochem.*, 100, p.1637-43, 1986.

TSUZUKI, S., FUSHIKI, T., KONDO, A., MURAYAMA, H., AND SUGIMOTO, E. Effect of a high-protein diet on the gene expression of a trypsin-sensitive, cholecystokinin-releasing peptide (monitor peptide) in the pancreas. *Eur. J. Biochem*, 199, p.245-52, 1991.

VAINSHTEIN, B. K. *Modern Crystallography. Fundamentals of crystals. Second Enlarged Edition.* Springer-Verlag, New York, NY, EUA, 1981.

VALLEE, F., KADZIOLA, A., BOURNE, Y., JUY, M., RODENBURG, K.W., SVENSSON, B., AND HASER, R. Barley α -amylase bound to its endogenous protein inhibitor BASI: Crystal structure of the complex at 1.9 D resolution. *Structure*, 6, p.649-59, 1998.

VAN DE LOCHT, A., LAMBA, D., BAUER, M., HUBER, R., FRIEDRICH, T., KRÖGER, B., HÖFFKEN, W., AND BODE, W. Two heads are better than one: Crystal structure of the insect derived double domain Kazal inhibitor rhodniin in complex with thrombin. *EMBO J.*, 14, p.5149-57, 1995.

VOET, D. & VOET, J. G. In *Biochemistry*, pp. 395–396, John Wiley & Sons Inc., New York, 1995.

WEEKS, C. M. & MILLER, R. The design and implementation of SnB v2.0. *J. Appl. Crystallog.*, 32, p.120–4, 1999.

WIEDOW, O., SCHRODER, J. M., GREGORY, H., YOUNG, J. A., CHRISTOPHERS, E. Elafin: an elastase-specific inhibitor of human skin. Purification, characterization, and complete amino acid sequence. *J Biol Chem.* 265(25), p.147915, 1990.

WILSON, K. S., BUTTERWORTH, S., DAUTER, Z., LAMZIN, V. S., WALSH, M., WODAK, S., PONTIUS, J., RICHELLE, J., VAGUINE, A., SANDER, C., HOOFT, R. W. W., VRIEND, G., THORNTON, J. M., LASKOWSKI, R. A., MACARTHUR, M. W., DODSON, E. J., MURSHUDOV, G., OLDFIELD, T. J., KAPTEIN, R., RULLMANN, J. A. C. Who checker the checkers? Four validation tools applied to eight atomic resolution structures (EU Validation Network) *J. Mol. Biol.* 276, p.417–36, 1998.

WITSCHI, H. & KENNEDY, A. R. *Carcinogenesis*, 10, p.2275- 7, 1989.

YANG, S. Q., WANG, C. I., GILLMOR, S. A., FLETTERICK, R. J., CRAIK, C. S. *J. Mol. Biol.* 279, p.94557, 1998.

ZHANG, J., COUSENS, L. S., BARR, P. J. & SPRANG, S. R. Three-dimensional structure of human basic fibroblast growth factor, a structural homolog of interleukin 1 beta. *Proc. Natl Acad. Sci. USA*, 88(8), p.3446–51, 1991.

ZHU, X., KOMIYA, H., CHIRINO, A., FAHAM, S., FOX, G. M., ARAKAWA, T., HSU, B. T. & REES, D. C. Three-dimensional structures of acidic and basic fibroblast growth factors. *Science*, 251(4989), p.90–3, 1991.

SITES INTERESSANTES

<http://delphi.phys.univ-tours.fr/Prolysis/>
<http://delphi.phys.univ-tours.fr/Prolysis/images.html>
<http://merops.sanger.ac.uk/>
<http://www-structmed.cimr.cam.ac.uk/Course/Overview/Overview.html>
http://www-structure.llnl.gov/crystal_lab/cystalmake.html
<http://www-structure.llnl.gov/Xray/xrayequipment.htm>
<http://www-structure.llnl.gov/Xray/101index.html>
<http://didaktik.physik.uni-wuerzburg.de/~pkrahmer/ntnujava/emWave/emWave.html>
<http://didaktik.physik.uni-wuerzburg.de/~pkrahmer/ntnujava/shm/shm.html>
<http://www.chem.pmf.hr/~dijana/procr/main.html>
<http://www.msg.ku.edu/~xraylab/notes/collect.html>
<http://www.yorvic.york.ac.uk/~cowtan/fourier/magic.html>
http://www.lure.u-psud.fr/sections/Xenon/tutorial/tutorial_eng.html
http://www.bmsc.washington.edu/scatter/AS_index.html
<http://www.ncbi.nlm.nih.gov/>
<http://www.rcsb.org/pdb>
<http://www.sanger.ac.uk/Software/Pfam/>
<http://www.ebi.ac.uk/swissprot/>
<http://www.expasy.ch/>
<http://www.ebi.ac.uk/clustalw/>
<http://shelx.uni-ac.gwdg.de/>
<http://www.lnls.br>

ANEXOS

ANEXO 1: ARTIGO DA HEXOQUINASE PII

“The high resolution crystal structure of yeast hexokinase PII with the correct primary sequence provides new insights into its mechanism of action.”

Paula R. Kuser, Sandra Krauchenco, Octávio A. C. Antunes & Igor Polikarpov

Journal of Biological Chemistry, v. 275 n. 27, p. 2081421, jul 2000.

ANEXO 2: ARTIGO DA MOUSE URINARY PROTEIN (MUP)

“Crystallization and preliminary diffraction studies of a recombinant Mouse Urinary Protein.”

Paula R. Kuser, Sandra Krauchenco, Andrea Fangel & Igor Polikarpov

Acta Crystallographica section D v.55, p. 1340-1, 1999.

ANEXO 3: ARTIGO DA CRISTALIZAÇÃO DE CTI

“Crystallization and preliminary x-ray diffraction analysis of a novel trypsin inhibitor from seeds of *Copaifera langsdorffii*.”

Sandra Krauchenco, José A. da Silva, Ronaldo A. P. Nagem, José R. Brandão Neto, Valéria P. Forrer, Renata Carmona e Ferreira, Maria R. L. Macedo, José C. Novello, Sérgio Marangoni & Igor Polikarpov

Acta Crystallographica section D v.57, p. 1316-8, 2002.

ANEXO 4: ARTIGO DA ESTRUTURA CRISTALOGRÁFICA DE DrTI

“Crystal structure of the Kunitz (STI) type inhibitor from *Delonix regia* seeds.”

Sandra Krauchenco, Silvana C. Pando, Sérgio Marangoni & Igor Polikarpov.

Biochemical and Biophysical Research Communications v.312, p.1303-8, Out 2003.

ANEXO 5: ARTIGO DA ESTRUTURA CRISTALOGRÁFICA DE CTI

“Crystal structure of an unusual Kunitz (STI) type trypsin inhibitor from *Copaifera langsdorffii* seeds.”

Sandra Krauchenco, Ronaldo A. P. Nagem, José A. da Silva, Sérgio Marangoni & Igor Polikarpov.

Aceito para publicação na Biochimie do Japão.

The High Resolution Crystal Structure of Yeast Hexokinase PII with the Correct Primary Sequence Provides New Insights into Its Mechanism of Action*

Received for publication, December 27, 1999, and in revised form, March 27, 2000
Published, JBC Papers in Press, April 3, 2000, DOI 10.1074/jbc.M910412199

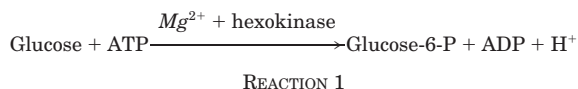
Paula R. Kuser‡, Sandra Krauchenco‡, Octávio A. C. Antunes§, and Igor Polikarpov‡¶

From the ‡Laboratório Nacional de Luz Síncrotron, Campinas, São Paulo, Brazil, Caixa Postal 6192-CEP 13083-970 and §Departamento Química Inorgânica, Instituto de Química, UFRJ, Rio de Janeiro, RJ, Brazil

Hexokinase is the first enzyme in the glycolytic pathway, catalyzing the transfer of a phosphoryl group from ATP to glucose to form glucose 6-phosphate and ADP. Two yeast hexokinase isozymes are known, namely PI and PII. The crystal structure of yeast hexokinase PII from *Saccharomyces cerevisiae* without substrate or competitive inhibitor is determined and refined in a tetragonal crystal form at 2.2-Å resolution. The folding of the peptide chain is very similar to that of *Schistosoma mansoni* and previous yeast hexokinase models despite only 30% sequence identity between them. Distinct differences in conformation are found that account for the absence of glucose in the binding site. Comparison of the current model with *S. mansoni* and yeast hexokinase PI structures both complexed with glucose shows in atomic detail the rigid body domain closure and specific loop movements as glucose binds. A hydrophobic channel formed by strictly conserved hydrophobic residues in the small domain of the hexokinase is identified. The channel's mouth is close to the active site and passes through the small domain to its surface. The possible role of the observed channel in proton transfer is discussed.

Hexokinase (HK),¹ or ATP:D-hexose 6-phosphotransferase, is a member of the kinase family of tissue-specific isoenzymes (EC 2.7.1.1). Reduction in its activity causes illnesses in humans like hemolytic anemia (1, 2) and cardiomyopathy (3). Mutations in glucokinase (hexokinase IV) are associated with early onset non-insulin-dependent diabetes mellitus (4, 5). Hexokinase has been a target for the development of efficient inhibitors in the search for new drugs against diseases caused by trypanosomes (6) and has been used in the construction of biosensors capable of detecting glucose (7).

Hexokinase catalyzes the transfer of an ATP γ -phosphate group to the 6-position on the glucose (Glc) ring as follows.



* This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico Grant 150088-5 and Fundação de Amparo à Pesquisa do Estado de São Paulo Grant 99/03387-4. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed. Tel.: 55 19 287 4520; Fax: 55 19 287 4632; E-mail: igor@lnls.br.

¹ The abbreviations used are: HK, hexokinase; OTG, *ortho*-toluoylglucosamine; r.m.s., root mean square.

Cloning the hexokinase genes from *Saccharomyces cerevisiae* has shown that there are two isoenzymes of hexokinase in yeast: PI and PII, with an overall homology in their amino acid sequences of about 76% (8, 9). Hexokinase PII is the predominant hexose kinase in *S. cerevisiae* grown on glucose (10) and is required for the catabolite repression, by glucose, of the expression of other genes (11–13). The yeast HKs are known to exist as phosphoproteins *in vitro* (14) and *in vivo* (15), with the dimer-monomer equilibrium affected by phosphorylation. The *in vivo* phosphorylation site has been identified as Ser¹⁵ (16). The crystallographic structures of yeast hexokinase PI complexed with glucose (PI-Glc) refined at 3.5-Å resolution (17, 18) and the hexokinase PII from yeast complexed with the competitive inhibitor, *ortho*-toluoylglucosamine (PII-OTG), determined at 2.1-Å resolution (19, 20), have already been reported.

The original structures of yeast hexokinases were determined without a primary amino acid sequence, which did not become available until 1985 (8, 9). The identity of the side chains was deduced from inspection of the electron density. In fact, only a 30% identity can be found between the primary sequence of the crystallographic models and the one obtained from cDNA sequencing.

In an attempt to define the structure and function of this enzyme more clearly, we solved the crystal structure of yeast hexokinase PII to 2.2-Å resolution, refining the model with the correct amino acid sequence. This is the first high resolution hexokinase structure solved without a bound substrate or competitive inhibitor. We compare it with the structures of human hexokinase determined to 2.8-Å resolution (21–23), *Schistosoma mansoni* hexokinase also solved to 2.8-Å resolution (24), and previous structures of yeast hexokinase PI and PII (17–20).

The primary sequence of yeast hexokinase PII is compared with the amino acid sequences of other hexokinases, and a possible role for the conserved amino acids is proposed. Glucose and ATP binding sites and a possible reaction mechanism are discussed. We demonstrate the existence of a channel, formed by the conserved hydrophobic residues Ile⁸⁵, Leu⁸⁷, Leu¹²⁷, Ile¹³¹, Leu¹³⁵, Met¹³⁹, Leu¹⁵³, Phe¹⁵⁵, Phe¹⁵⁷, Phe¹⁷⁸, Leu¹⁹², and Ile¹⁹⁶, which passes from the active site of the enzyme through the small domain to the exterior of the protein. Its possible role as a proton sink is discussed.

EXPERIMENTAL PROCEDURES

Crystallization—The protein material used for crystallization was purchased from Sigma and further purified using an anion exchange column. Crystallization was carried out using the sitting-drop vapor diffusion method. The drops were composed of 3 μ l of protein solution and 3 μ l of crystallization solution. The volume of the reservoir was 1 ml. Crystal screenings were carried out at room temperature, varying the concentration of precipitant from 1.6 to 2.4 M ammonium sulfate in 100 mM potassium phosphate buffer. Needles were observed under several conditions, and some data were collected for these crystals,

TABLE I
Statistics of data collection and refinement

Parameters	Values
Data collection	
Resolution limit of data (Å)	2.2
Number of measurements	53,732
Number of measurements > 1σ	48,332
Number of unique reflections	28,523
Completeness of data (last shell) (%)	95.1 (96.8)
R_{sym}^a	11.3
Refinement	
Resolution range for refinement (Å)	13.00–2.2 Å
Total number of atoms	3671
Total number of solvent molecules	442
R -factor ^b	16.27
R_{free}^c	24.60
Average B factor	
All atoms (Å ²)	23.41
Solvents (Å ²)	43.41
r.m.s. deviations from ideal geometry	
bonds (Å)	0.011
bond angles (degrees)	2.1
Ramachandran plot statistics	
Residues in core Ramachandran (%)	90.5
Residues in disallowed regions (%)	0.0

^a $R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I \times 100$.

^b Crystallographic R -factor = $\sum (|F_{\text{obs}}| - |F_{\text{calc}}|) / \sum F_{\text{obs}}$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes respectively.

^c R_{free} is the crystallographic R -Factor calculated for a subset of randomly selected reflections (5%) not used in the phasing process.

diffracting to 2.8-Å resolution. A single crystal, from which diffraction data were collected, had an elongated bipiramidal shape, unlike the more usual thin needles. Its dimensions were $0.25 \times 0.4 \times 1.0$ mm.

X-ray Diffraction Data—Diffraction data to 2.2 Å were collected at room temperature (+27 °C) on a MAR image plate system at the dedicated protein crystallography beamline of the Brazilian synchrotron light source (25, 26) and processed with the programs DENZO and SCALEPACK (27). Data were collected on frames of 1.0° rotation with the same amount of x-ray dose per frame. The crystal-to-detector distance was 200 mm. A total of 53,732 reflections were measured and merged to 24,924 unique reflections, with a completeness of 94.9% for all data to 2.2 Å (Table I). The crystal belonged to tetragonal space group I4, with cell constants of $a = b = 142.81$ Å and $c = 58.46$ Å. Estimation of the solvent content indicated that the crystal contained one molecule per asymmetric unit with $V_m = 2.7$ Å Da⁻¹. The structure was solved using Patterson search techniques. Rotational and translational searches were performed with AMoRe (28) using data to 2.2 Å and a P152K mutant of yeast hexokinase (29) as the starting model. A unique solution was found after rigid body refinement, with a correlation coefficient of 53.5% and an R -factor of 44.8%.

Refinement and Model Building—Repeated cycles of rebuilding in O (30) and refinement using REFMAC (62) with insertion of water molecules using ARPP (32) lowered the R -factor to a final value of 16.3% and R -free (33), calculated for 5% randomly chosen reflections that were not included in the refinement, to 24.6%. The root mean square (r.m.s.) deviation of bond lengths was 0.011 Å, and r.m.s. deviation of bond angles was 2.1°. A summary of the refinement statistics, together with other crystallographic and final model parameters, is given in Table I. Analysis of the final model using PROCHECK (40) indicates that 90.5% of the residues fall within “core” regions of the Ramachandran plot (34, 35). No residues are found in disallowed regions. The G -factor of the model is -0.06 . The final model contains residues 18–486 and 442 water molecules. No continuous density can be observed for the first 17 residues of the N terminus. Yeast HK is susceptible to limited proteolysis by endogenous proteases, resulting in loss of the first 11 N-terminal amino acid residues (36, 37), which, together with the thermal disorder, is probably the reason for the lack of electron density at the N terminus.

The average thermal parameter for the model is 23.41 Å², with the highest B value of 72.44 Å² for the side chain of residue Asp¹¹⁶, which is located in a loop between β -strand β_4 and α -helix α_3 . The final $(2F_o - F_c)$ electron density, calculated using data to 2.2 Å, is well defined for most residues.

RESULTS AND DISCUSSION

Secondary and Tertiary Structure—The molecule has a palm shape, with approximate dimensions of $59 \times 78 \times 54$ Å³, and is in its open conformation. It has the same α/β fold observed in other hexokinase structures (17–24). The polypeptide chain of 486 residues is distinctly folded into two domains of unequal size: the large and small domains. These are separated by a deep cleft containing the residues making up the enzyme active site.

The secondary structure was analyzed using the PROMOTIF (39) program. In all, there are 14 α -helices, 13 β -strands, and three 3_{10} -helices (Fig. 1). The large domain (residues 13–76 and 212–457) comprises a six-stranded mixed β -sheet (β_1 , β_9 , β_{10} , β_{11} , β_{12} , and β_{13}) and a number of additional α -helices. The mixed sheet has $-3,1,1,-3x,-1x$ topology in the Richardson notation (38). On one side, the sheet packs against the small domain, and on the other side it is shielded by several α -helices. The domain also contains a particularly long helix, α_{11} (36 residues and 53.35 Å in length), which exhibits a slight bend.

The small domain comprises residues 77–211 and 458–486, and its dominant feature is a five-stranded mixed β -sheet (β_2 , β_3 , β_4 , β_5 , and β_8), with a sheet topology of $-1,-1,3x,1x$. The sheet is flanked by two helices on one side and by one helix on the other. The domain also has an additional β -sheet formed by two antiparallel strands (β_6 and β_7).

Two regions of distortion were identified in the β -sheets, both classified as classic parallel β -bulges, one involving β -strands β_9 and β_{12} and the other involving β_5 and β_8 . Five β -hairpins are also present in the structure, between antiparallel β -strands $\beta_2:\beta_3$, $\beta_3:\beta_4$, $\beta_6:\beta_7$, $\beta_9:\beta_{10}$, and $\beta_{10}:\beta_{11}$.

All these secondary structure elements match those present in the human HK structure apart from three additional 3_{10} -helices (α_5 , 245–251, α_5 , 269–273; α_{12} , 445–449) found in the yeast HK structure. The structure of *S. mansoni* HK shows the same secondary structural pattern. Direct comparison with the previous HK structures (17–20) can only be somewhat approximate due to the limited refinement and the uncertain amino acid composition of those earlier models. However, such an approximate comparison suggests that most of the secondary structural elements in the current structure have their equivalents in the previous ones.

The superposition of the yeast and *S. mansoni* hexokinase structures (Fig. 2) clearly shows the *open conformation* of the polipeptide chain (17–20).

Conserved Amino Acid Residues—The primary sequence alignment of a selection of proteins from the hexokinase family is shown in Fig. 3. This clearly demonstrates extensive similarity between the N- and C-terminal halves of type I human hexokinase, rat hexokinase, and hexokinase from *S. mansoni* and between these and yeast hexokinase, consistent with the gene duplication-fusion concept proposed by Colowick (36). Approximately 34% of the amino acid residues are conserved in all members of the family, and 13% are perfect matches. The strong conservation of these residues implies their relevance to biological function. Various strictly conserved amino acid residues are present in the binding site. A high number of glycine residues are also conserved (Gly⁷⁶, Gly⁸⁰, Gly⁸⁸, Gly⁸⁹, Gly¹⁵⁴, Gly²³³, Gly²³⁵, Gly²⁹⁷, Gly³⁰⁷, Gly⁴¹⁸, and Gly⁴⁶¹). These residues are located at the ends of β -strands or α -helices, changing the direction of the chain. Their conservation probably gives the hexokinase molecule the flexibility necessary for binding glucose and ATP. Some of them are directly involved in the active site formation and are essential for enzymatic activity (see discussion below).

Comparison of the HK Structures and the Conformational Changes—Fig. 2a shows a comparison of yeast hexokinase PII

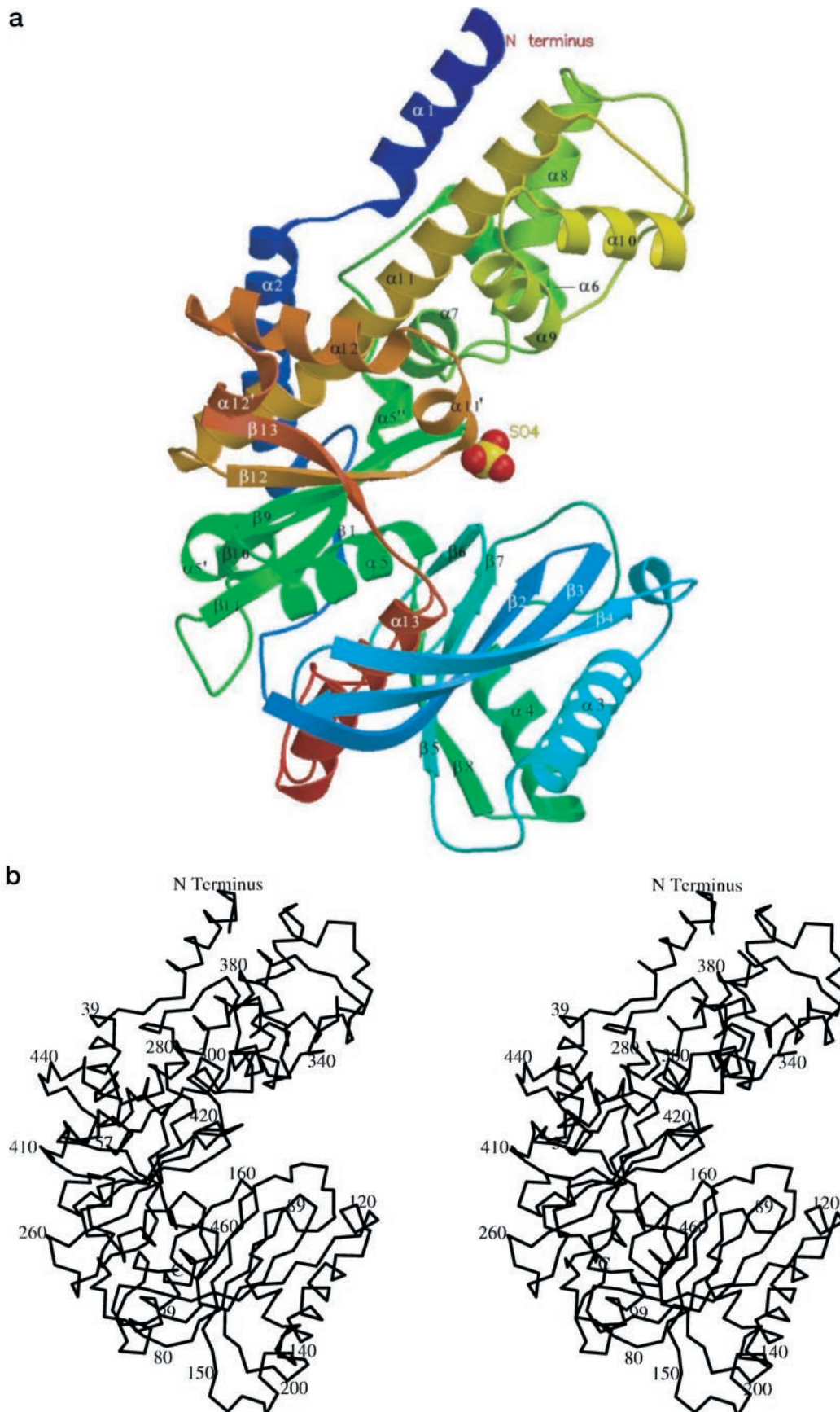


FIG. 1. **Schematic diagram of the yeast hexokinase PII model (a) and stereo C- α trace (b).** The chain is color-coded from *blue* (N terminus) to *red* (C terminus). This model is in an open conformation, and the active site is located between the two domains. Secondary structure assignments are made in accordance with the method of Kabsh and Sander (60). This figure was generated using MOLSCRIPT (61).

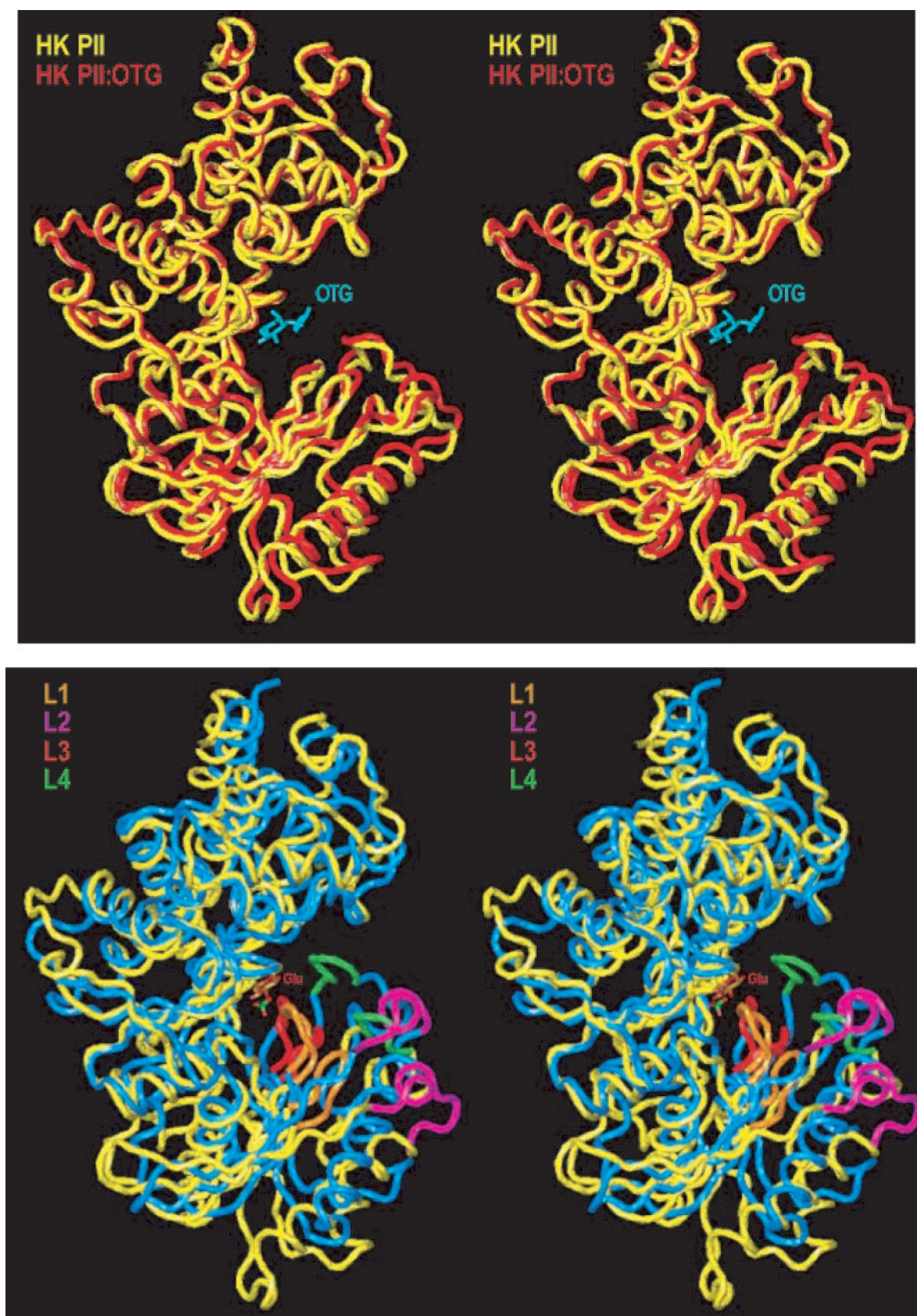


FIG. 2. *a*, stereoview C- α overlay of refined yeast hexokinase PII (yellow) and yeast hexokinase PII-OTG (red). Subtle movements of the chain in the small domain indicate a slightly more closed conformation of the PII-OTG model. *b*, stereoview C- α overlay of refined yeast hexokinase PII (yellow) and *S. mansoni* hexokinase (blue). The major structural differences are the closure of the active site and the bending of loops L1, L2, L3, and L4. Superposition was performed using alignment of residues from the large domain.

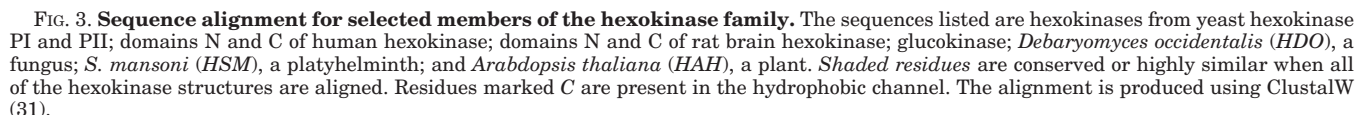
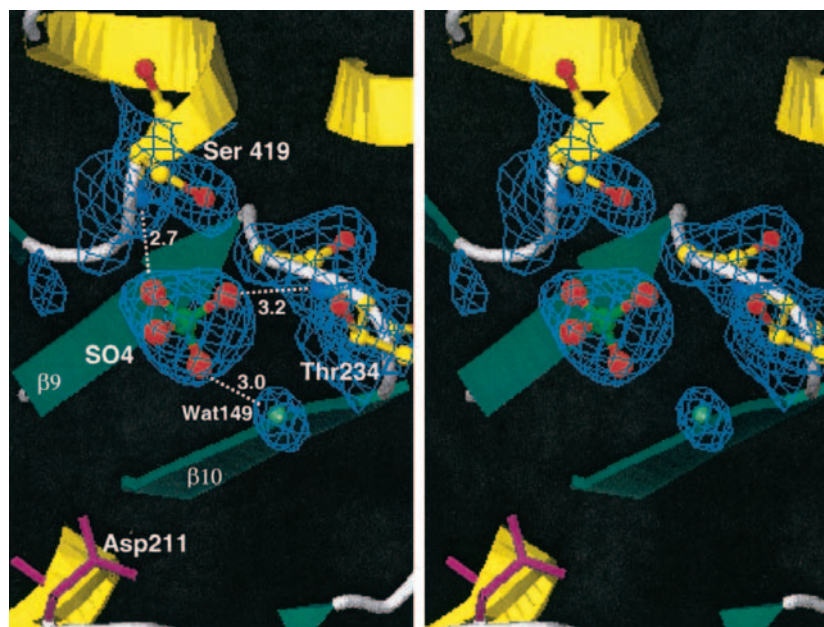


FIG. 4. The sulfate ion bound to the active site of yeast hexokinase. A sulfate molecule, hydrogen-bonded to residues Thr²³⁴ and Ser⁴¹⁹ and to water molecule W149 is shown as a ball-and-stick model. $2F_o - F_c$ density for the region around the sulfate anion is shown in blue. The sulfate molecule is located at the entrance to the active site.



with the previously determined PII-OTG model. An r.m.s. deviation calculated using the LSQMAN program (41) for 452 C- α atom pairs is 0.93 Å. The main difference is that the present structure adopts a slightly more open conformation, suggesting that OTG binding induces a slight closure of the cleft between two domains.

A comparison with hexokinase from *S. mansoni*, determined as a complex with glucose, reveals a movement of the domains closing the cleft (Fig. 2b). The overall r.m.s. deviation, calculated for 414 C- α atom pairs is 4.7 Å. However, if the two domains are considered separately, the superposition of the large domains gives an r.m.s. deviation of 1.8 Å for 320 C- α atom pairs, while the superposition of the small domains gives an r.m.s. deviation of 2.2 Å for 261 C- α atom pairs. Similar changes in conformation have been observed in the yeast PI-Glc complex (17) with an r.m.s. deviation of 8.1 Å for 409 C- α atom pairs.

When the large domains of these molecules are superposed, it becomes clear that, in addition to the rigid body closure of the small domain, four peptide segments, namely residues 87–92 (L1), 115–124 (L2), 158–163 (L3), and 174–178 (L4) (yeast hexokinase PII numbering), move forward to embrace the binding site. These loops move by as much as 8 Å measured at C- α atom positions. The movement of these loops seems to be functionally important to complete the formation of the glucose binding site and to preform the nucleotide binding site. Two of the loops, namely L1 (⁸⁷LGGTN⁹¹) and L4 (¹⁷⁴WTKGF¹⁷⁸), are composed of amino acid residues conserved in all but two of the hexokinase sequences listed in Fig. 3. In the N-terminal regulatory domains of rat and human hexokinases, residues Thr⁹⁰ and Asn⁹¹ of the loop L1 are substituted by serines, whereas the residue Gly¹⁷⁷ of the loop L4 is substituted by an arginine. These substitutions might be directly related to the lack of enzymatic activity of these domains.

In addition to the significant conformational changes of the loops involved in glucose and ATP binding, differences in the conformations of the external loops are also observed, mainly in the positions of insertions and deletions in respective amino acid sequences.

Comparison of the current structure with the models of hexokinase complexed with substrate provides precise information about the conformational changes this enzyme undergoes upon substrate binding and confirms the “induced fit” mechanism

theory (42) and previous observations that initial recognition of glucose by HK is followed by closure of the cleft altering the enzyme-substrate interaction (43).

The Glucose Binding Site—The glucose binding site on our model is in close agreement with the binding site as previously elucidated (17–20). Asp²¹¹, identified as a catalytic base, makes a hydrogen bond with the hydroxyl at position 6 of glucose in both open and closed forms of HK, whereas side chains of the amino acid residues Asn²³⁷, Glu²⁶⁹, and Glu³⁰² are at hydrogen bonding distances from the O-4, O-3, and O-1 atoms of glucose, respectively. Initial glucose recognition followed by the closure of the cleft between the two HK domains brings amino acid residues of the small domain, notably Thr¹⁷² and Lys¹⁷³, within hydrogen interaction distances of the substrate.

Crystallographic structures clearly show that in the closed conformation Ser¹⁵⁸ interacts with hydroxyl group 3 of the glucose molecule via a carboxyl oxygen (distance 3.22 Å in the *S. mansoni* hexokinase structure), which confirms the results of mutational studies (44); mutation of Ser¹⁵⁸ to Ala impairs HK catalytic activity, implying functional significance of the hydrogen bonding via the side chain. Ser¹⁵⁸ is also strictly conserved among HK sequences. It has also been shown that the binding of certain types of glucose inhibitors, such as D-xylitol or D-xylose, promotes the short lived hydrolytic activity of the enzyme followed by inactivation of the enzyme via autophosphorylation (36, 45, 46). The autophosphorylation site has recently been identified as Ser¹⁵⁸ (47). Taken together, these observations support the idea that additional conformational changes might occur upon binding of ATP (48, 49) and that Ser¹⁵⁸ might play an important role in the process of the phosphate transfer.

As observed by Steitz *et al.* (17–20), in the open conformation OTG forms few hydrogen bonds or Van der Waals contacts with the amino acid residues of the small domain. Projection of OTG into the *S. mansoni* model shows that the glucosamine position is very close to the site occupied by glucose, but its toluoyl group prevents complete domain closure. However, there is room for a significantly larger closure of the HK domains than the comparison of yeast hexokinase PII and PII-OTG models reveals. Taking this into account, we believe that the steric clashes of the OTG toluoyl group with the amino acid residues of the small domain prevent formation of an energetically favorable closed conformation of PII-OTG complex. Recent scan-

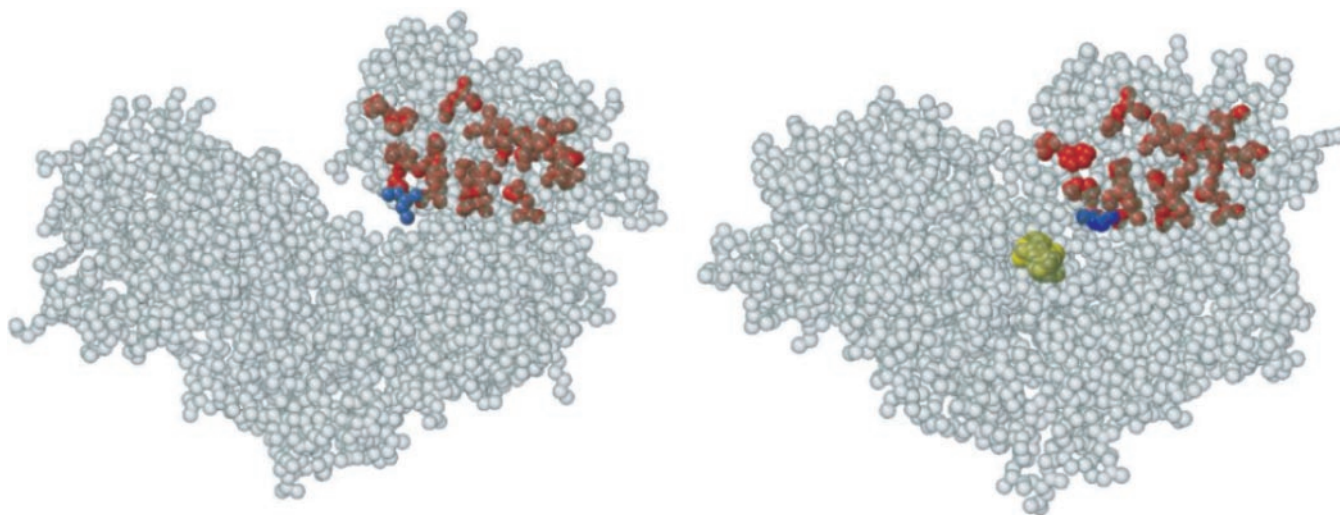


FIG. 5. **The conserved hydrophobic channel.** Amino acids forming the hydrophobic channel are displayed in red, and the residue in blue is serine 158. The channel is also highlighted in the *S. mansoni* hexokinase to show that in the closed conformation the channel end is close to the active site. The glucose molecule is shown in yellow.

ning calorimetry measurements of the D-glucose binding to HK have demonstrated that this process is entropy-driven and that the binding enthalpy is zero (50). The OTG toluoyl group interactions may make it impossible to trap the hexokinase in the closed conformational state, with the open conformation being thermodynamically more favorable.

The ATP Binding Site—Crystallographic analysis of a binary complex of yeast HK with 8-bromo-AMP and modeling of ATP into the active site of the enzyme (49) indicated that amino acid residues 344–348 and 422–424 interact with the adenine ring of ATP. Comparison of the three-dimensional structures of the 44-kDa ATPase fragment of 70-kDa heat shock cognate (HSC70) protein (51), of actin (52), and of HK led to the identification of an ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and HSC70 proteins (53). The ATP binding pattern was defined as consisting of five sequence motifs: phosphate 1 (residues 82–103), connect 1 (residues 203–223), phosphate 2 (residues 229–248), adenosine (residues 411–439), and connect 2 (residues 453–473) with a number of Gly, Ser, Thr, and Asp residues being strictly conserved between all of the compared sequences.

The present model of yeast hexokinase was determined in the absence of any substrate in the active site. However, in the last stages of refinement a ($F_{\text{obs}} - F_{\text{calc}}$) difference map exhibited pronounced electron density within the catalytic cleft, and this density was modeled as a sulfate molecule. Sulfate was present in 2 M concentration in the crystallization conditions. Its well defined density is located inside the cleft, close to the position of a sulfate ion observed in the *S. mansoni* hexokinase and PI-Glc models. The sulfate ion was inserted into the model and refined with an occupancy of 1.0. The thermal parameters are comparable with the average thermal parameter for solvent molecules and similar to the B value of water 149 to which the sulfate ion is bound. The sulfate group makes a hydrogen bond to the main chain nitrogens of residues Ser⁴¹⁹ and Thr²³⁴ (Fig. 4). Remarkably, these are the strictly conserved residues from the phosphate 2 (Thr²³⁴) and adenosine (Ser⁴¹⁹) ATPase pattern recognition motifs (54). Based on these motifs, the ATP position was modeled, by superposing the actin-ATP model onto the present HK structure. Contrary to the suggestion in Ref. 19, the sulfate ion does not occupy the position of the β - or γ -phosphate. Rather, it occupies a position close to the α -phosphate of ATP. If ATP does indeed bind to HK in the modeled position and conformation, its γ -phosphate will be 3.7 Å from

the hydroxyl-6 of glucose. Interestingly, the β -phosphate interacts only with amino acid residues of the small domain. This means that after the enzymatic reaction has taken place and the HK cleft opens, the small domain will drag ADP away from the active site, opening the way to release Glc-6-P. This is consistent with the view that glucose binds first and then ATP, whereas ADP is released first, followed by Glc-6-P (36).

The presence of a sulfate ion in the open conformation of hexokinase proves that it does not provoke significant conformational change upon binding to the enzyme. The fact that the same sulfate/phosphate anion binding site was repeatedly found in yeast (17–20), human (21–23), and *S. mansoni* (24) hexokinase models shows that the HK ATP binding site is able to bind monophosphates. This may have functional importance. It is known, for example, that glucose binding is strongly promoted in the presence of 0.05 M phosphate (54). Phosphate binding might therefore somehow restrict the conformations of the amino acid residues in the glucose binding site, facilitating the binding of glucose.

The Conserved Hydrophobic Channel—Most of the strictly conserved amino acid residues of hexokinase, when viewed in a space-filling representation, appear at the cleft between the two domains and form the glucose and ATP binding sites. However, a number of hydrophobic residues belonging to the small domain form a channel that begins close to the active site, goes through the entire small domain, and ends up at the surface of the protein approximately 30 Å away from the Glc/ATP binding site (Fig. 5). These residues are Ile⁸⁵, Leu⁸⁷, Leu¹²⁷, Ile¹³¹, Leu¹³⁵, Met¹³⁹, Leu¹⁵³, Phe¹⁵⁵, Phe¹⁵⁷, Phe¹⁷⁸, Leu¹⁹², and Ile¹⁹⁶. The cross-section of the channel has an elongated shape and is about 2.5–3 Å wide. The length of the channel is approximately 25 Å. The surface charges at its entrance are predominantly negative.

What might be the function of this channel? We speculate that it might act as a tunnel for the proton generated in the glucose phosphorylation reaction. The channel is too small for the transport of any other substrate or product of reaction (Glc, ATP, ADP, or Glc-6-P) or an inorganic phosphate. The fact that it is connected to the active site area and is lined by strictly conserved residues suggests its importance in the function of the enzyme.

It is clear that glucose binding to HK induces substantial conformational changes. Loops, formed by amino acid residues 87–92, 115–124, 158–163, and 174–178, forming the mouth of

the conserved hydrophobic channel, close up over the active site, bringing the entrance to the channel into close proximity with the ligand binding sites. The closed active site conformation is probably completed after additional conformational changes that accompany ATP binding. Nucleophilic attack of the glucose's 6-hydroxyl group by the γ -phosphate of ATP is promoted by Asp²¹¹ and Lys¹⁷³ and is followed by "in-line" transfer of the γ -phosphate group to glucose, presumably through the dissociative transitional state (55) and liberation of a proton.

How does release of the reaction products occur? The driving force of product release is probably the repulsion between Glc-6-P and the β -phosphate of ADP. Since the β -phosphate is bound to residues of the small domain, its repulsion from the γ -phosphate, now bound to glucose, would drive the HK open. Repulsion of these two phosphates will depend on the local proton concentration. Low active site local pH would shield electrostatic repulsion between the negative charges of the phosphates and impede release of ADP and Glc-6-P from the HK active site. Proton transfer through the hydrophobic channel might provide the means for decreasing the concentration of protons in the active site cavity, increasing ADP and Glc-6-P mutual repulsion and facilitating the release of the reaction products. In a way, the putative channel function might be similar and opposite to that of the ATPase synthase proton channel (56–59).

If this hypothesis is valid, the disruption of the channel should impede the reaction product release, whereas the inversion of the proton flux through the channel would change the reaction balance increasing the phosphate transferase activity of HK. In this situation, ATP and glucose should be produced from ADP and Glc-6-P by HK. Site-directed mutagenesis of the strictly conserved hydrophobic amino acids forming the channel will probably tell whether our hypothesis is correct and will shed light on the nature, purpose, and function of the hydrophobic channel.

Acknowledgments—We thank Prof. Maria Lucia Bianconi and Prof. Jose Abrahão Neto for the kind gift of some of the enzyme and Alexander Golubev, Ricardo Aparício, and Elisabete de Souza for help with crystallization. We are indebted to Dr. Roman Laskowski, Dr. Bill Boys, and Dr. Stefan Kycia for critically reading the manuscript and correcting the English.

REFERENCES

- Valentine, W. N., Oski, F. A., Paglia, D. E., Baughan, M. A., Schneider, A. S., and Naiman, J. L. (1967) *N. Engl. J. Med.* **276**, 1–11
- Magnani, M., Stocchi, V., Cucchiaroni, L., Novelli, G., Lodi, S., Isa, L., and Fornaini G. (1985) *Blood* **66**, 690–697
- Barrie, S. E., Saad, E. A., Ubatuba, S., Da Silva Lacaz, P., and Harris, P. (1979) *Res. Commun. Chem. Pathol. Pharmacol.* **23**, 375–381
- Vionnet, N., Stoffel, M., Takeda, J., Yasuda, K., Bell, G. I., Zouali, H., Lesage, S., Velho, G., Iris, F., Passa, P., Froguel, P., and Cohen, D. (1992) *Nature* **356**, 721–722
- Gupta, B. L., Nehal, M., and Baquer, N. Z. (1997) *Indian J. Exp. Biol.* **35**, 792–795
- Willson, M., Alric, I., Perie, J., and Sanejouand, Y. H. (1997) *J. Enzyme Inhibition* **12**, 101–121
- Cheng, Q., and Stevens, R. C. (1997) *Adv. Mater.* **9**, 481–483
- Kopetzki, E., Entian, K., and Mecke, D. (1985) *Gene (Amst.)* **39**, 95–102
- Fröhlich, K., Entian, K., and Mecke, D. (1985) *Gene (Amst.)* **36**, 105–111
- Gancedo, J. M., Clifton, D., and Fraenkel, D. G. (1977) *J. Biol. Chem.* **252**, 4443–4444
- Entian, K.-D. (1980) *Mol. Gen. Genet.* **178**, 633–637
- Rose, M., Albige, W., and Entian, K.-D. (1991) *Eur. J. Biochem.* **199**, 511–518
- De Winder, J. H., Crauwels, M., Hohmann, S., Thevelein, J. M., and Winderickx, J. (1996) *Eur. J. Biochem.* **241**, 633–643
- Fernandez, R., Herrero, P., Fernandez, M. T., and Moreno, F. (1986) *J. Gen. Microbiol.* **132**, 3467–3472
- Vojtek, A. B., and Fraenkel, D. G. (1990) *Eur. J. Biochem.* **190**, 371–375
- Behlke, J., Heidrich, K., Naumann, M., Muller, E.-C., Otto, A., Reuter, R., and Kriegel, T. (1998) *Biochemistry* **37**, 11989–11999
- Steitz, T. A., Fletterick, R. J., Anderson, W. F., and Anderson, Ch. M. (1976) *J. Mol. Biol.* **104**, 197–222
- Bennet, W. S., Jr., and Steitz, T. A. (1980) *J. Mol. Biol.* **140**, 183–209
- Anderson, C. M., Stenkamp, R. E., McDonald, R. C., and Steitz, T. A. (1978) *J. Mol. Biol.* **123**, 15–34
- Anderson, C. M., Zucker, F. H., and Steitz, T. A. (1979) *Science* **204**, 375–380
- Aleshin, A. E., Zeng, C., Bartunik, H. D., Fromm, H. J., and Honzatko, R. B. (1998) *J. Mol. Biol.* **282**, 345–357
- Aleshin, A. E., Zeng, C., Bourenkov, G. P., Bartunik, H. D., Fromm, H. J., and Honzatko, R. B. (1998) *Structure* **6**, 39–50
- Aleshin, A. E., Fromm, H. J., and Honzatko, R. B. (1998) *FEBS Lett.* **434**, 42–46
- Mulichack, A. M., Wilson, J. E., Padmanabhan, K., and Garavito, R. M. (1998) *Nat. Struct. Biol.* **5**, 555–560
- Polikarpov, I., Perles, L. A., de Oliveira, R. T., Oliva, G., Castellano, E. E., Garrat, R. C., and Craievich A. (1998) *J. Synchrotron Rad.* **5**, 72–76
- Polikarpov, I., Oliva, G., Castellano, E. E., Garrat, R. C., Arruda, P., Leite, A., and Craievich A. (1998) *Nuclear Instruments and Methods* **405**, 159–164
- Otwinowski Z. (1993) in *Proceedings of the CCP4 Study Weekend: Data collection and Processing* (Sawyer, L., Isaacs, N., and Bailey, S., eds) pp. 56–62, Daresbury Laboratory, Warrington, United Kingdom
- Navaza, J. (1994) *Acta Crystallogr. Sec. A* **50**, 157–163
- Polikarpov, I., Reibner, C., Beecken, V., Jacob, L., Rose, M., Entian, K.-D., and Bartunik, H. D. (1993) *HASYLAB Annual Report*, pp. 777–784, HASYLAB, Hamburg, Germany
- Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sec. A* **47**, 110–119
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680
- Lamzin, V. S., and Wilson, K. S. (1997) *Methods Enzymol.* **277**, 269–305
- Brünger, A. T. (1992) *Nature* **355**, 472–475
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283–291
- Ramachandran, N. G., and Sasisekharan, V. (1968) *Adv. Protein Chem.* **23**, 283–437
- Kleywegt, G. J., and Jones T. A. (1996) *Structure* **4**, 1395–1400
- Colowick, S. P. (1973) in *The Enzymes* (Boyer, P. D., ed) Vol. 9, pp. 1–48, Academic Press, Inc., New York
- Schmidt, J. J., and Colowick, S. P. (1969) *Arch. Biochem. Biophys.* **158**, 471–477
- Richardson, J. S. (1981) *Adv. Protein. Chem.* **34**, 167–339
- Hutchinson, E. G., and Thornton J. M. (1996) *Protein Sci.* **5**, 212–220
- Kleywegt, G. J. (1996) *Acta Crystallogr. Sec. D* **52**, 842–857
- Koshland, D. E. (1959) in *The Enzymes* (Boyer, P. D., Lardy, H. and Myrbäck, K., eds.) 2nd Ed., Vol. 1, pp. 305–346, Academic Press, Inc., New York
- Wilson, J. E. (1995) *Rev. Physiol. Biochem. Pharmacol.* **126**, 65–198
- Arora, K. K., Filburn, C. R., Pedersen, P. L. (1991) *J. Biol. Chem.* **266**, 5359–5362
- DelaFuente, G., and Sols, A. (1970) *Eur. J. Biochem.* **16**, 234–239
- DelaFuente, G. (1970) *Eur. J. Biochem.* **16**, 240–243
- Heidrich, K., Otto, A., Behlke, J., Rush, J., Wenzel, K. W., and Kriegel, T. (1997) *Biochemistry* **36**, 1960–1964
- Shoham, M., and Steitz, T. A. (1980) *J. Mol. Biol.* **140**, 1–14
- Shoham, M., and Steitz, T. A. (1980) *Biochim. Biophys. Acta* **705**, 380–384
- Catanzaro, F., Ganbuti, A., Graziano, G., and Barobe, G. (1997) *J. Biochem. (Tokyo)* **121**, 568–577
- Flaherty, K. M., DeLuca-Flaherty, C., McKay, D. B. (1990) *Nature* **346**, 623–628
- Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F., Holmes, K. C. (1990) *Nature* **347**, 37–44
- Bork, P., Sander, C., and Valencia, A. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 7290–7294
- Gazith, J., Schulze, I. T., Gooding, R. H., Womack, F. C., and Colowick, S. P. (1968) *Ann. N.Y. Acad. Sci.* **151**, 307–331
- Jones, J. P., Weiss, P. M., and Cleland, W. W. (1991) *Biochemistry* **30**, 3634–3639
- Futai, M., Noumi, T., and Maeda, M. (1989) *Annu. Rev. Biochem.* **58**, 111–136
- Senior, A. E. (1988) *Physiol. Rev.* **68**, 177–231
- Lewis, M. L., Chang, J. A., and Simoni R. D. (1990) *J. Biol. Chem.* **265**, 10541–10550
- Cain, B. D., Simoni, R. D. (1989) *J. Biol. Chem.* **264**, 3292–3300
- Kabsch, W., and Sander, C. (1983) *Biopolymers* **22**, 2577–2637
- Kraulis, J. (1991) *J. Appl. Crystallogr.* **24**, 946–950
- Murshudor, G. N., Vagin, A. A., Dodson, E. J. (1997). *Acta Crystallogr. Sec. D* **50**, 240–255

Crystallization and preliminary X-ray diffraction analysis of a novel trypsin inhibitor from seeds of *Copaifera langsdorffii*

S. Krauchenco,^a J. A. Silva,^b
R. A. P. Nagem,^a J. R. Brandão
Neto,^a V. P. Forrer,^a R. Carmona
e Ferreira,^a M. L. R. Macedo,^c
J. C. Novello,^b S. Marangoni^b and
I. Polikarpov^{a*}

^aLaboratório Nacional de Luz Síncrotron, Campinas, SP, Brazil, ^bDepartamento de Bioquímica, Instituto de Biologia, Unicamp, Campinas, SP, Brazil, and ^cDepartamento de Ciências Naturais, CEUL, Universidade Federal de Mato Grosso do Sul, MS, Brazil

Correspondence e-mail: igor@lnls.br

A novel trypsin inhibitor isolated from seeds of *Copaifera langsdorffii* was purified to homogeneity and crystallized. Crystals suitable for X-ray analysis were grown using the hanging-drop vapour-diffusion method at 291 K in sodium acetate buffer at pH values near 4.3 using PEG 4000 as precipitant. The crystals presented symmetry compatible with the space group $P4_12_12$ or $P4_32_12$, with unit-cell parameters $a = b = 58.71$, $c = 93.75$ Å, and diffracted to 1.83 Å resolution at the synchrotron source.

Received 15 February 2001
Accepted 3 July 2001

1. Introduction

Proteinase inhibitors may be defined as proteins capable of strongly inhibiting hydrolytic enzymes both *in vitro* and *in vivo* by forming stoichiometric and stable complexes (Breddam *et al.*, 1991; Bode & Huber, 1991). They are present in multiple forms in numerous tissues and fluids of plants, animals and microorganisms (Laskowski & Kato, 1980; Richardson, 1991). Of the inhibitors presents in plants, many are active against exogenous rather than endogenous enzymes, suggesting that they play a role in plant defence, conferring a broad spectrum of resistance to pests and pathogens (Bowles, 1990; Broadway & Duffey, 1986, 1988; Ryan, 1990; Christeller *et al.*, 1992; Shewry & Lucas, 1997; Vigers *et al.*, 1991).

Animal tests and medical experiments have shown that proteinase inhibitors of certain types are anticarcinogenic (Yavelow *et al.*, 1983; Troll *et al.*, 1987; Troll & Kennedy, 1989; Kennedy, 1994; Liener, 1995). The anticarcinogenic properties include the ability to reduce oxygen-radical formation (Yavelow *et al.*, 1982; Frenkel *et al.*, 1987), to suppress the growth of chemical-induced colon and anal gland tumours in rats (Billings *et al.*, 1990), breast tumours in rats and humans (Troll *et al.*, 1980; Tamir *et al.*, 1990) and lung tumours in mice (Witschi & Kennedy, 1989), to suppress chemical- or radiation-induced cell transformation (Billings *et al.*, 1987, 1989) and to reduce spontaneous chromosome abnormality (Afzal *et al.*, 1989). Epidemiological studies suggest that human populations which are known to have high concentration of certain proteinase inhibitors, mainly Bowman–Birk and Kunitz families, in their diet have lower rates of colon, breast, prostate and skin cancers (Correa, 1981; Kennedy, 1998; Chen *et al.*, 1992).

Proteinase inhibitors are divided according to the class of the enzyme they inhibit. Sequencing and X-ray crystallographic studies have shown that the inhibitors of serine proteinases can be further subdivided into several families. The major criteria for establishing a family are extensive homology among its members, topological relationships between the disulfide bridges and location of the reactive site (Laskowski & Kato, 1980; Weder, 1992; Ryan, 1990; Bowles, 1990; Bode & Huber, 1992; Birk, 1985, 1994). Of the many types of trypsin inhibitors, the most important are the Kunitz and the Bowman–Birk proteinase inhibitor families. The Bowman–Birk inhibitors have molecular weights of 8–10 kDa and usually possess seven disulfide bridges. The Kunitz-type inhibitors usually have a molecular weight of about 20 kDa and only two disulfide bridges.

An unusual proteinase inhibitor forming a heterodimer of two non-covalently linked polypeptide chains with molecular weights of 11 and 9 kDa was extracted from the seeds of the *C. langsdorffii* tree (Leguminosae, Caesalpinioideae). Here, we present the results of purification, N-terminal characterization, crystallization and preliminary X-ray diffraction studies of this inhibitor.

2. Experimental

2.1. Protein extraction and purification

The whole seeds were ground in an electric mill and the heavily pigmented fragments of the tegument were separated by suspension in chloroform [1:3(w/v)]. A crude inhibitor preparation was obtained by extraction with 10 mM potassium phosphate buffer pH 7.6 [1:4(w/v)] for 4 h at 277 K. Centrifugation of the extracts was performed at 10 000 rev min⁻¹ at 277 K for 30 min and was followed by

ammonium sulfate precipitation. The precipitate was dialyzed for 24 h at 277 K against distilled water and freeze-dried. All the following purification steps were performed at room temperature.

The precipitate was dissolved in 50 mM Tris-HCl buffer pH 8.0 and applied to a Sepharose DEAE column (1.6 × 19 cm) equilibrated with 50 mM Tris-HCl buffer pH 8.0. After 200 ml of buffer elution, a column was eluted with a concentration gradient of sodium chloride from 0 to 0.5 M. The flow rate was 30 ml h⁻¹, the collection

volume was 3.0 ml and the protein absorption was monitored at 280 nm. The inhibitory fraction (DI) was dialyzed and freeze-dried. This fraction was applied to a Sepharose 4-B-anhydrotrypsin affinity column using 50 mM Tris buffer pH 8.0. The column was washed with the same buffer and the inhibitors were eluted with 1 mM HCl in 500 mM NaCl. The biological activity of the inhibitor was detected using bovine trypsin and *N*- α -benzoyl-DL-arginine *p*-nitroanilide (BAPNA) as described by Erlanger *et al.* (1961). The purity of the inhibitor was checked by SDS-PAGE (Fig. 1).

2.2. N-terminal characterization

N-terminal sequencing was performed by Edman degradation (Edman & Begg, 1967). 40 nmol of purified inhibitor reduced with 1 M DTT was transferred to a PVDF membrane and applied on the automatic sequencer Procise 491 (Applied Biosystems). Two different N-terminal sequences have been derived, one of which, corresponding to 11 kDa chain, shows a high

degree of similarity with the members of the Kunitz family of inhibitors (Table 1). No homology was found, however, for the 9 kDa chain N-terminal amino-acid sequence.

2.3. Crystallization and X-ray data collection

All attempts to crystallize the separate polypeptide chains of the protein were unsuccessful. The intact protein inhibitor, however, was readily crystallizable. The crystals of the whole inhibitor used for data collection were grown by hanging-drop vapour diffusion at 291 K in 0.1 M sodium acetate buffer at a pH near 4.3 using PEG 4000 (20–25%) as a precipitant (Fig. 2). Drops consisted of equal volumes of protein at a concentration of 10 mg ml⁻¹ and reservoir solution. Small crystals grew after 4 d. The presence of both chains in the crystallized material was confirmed by SDS-PAGE gel (Fig. 1).

X-ray diffraction data was collected from crystals mounted in a rayon loop, immersed for 30 s in a cryocooling solution (20% ethylene glycol mixed with the mother liquor) and flash-cooled to 80 K in a cold nitrogen stream (Fig. 3). Data collection was performed at the Protein Crystallography beamline (Polikarpov, Oliva *et al.*, 1997; Polikarpov, Perles *et al.*, 1997) at the Brazilian National Synchrotron Light Laboratory (Campinas, SP, Brazil) using a MAR345 image plate. Data was autoindexed and integrated with the program *DENZO* (Otwinowski, 1993). Scaling and merging of data were performed with the program *SCALEPACK* (Otwinowski, 1993).

3. Results and discussion

The results from SDS-PAGE (Fig. 1) and N-terminal amino-acid sequencing (Table 1) indicate that the inhibitor under study has two chains, with molecular weights of approximately 11 and 9 kDa, that are non-covalently linked together to form a heterodimer of approximately 20 kDa. These two domains are separated under reducing conditions with 0.1 M DTT and have different N-terminal amino-acid composition.

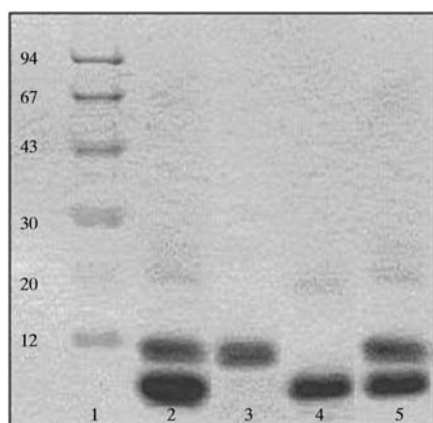


Figure 1

SDS-PAGE using a 16.5% tricine gel. Lane 1 shows the molecular-weight standards: phosphorylase (MW = 94 kDa), albumin (MW = 66 kDa), egg albumin (MW = 43 kDa), carbonic anhydrase (MW = 30 kDa), pancreas trypsin inhibitor (MW = 20 kDa) and cytochrome (MW = 12 kDa). Lane 2 shows the inhibitor extracted from *C. langsdorffii* seeds and reduced with 0.1 M DTT. Lanes 3 and 4 show the two non-covalently bound parts of the molecule separated with 0.1 M DTT and purified by HPLC using a reverse-phase column. Lane 5 shows the inhibitor from dissolved crystal.

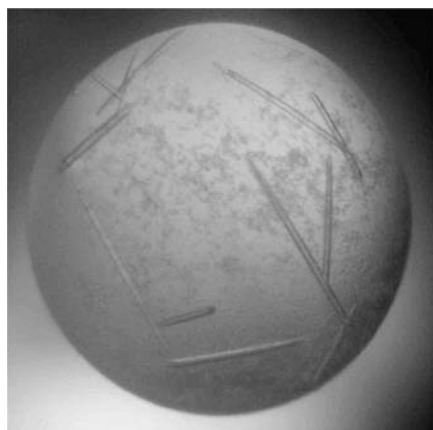


Figure 2

Crystals of *C. langsdorffii* protease inhibitor. The crystals were grown by the hanging-drop vapour diffusion method at 291 K in 0.1 M sodium acetate buffer at pH values near 4 using PEG 4000 (20–25%) as precipitant.

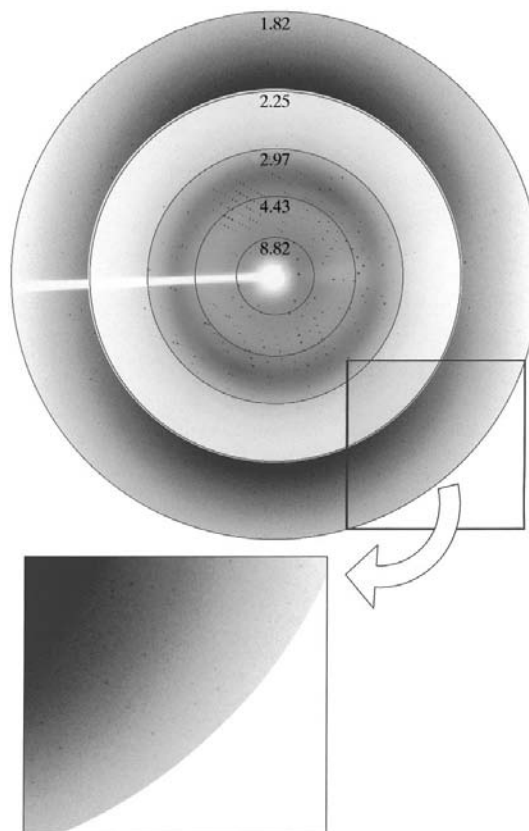


Figure 3

Diffraction pattern of the crystal. The outer area of the diffraction image (between resolution rings of 2.25 and 1.82 Å) is shown about four times more saturated than the inner part of the image. The close-up of the outer part of the image depicts reflections extending to a maximum resolution of 1.83 Å.

Table 1

N-terminal amino-acid sequences of the two inhibitor polypeptide chains separated with 0.1 M DTT and isolated by purification on a HPLC reverse-phase column.

The table also shows part of the soybean trypsin inhibitor primary structure, which presents 50% identity with chain I of the *C. langsdorffii* trypsin inhibitor. No significant homology was found for the chain II. *, identical; :, similar; ., different.

Inhibitor chains separated	N-terminal sequence
TI chain I (11 kDa)	RLVDNKGKPIENDGAEYYILPAVR -::***:*.*:.*.*.*****:*
STI (residues 2–23)	VLDTDGNPLRN–GGTYIILPAIR
TI chain II (9 kDa)	WQLPSVTGVGNPKVSAFGGPF

An X-ray diffraction data set (Table 2) was collected from a flash-frozen crystal measuring $0.1 \times 0.07 \times 0.03$ mm using synchrotron radiation with a wavelength of 1.38 Å. Diffraction pattern showed tetragonal Laue symmetry and systematic absences indicated that the crystal is compatible with the space group $P4_32_12$ or $P4_32_12$. The small size of the crystal limited diffraction intensities and the final resolution range. In order to optimize the X-ray diffraction data collection, the synchrotron-radiation wavelength was changed to 1.535 Å (Polikarpov, Teplyakov *et al.*, 1997). A second data set was collected from a different single crystal and evaluated. Its maximum resolution range extended to 1.83 Å. The crystal belonged to the same tetragonal space group. The unit-cell parameters were determined to be $a = b = 58.71$, $c = 93.75$ Å.

The calculated cell volume is 3.25×10^5 Å³. Assuming the molecular weight of the protein to be 20 kDa, the calculated Matthews coefficient (V_M ; Matthews, 1968) was 2.03 Å³ Da^{−1}, indicating the presence of one molecule in the asymmetric unit. A number of crystallographic models of Kunitz-type trypsin inhibitors available in Protein Data Bank were tested as search models for molecular replacement. However, all attempts to find molecular-replacement solution were unsuccessful. A systematic search for heavy-atom derivatives is currently under way. The quick cryoderivatization procedure recently

Table 2

Data collection and processing statistics.

Statistical values for the highest resolution shell are shown in parentheses.

Space group	$P4_32_12$
Unit-cell dimensions (Å)	
$a = b$	58.86
c	93.83
Wavelength (Å)	1.535
Crystal-to-detector distance (mm)	100
Resolution range (Å)	25.3–1.83 (1.87–1.83)
No. of reflections	123604
No. of unique reflections	15024
Redundancy	4.0
$\langle I/\sigma(I) \rangle$	23.6 (3.3)
Mosaicity (°)	0.5
Multiplicity	8.2 (5.4)
No. of images collected	100
Oscillation per image (°)	1.3
Completeness (%)	99.7 (97.6)
R_{merge} (%)	8.9 (50.8)

$$\dagger R_{\text{merge}} = \sum_{hkl} |I - \langle I \rangle| / \sum_{hkl} I.$$

introduced by Dauter *et al.* (2000) will be applied.

Financial support from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) via projects 99/03387-4 and 00/02317-1 and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) is acknowledged.

References

- Afzal, V., Wiencke, J. K. & Wolff, S. (1989). *Carcinogenesis*, **10**, 1193–1196.
- Billings, P. C., Morrow, A. R., Ryan, C. A. & Kennedy, A. R. (1989). *Carcinogenesis*, **10**, 687–691.
- Billings, P. C., Newberne, P. M. & Kennedy, A. R. (1990). *Carcinogenesis*, **11**, 1083–1086.
- Billings, P. C., St Clair, W., Ryan, C. A. & Kennedy, A. R. (1987). *Carcinogenesis*, **8**, 809–812.
- Birk, Y. (1985). *Int. J. Pept. Protein Res.* **25**, 113–131.
- Birk, Y. (1994). *Arch. Latinoam. Nutr.* **44**, 26S–30S.
- Bode, W. & Huber, R. (1991). *Biomed. Biochim. Acta*, **50**, 437–446.
- Bode, W. & Huber, R. (1992). *Eur. J. Biochem.* **204**, 433–451.
- Bowles, D. J. (1990). *Annu. Rev. Biochem.* **59**, 873–907.
- Breddam, K., Widmer, F. & Meldal, M. (1991). *Int. J. Pept. Protein Res.* **37**, 153–160.
- Broadway, R. M. & Duffey, S. S. (1986). *J. Insect Physiol.* **8**, 673–680.
- Broadway, R. M. & Duffey, S. S. (1988). *J. Insect Physiol.* **34**(12), 1111–1117.
- Chen, P., Rose, J., Love, R., Wei, C. H. & Wang, B. (1992). *J. Biol. Chem.* **267**, 1990–1994.

- Christeller, J. T., Laing, W. A., Markwick, N. P. & Burgess, E. P. J. (1992). *Insect Biochem. Mol. Biol.* **22**, 735–746.
- Correa, P. (1981). *Cancer Res.* **41**, 3685–3690.
- Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst. D* **56**, 232–237.
- Edman, P. & Begg, G. (1967). *Eur. J. Biochem.* **1**, 80–91.
- Erlanger, B. F., Kolowsky, N. & Cohen, N. (1961). *Arch. Biochem. Biophys.* **95**, 271–278.
- Flavin, D. F. (1982). *Vet. Hum. Toxicol.* **24**, 25–28.
- Frenkel, K., Chrzan, K., Ryan, C. A., Wiesner, R. & Troll, W. (1987). *Carcinogenesis*, **8**, 1207–1212.
- Kennedy, A. R. (1994). *Cancer Res.* **54**, 1999S–2005S.
- Kennedy, A. R. (1998). *Pharmacol. Ther.* **78**, 167–209.
- Laskowski, M. Jr & Kato, I. (1980). *Annu. Rev. Biochem.* **49**, 593–626.
- Liener, I. E. (1995). *J. Nutr.* **125**, 744S–759S.
- Lorenzi, H. (1992). *Arvores Brasileiras – Manual de Identificação e Cultivo de Plantas Nativas do Brasil*. Editora Platarum LTDA.
- Matthews, B. M. (1968). *J. Mol. Biol.* **33**, 491–497.
- Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
- Polikarpov, I., Oliva, G., Castellano, E. E., Garratt, R., Arruda, P., Leite, A. & Craievich, A. (1997). *Nucl. Instrum. Methods A*, **405**, 159–164.
- Polikarpov, I., Perles, L. A., de Oliveira, R. T., Oliva, G., Castellano, E. E., Garratt, R. & Craievich, A. (1997). *J. Synchrotron Rad.* **5**, 72–76.
- Polikarpov, I., Teplyakov, A. & Oliva, G. (1997). *Acta Cryst. D* **53**, 734–737.
- Pusztai, A. (1989). *Toxicants of Plant Origin*, Vol. 2, edited by P. R. Sheek, pp. 29–71. Boca Raton, Florida: CRC Press.
- Richardson, M. (1991). *Methods Plant. Biochem.* **5**, 259–305.
- Ryan, C. A. (1990). *Annu. Rev. Phytopathol.* **28**, 425–449.
- Shewry, P. R. & Lucas, J. A. (1997). *Adv. Bot. Res.* **26**, 135–192.
- Tamir, S., Kadner, S., Katz, J. & Finlay, T. H. (1990). *Endocrinology*, **127**, 1319–1328.
- Troll, W. & Kennedy, A. R. (1989). *Cancer Res.* **49**, 499–502.
- Troll, W., Wiesner, R. & Frenkel, K. (1987). *Adv. Cancer Res.* **49**, 265–283.
- Troll, W., Wiesner, R., Shellabarger, C. J., Holtzman, S. & Stone, J. P. (1980). *Carcinogenesis*, **1**, 469–472.
- Vigers, A. J., Roberts, W. K. & Selitrennikoff, C. P. (1991). *Mol. Plant-Microbe Interact.* **4**, 315–323.
- Weder, J. K. P. (1992). *Annu. Rev. Biochem.* **16**, 239–279.
- Witschi, H. & Kennedy, A. R. (1989). *Carcinogenesis*, **10**, 2275–2277.
- Yavelow, J., Finlay, T. H., Kennedy, A. R. & Troll, W. (1983). *Cancer Res.* **43**, 2454–2459.
- Yavelow, J., Gidlund, M. & Troll, W. (1982). *Carcinogenesis*, **3**, 135–138.



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Biochemical and Biophysical Research Communications 312 (2003) 1303–1308

BBRC

www.elsevier.com/locate/ybbrc

Crystal structure of the Kunitz (STI) type inhibitor from *Delonix regia* seeds

Sandra Krauchenco,^a Silvana C. Pando,^b Sérgio Marangoni,^c and Igor Polikarpov^{a,*}

^a Instituto de Física de São Carlos, USP, Av. Trabalhador Saocarlense, 400, CEP 13560-970, São Carlos, SP, Brazil

^b Instituto de Biologia, Universidade Estadual do Centro Oeste, Guarapuava, PR, Brazil

^c Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brazil

Received 24 October 2003

Abstract

The three-dimensional structure of a novel Kunitz (STI) family member, an inhibitor purified from *Delonix regia* seeds (DrTI), was solved by molecular replacement method and refined, respectively, to R_{factor} and R_{free} values of 21.5% and 25.3% at 1.75 Å resolution. The structure has a classical β -trefoil fold, however, differently from canonical Kunitz type (STI) inhibitors, its reactive site loop has an insertion of one residue, Glu68, between the residues P1 and P2. Surprisingly, DrTI is an effective inhibitor of trypsin and human plasma kallikrein, but not of chymotrypsin and tissue kallikrein. Putative structural grounds of such specificity are discussed. © 2003 Elsevier Inc. All rights reserved.

Keywords: Kunitz-type inhibitor; Kallikrein inhibitor; X-ray structure; β -Trefoil fold; Flamboyant; *Delonix regia*

Proteinase could be inhibited by various compounds including proteins that act by forming stoichiometric and stable complexes with proteolytic enzymes, inhibiting their activity, and preventing the unwanted proteolysis [1]. Presence of proteinase inhibitors in plants and seeds frequently accounts for the low nutritive value of raw vegetarian meals [2]. Epidemiological studies suggest that human populations which are known to consume food with high concentration of proteinase inhibitors in their diet have lower rates of colon, breast, prostate, and skin cancers [3].

The protein proteinase inhibitors are divided into families according to the class of proteolytic enzymes inhibited, extensive sequential and structural homology among the members, and the locations of disulfide bridges and the reactive site [1]. The Kunitz type inhibitors of serine-proteinases are subdivided into two families. The Kunitz bovine pancreatic trypsin inhibitor (BPTI) family members have a molecular mass of about 6.5 kDa and three disulfide bridges. The Kunitz soybean trypsin inhibitor (STI) family is presented by proteins

with a molecular mass of about 20 kDa that contain two disulfide bridges.

Here we discuss the crystallographic structure solution of a new member of the Kunitz (STI) family inhibitors, purified from *Delonix regia* seeds (DrTI). DrTI has a molecular mass of 22 kDa and two disulfide bridges. Biochemical studies clearly demonstrated that it is an effective inhibitor of trypsin and human plasma kallikrein with K_i values of 21.9 and 5.3 nM respectively, but it is not active toward chymotrypsin or tissue kallikrein [4]. Present crystallographic structure reveals a molecular basis of such specificity.

Materials and methods

Crystallization. DrTI was purified as described previously [4,5]. The lyophilized inhibitor was dissolved in 25 mM sodium phosphate buffer, pH 8.0, at a concentration of 10 mg/ml. The crystals suitable to the data collection were grown by the hanging-drop vapor diffusion method at 291K in 0.1 M sodium phosphate buffer, pH 8.5, with 0.25 M ammonium sulfate and PEG 8000 (20–25%) as precipitants.

Data collection and processing. X-ray data were collected from a single cryo-cooled crystal at the Protein Crystallography beamline of the Brazilian National Synchrotron Light Laboratory (Campinas, SP, Brazil) [6]. The synchrotron-radiation wavelength was set to 1.38 Å to improve the signal-to-noise ratio in diffraction data [7]. The DrTI

*Corresponding author. Fax: +55-16-273-9881.

E-mail address: ipolikarpov@ifsc.usp.br (I. Polikarpov).

Table 1

Summary of data collection and refinement statistics (values for the highest resolution shell are shown in parentheses)

Space group	P2 ₁ 2 ₁ 2 ₁
Cell dimensions (Å)	<i>a</i> = 32.13, <i>b</i> = 67.25, <i>c</i> = 72.06
Crystal-to-detector distance (mm)	130
Resolution range (Å)	30.0–1.75 (1.89–1.75)
No. of reflections	75,603
No. of unique reflections	16,616
Redundancy	4.6 (3.2)
$\langle I/\sigma(I) \rangle$	9.0 (2.4)
Mosaicity (°)	0.6
Completeness (%)	82
^a <i>R</i> _{merge} (%)	6.1 (38.1)
Refinement statistics	
No. of protein atoms	1417
No. of water molecules	183
^b <i>R</i> _{factor} (%)	21.5
^c <i>R</i> _{free} (%)	25.3
Average <i>B</i> _{factor}	
Main-chain (Å ²)	32.5
Side-chains (Å ²)	34.8
Water molecules (Å ²)	45.4
All atoms	34.9
r.m.s. deviations from ideal geometry	
Bonds (Å ²)	0.0050
Bond angles (°)	1.2
Ramachandran plot statistics	
Residues in most favored regions	134 (88.2%)
Residues in additional allowed regions	13 (8.6%)
Residues in generously allowed regions	5 (3.3%)
Residues in disallowed regions	0 (0%)

^a *R*_{merge} = $\sum_{hkl} |I - \langle I \rangle| / \sum_{hkl} I$.

^b *R*_{factor} = $\sum (|F_{obs.}| - |F_{calc.}|) / \sum F_{obs.}$, where *F*_{obs.} and *F*_{calc.} are the observed and calculated structure factor amplitudes, respectively.

^c *R*_{free} is the *R*_{factor} calculated for a subset of 5% of randomly selected reflections.

level. However, the electron density for the residues 180–185 of the C-terminal portion of the protein is somewhat weak and fragmented. The deviations from ideal bonds and angles are small (Table 1) and there are no residues in the disallowed region of the Ramachandran plot. The loop regions, including the reactive loop, and the C-terminal portion are characterized by high temperature factor values than structurally more constrained parts of molecule.

Overall structure

The structure of DrTI, of a β-trefoil fold, consists of 12 antiparallel β-strands connected by long loops that form six two-stranded β-hairpins [14]. Three of the β-hairpins form a barrel structure while the remaining three form a triangular cap on the barrel (Fig. 1). Three structural repeats related by a pseudo-threefold axis of symmetry oriented parallel to the barrel axis are a characteristic feature of this fold. The repetition unit consists of about 60 aminoacid residues and is divided into two approximately equal parts: 30 aminoacids of four sequential β-strands and the other 30 forming the connecting loops. Each of the 12 β-strands of the fold is labeled according to the McLachlan classification [15] (Fig. 1).

The repeats do not bear detectable sequence similarity but are readily apparent following structure superposition. The superposition of structurally equivalent Cα atoms of the β-strands gives an r.m.s. deviation of 1.77 Å between subdomains A and B, 1.53 Å between subdomains A and C, and 1.59 Å between subdomains B and C.

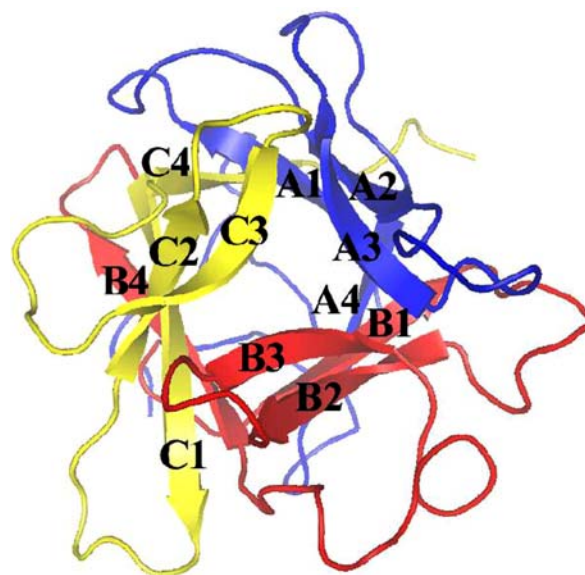


Fig. 1. The overall structure of DrTI. The protein has the β-trefoil fold formed by the three structural repeats A–C, painted in blue, red, and yellow, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

crystal diffracted beyond 1.75 Å resolution. Data were integrated and scaled with DENZO and SCALEPACK [8] (Table 1).

Structure solution and refinement. The DrTI structure was solved by molecular replacement method using the program AMoRe [9]. Kunitz STI family inhibitors from *Glycine max* seeds [10] (STI, PDB code 1AVU) and from *Erythrina caffra* seeds [11] (ETI, PDB code 1TIE) were used as search models and resulted in essentially the same structure solution. Several cycles of positional and restrained isotropic B-factor refinement with REFMAC [9], intercalated with simulated annealing using CNS package [12], were performed. The program O was used to analyze and correct the model [13]. The water molecules were added according to the criteria that each water molecule must make at least one stereochemically reasonable hydrogen bond, that it should be well defined in $(2m|F_{obs.}| - D|F_{calc.}|)$ and $(m|F_{obs.}| - D|F_{calc.}|)$ electron density maps. The data collection and refinement statistics is reported in Table 1.

Results

Quality of the model

Essential part of the crystallographic model is well defined by electron density calculated at the 2.5–3σ

110 Comparison with homologous structures

111 DrTI shares the β -trefoil fold with several super-
 112 families of proteins, such as cytokines [16–19], ricin
 113 B-like lectins [20–22], agglutinins [23], actin-crosslinking
 114 proteins [24,25], and STI-like proteins, that include the
 115 Kunitz (STI) family of inhibitors and tetanus [26] and
 116 botulinum [27] neurotoxins. Although these superfami-
 117 lies of proteins demonstrate considerable structural ho-
 118 mology, they have little sequence similarity, different
 119 functions, and occur in different subcellular localiza-
 120 tions. In spite of drastic differences between these su-
 121 perfamilies, there is enough statistical evidence to
 122 conclude that the β -trefoil fold is a result of a double
 123 gene duplication and thus these proteins could have
 124 arisen from a common ancestor [28].

125 DrTI is a member of the Kunitz (STI) family and is
 126 similar to other plant serine proteinase inhibitors such as
 127 STI [10], ETI [11], and trypsin inhibitor from *Copaifera*
 128 *langsdorffii* seeds (CTI) [29].

129 The structural homology is clear and a superposition
 130 is particularly good in the conserved β -strand regions.
 131 Large variations observed in the loop regions are, in
 132 part, due to the insertions or deletions of residues in
 133 these parts of the proteins.

134 The reactive loop

135 Structure and amino acid comparison of DrTI with
 136 other inhibitors of the Kunitz (STI) family defines its
 137 reactive site loop from the residue Ser65(P4) to the
 138 residue Ile73(P4'). Fig. 2 shows the superposition of the
 139 residues of the DrTI reactive site loop overlapped with
 140 the final (2Fo–Fc) electron density map at 1.2 σ level.

141 The structural comparison shows that only STI, ETI,
 142 and WCI have almost exactly the same conformation
 143 for the reactive loop from residue P4 to P4', which is
 144 known as *canonical conformation*. The reactive loop
 145 conformation of the albumin WBA [30], the α -amylase,
 146 and the subtilisin inhibitor BASI [31] is significantly
 147 different from the canonical conformation (Fig. 3A).

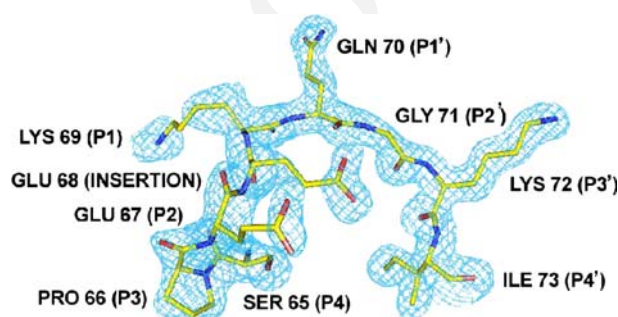


Fig. 2. The final (2Fo–Fc) electron density map, contoured at 1.0 σ , around the reactive loop of DrTI structure. The residues Ser65–Ile73 (P4–P4') are shown.

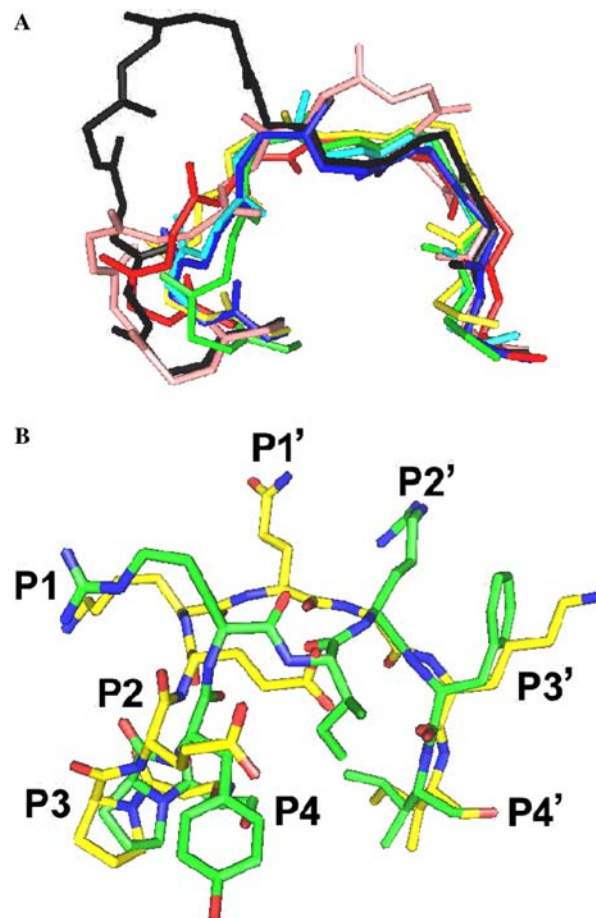


Fig. 3. (A) The structural superposition of the reactive loop backbones (residues P4–P4') of the Kunitz STI family members. DrTI is painted in red and STI in blue, ETI in yellow, chymotrypsin inhibitor WCI in green, CTI in cyan, BASI in salmon, and WBA in black. (B) The superposition of the reactive loops of DrTI (painted in yellow) with STI (in green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

The superposition between the reactive site loops of DrTI and STI (Fig. 3B) shows that an insertion of one residue, Glu68, between the residues P1 and P2 distorts DrTI reactive loop as compared to the canonical conformation. This deformation could explain the reduced DrTI inhibitory activity against trypsin. DrTI inactivates trypsin with a K_i value of 21.9 nM [5], while canonical Kunitz (STI) type trypsin inhibitors, STI or ETI, have K_i values in the 1 nM range. Furthermore, the presence of two negatively charged residues, Glu67 and Glu68, in DrTI reactive loop seems to be related to the specificity of this inhibitor toward human plasma kallikrein (HPK) but not tissue kallikrein (TK).

Interaction of DrTI with trypsin and HPK

The kallikreins are a sub-family of serine proteases, which are divided into two groups, plasma (EC 3.4.21.34) and tissue (EC 3.4.21.35) kallikreins. These

groups are significantly different in molecular weight, substrate specificity, immunological characteristics, gene structure, and a type of the kinin released. Plasma kallikrein is expressed solely in a liver and is involved in blood clotting, fibrinolysis, regulation of blood pressure, and inflammatory reactions. Tissue kallikreins are a large group of enzymes which have substantial similarities at gene and protein level. Tissue kallikreins are involved in the post-translational processing of the polypeptides (like kininogen) and releasing potential biologically active peptides (e.g., kinin). [32] Although the sequences of both human plasma kallikrein (HPK) and tissue kallikrein (TK) are determined, only the TK three-dimensional structure is known.

In an attempt to find a structural rationale as to why DrTI is an effective inhibitor of HPK with a K_i value of 5.25 nM, but does not present any inhibitory activity against TK, we modeled the DrTI putative interactions with trypsin and HPK.

First, the DrTI structure was superimposed with the STI–trypsin complex structure (PDB code 1AVW) [10] and the putative contacts between DrTI–trypsin can be analyzed. The intermolecular contacts involve not only residues of the reactive site, but also residues in other parts of the molecules forming a molecular interface between the enzyme and the inhibitor.

As a second step, the structures of chymotrypsin (PDB code 5CHA) [33] and TK (PDB code 2KAI) [34]

were superimposed with the trypsin structure in the DrTI–trypsin putative complex. Intriguingly, TK, as compared to trypsin, has various insertions at the molecular interface region. For example, the insertions of Ala–Asp–Gly–Lys between residues Ser97(TK) and Asp98(TK) and of Pro between the residues Thr219(TK) and Cys220(TK) significantly modify the molecular shape of TK. This might decrease surface complementarities between inhibitor and enzyme and contribute to the lack of inhibitory activity of the former against the latter.

Finally, as the HPK structure is not known, its sequence (swiss-prot code P03952) [35] was aligned with the other enzymes (Fig. 4). Only the part of the HPK sequence that presents homology with serine proteinases (residues 360–646) was used in the alignment. Analysis of the alignment shows that about half of the residues in contacts regions are conserved between all the enzymes or conservatively substituted between trypsin and HPK. For example, trypsin residue Ser37 is substituted by Ala422 in HPK, which presumably would not cause significant distortions at the DrTI molecular interface, but is substituted by Tyr37 in TK. Its significantly larger side-chain would provoke sterical hindrance in the putative interactions with Lys5 and Lys72 of DrTI since these residues make contacts with trypsin residue Ser37. Among all substitutions, there is one that calls special attention, namely Gln192(trypsin) into Lys583(HPK)

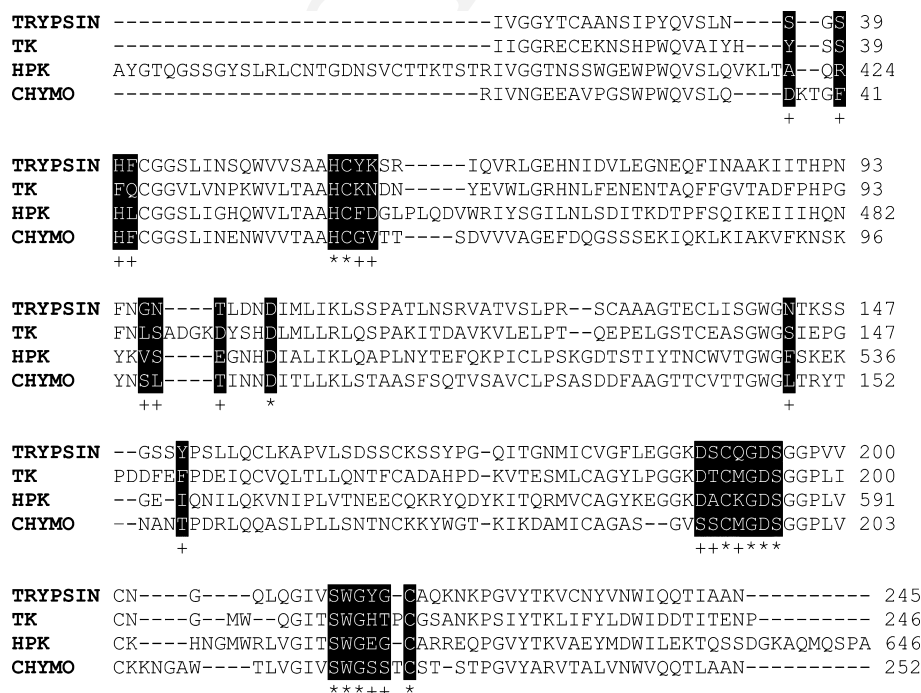


Fig. 4. Structure guided sequence alignment of the trypsin (PDB code 1AVW) with chymotrypsin (PDB code 5CHA), tissue kallikrein (PDB code 2KAI), and human plasma kallikrein (swiss-prot code P03952). Only the part of the sequence of the HPK that presents homology with serine proteinases (360–646) was used in the alignment. The residues that are involved in putative contacts with DrTI are in black boxes. The (*) indicates conserved contact residues and the (+) indicates residues that are not conserved.

and into Met192(TK). Trypsin residue Gln192 is putatively involved in a number of contacts including the residues Ser65, Glu68, Lys69, and Gln70 of the DrTI reactive loop. The substitution Gln192 (trypsin) into Met192(TK) replaces the amide side chain group by the more inert methyl group. But the substitution Gln192(trypsin) by Lys583(HPK) could open a possibility of new contacts both with the residues Glu67 and Glu68 of the DrTI reactive loop. These observations go in line with the experimental fact that DrTI is efficient as HPK inhibitor but not as TK inhibitor. Furthermore, HPK, unlike TK, does not have insertions at the molecular interface region, yet another factor that might favor inhibitory activity of DrTI against HPK.

Discussion

There are an intricate number of factors that contribute to the specificity and the efficiency of protease inhibitors. The composition and the conformation of their reactive loop site are clearly important, but various other aspects of the molecular structure must be considered as well. This is because the inhibitor reactive loop interacts not only with the enzyme, but there are other important areas of contact between these two proteins. In a way, the entire molecule must be considered.

The global architecture, for example, is essential to the inhibitor stability, rigidity, and efficiency. The reactive loop site of Kunitz (STI) family of inhibitors is located between the repeats A and B. The cleavage of scissile peptide bond theoretically could provoke the separation of the motif A of the motifs B and C. However, the motifs A–C are held together by an extensive intramolecular hydrogen bonding network. There has been substantial NMR evidence that this network, which does not exist in a normal substrate vulnerable to proteolysis, remains intact in the cleaved inhibitor [36]. These numerous contacts stabilize the newly formed N-terminus and maintain it in an optimal orientation for nucleophilic attack on the acyl-enzyme, favoring the religation. Moreover, the positioning of the amine also sterically hinders the hydrolytic water molecule from achieving the necessary proximity to the histidine base for nucleophilic activation.

The composition of the reactive loop determines the specificity of the inhibitor. The amino acid residue in the P1 position, which fits in the enzyme reactive site pocket, is of particular importance. Besides the specificity, the loop conformation also determines the efficiency of the inhibitor as in the substrate-like inhibitors, in which much closer to the canonical conformation the reactive loop is, the more effective is the inhibitor.

The lack of a hydrophobic residue in the P1 position of DrTI reactive loop, which is indispensable for the

inhibition of chymotrypsin [37], can explain why DrTI has no inhibitory activity against this enzyme. The DrTI activity against trypsin is granted by the presence of a positively charged residue, Lys69, at P1 position which is essential. However, compared to canonical trypsin inhibitors, the reactive loop of DrTI is distorted due to an insertion of Glu68 between the residues P1 and P2. This weakens the inhibitory activity of DrTI against trypsin. DrTI inhibition constant K_i against trypsin is 21.9 nM, whereas the canonical inhibitors from the family have K_i close to 1 nM.

DrTI is an effective inhibitor of HPK despite it having no inhibitory activity against TK. It is a combination of two crucial factors that seems to contribute to DrTI specificity. The first one is the presence of two glutamic acids at the DrTI reactive loop. Analysis of the regions that are involved in the interactions between the inhibitor and enzyme demonstrates that there is a replacement of the Gln192 in trypsin molecule by Lys583 in HPK. In addition to putative interactions with the residues Ser65, Glu68, Ser69, and Gln70 of DrTI reactive loop, Lys583 may be involved in a new electrostatic interaction with the side chain of Glu67 and Glu68. In the TK structure, the Gln192 (trypsin) is replaced by Met192, which cannot participate in a similar charge–charge interactions. The second crucial factor is the presence of insertions in the TK molecule. These insertions modify TK molecular surface and might hinder the interactions between inhibitor and enzyme. In contrast, HPK has a close sequence similarity with trypsin and its contact surface area with DrTI is presumably preserved. Although our modeling studies cannot assure that other structural reasons do not exist, they provide a basis for site-directed mutagenesis studies. Furthermore, they are in line with observed specificity of DrTI inhibitory activity.

Acknowledgments

We thank Valéria Peyrl Forrer and Jackson Antônio Lamounier Camargo Resende for help with crystallization and José Ribeiro Brandão Neto for help with data collection and processing. Financial support from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, Grants 99/04497-8 and 99/03387-4) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) is acknowledged.

References

- [1] M. Laskowsky Jr., I. Kato, Protein inhibitors of proteinases, *Annu. Rev. Biochem.* 49 (1980) 593–626.
- [2] I.E. Liener, Effects of processing on antinutritional factors in legumes: the soybean case, *Arch. Latinoam. Nutr.* 44 (4) (1996) 48s–54s.
- [3] P. Correa, Epidemiological correlations between diet and cancer frequency, *Cancer Res.* 41 (9) (1981) 3685–3690.

- 326 [4] S.C. Pando, M.L.V. Oliva, C.A.M. Sampaio, L. Di Ciero, J.C.
327 Novello, S. Marangoni, Primary sequence determination of a
328 Kunitz inhibitor isolated from *Delonix regia* seeds, *Phytochemis-*
329 *try* 57 (2001) 625–631.
- 330 [5] I. Polikarpov, A.M. Golubev, L.A. Perles, S.C. Pando, J.C.
331 Novello, S. Marangoni, Purification, crystallization and prelimin-
332 ary crystallographic study of a Kunitz-type trypsin inhibitor from
333 *Delonix regia* seeds, *Acta Crystallogr. D* 55 (1999) 1611–1613.
- 334 [6] I. Polikarpov, G. Oliva, E.E. Castellano, R. Garratt, P. Arruda,
335 A. Leite, A. Craievich, Protein crystallography station at LNLS,
336 The Brazilian National Synchrotron Light Source, *Nucl. Instr.*
337 *Methods A* 405 (1998) 159–164.
- 338 [7] I. Polikarpov, A. Teplyakov, G. Oliva, The ultimate wavelength
339 for protein crystallography, *Acta Crystallogr. D* 53 (1997) 734–
340 737.
- 341 [8] Z. Otwinowski, Data collection and processing, in: L. Sawyer, N.
342 Isaacs, S. Bailey (Eds.), *Proceedings of the CCP4 Study Weekend*,
343 SERC Daresbury Laboratory, Daresbury, England, 1993, pp. 56–
344 62.
- 345 [9] Collaborative Computational Project Number 4, The CCP4 Suite:
346 Programs for Protein Crystallography, *Acta Crystallogr. D* 50
347 (1994) 760–763.
- 348 [10] H.K. Song, S.W. Suh, Kunitz-type soybean trypsin inhibitor
349 revisited: refined structure of its complex with porcine trypsin
350 reveals an insight into the interaction between a homologous
351 inhibitor from *Erythrina caffra* and tissue-type plasminogen
352 activator, *J. Mol. Biol.* 275 (1998) 347–363.
- 353 [11] S. Onesti, P. Brick, D.M. Blow, Crystal structure of a Kunitz-type
354 trypsin inhibitor from *Erythrina caffra* seeds, *J. Mol. Biol.* 217
355 (1990) 153–176.
- 356 [12] A.T. Brünger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros,
357 R.W. Grosse-Kunstleve, J. Jiang, J. Kuszewski, M. Nilges, N.S.
358 Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren,
359 Crystallography & NMR system: a new software suite for
360 macromolecular structure determination, *Acta Crystallogr. D* 54
361 (1998) 905–921.
- 362 [13] T.A. Jones, M. Bergdoll, M. Kjeldgaard, C. Bugg, in: S. Ealick
363 (Ed.), *Crystallographic and Modeling Methods in Molecular*
364 *Design*, Springer, Berlin, 1990, pp. 189–195.
- 365 [14] A.G. Murzin, A.M. Lesk, C. Chothia, Beta-trefoil fold, *J. Mol.*
366 *Biol.* 223 (1992) 531–543.
- 367 [15] A.D. McLachlan, Threefold structural pattern in the soybean
368 trypsin inhibitor (Kunitz), *J. Mol. Biol.* 133 (1979) 557–563.
- 369 [16] X. Zhu, H. Komiya, A. Chirino, S. Faham, G.M. Fox, T.
370 Arakawa, B.T. Hsu, D.C. Rees, Three-dimensional structures of
371 acidic and basic fibroblast growth factors, *Science* 251 (1991)
372 90–93.
- 373 [17] J.P. Priestle, H.P. Schär, M. Grütter, Crystal structure of the
374 cytokine interleukin-1 β , *EMBO J.* 7 (2) (1988) 339–343.
- 375 [18] B.J. Graves, M.H. Hatada, W.A. Hendrickson, J.K. Miller, V.S.
376 Madison, Y. Satow, Structure of interleukin 1 α at 2.4 Å resolu-
377 tion, *Biochemistry* 29 (11) (1990) 2679–2689.
- 378 [19] H. Schreuder, C. Tardif, S. Trump-Kallmeyer, A. Soffientini, E.
379 Sarubbi, A. Akeson, T. Bowlin, S. Yanofsky, R.W. Barrett, A new
380 cytokine–receptor binding mode revealed by the crystal structure
381 of the IL-1 receptor with an antagonist, *Nature* 386 (6621) (1997)
382 194–200.
- 383 [20] E. Rutenber, B.J. Katzin, S. Ernst, E.J. Collins, D. Mlsna, M.P.
384 Ready, J.D. Robertus, Crystallographic refinement of ricin to
385 2.5 Å, *Proteins* 10 (3) (1991) 240–250.
- 386 [21] V. Notenboom, A.B. Boraston, S.J. Williams, D.G. Kilburn, D.R.
387 Rose, High resolution crystal structures of the lectin-like xylan
binding domain from *Streptomyces lividans* xylanase 10 Å with
bound substrates reveal a novel mode of xylan binding, *Biochem-*
istry 41 (13) (2002) 4246–4254.
- [22] Y. Liu, A.J. Chirino, Z. Misulovin, C. Leteux, T. Feizi, M.C.
Nussenzweig, P. Bjorkman, Crystal structure of the cysteine-rich
domain of mannose receptor complexed with a sulfated carbohy-
drate ligand, *J. Exp. Med.* 191 (7) (2000) 1105–1115.
- [23] T.R. Transue, A.K. Smith, H. Mo, I.J. Goldstein, M.A. Saper,
Structure of benzyl T-antigen disaccharide bound to *Amaranthus*
caudatus agglutinin, *Nat. Struct. Biol.* 4 (10) (1997) 779–783.
- [24] J. Habazettl, D. Gondol, R. Wilschek, J. Otlewski, M. Schle-
icher, T.A. Holak, Structure of hisactophilin is similar to interleu-
kin-1b and fibroblast growth factor, *Nature* 359 (6398) (1992)
855–858.
- [25] S. Ono, Y. Yamakita, S. Yamashiro, P.T. Matsudaira, J.R.
Gnarra, T. Obinata, F. Matsumura, Identification of an actin
binding region and a protein kinase C phosphorylation site on
human fascin, *J. Biol. Chem.* 272 (4) (1997) 2527–2533.
- [26] P. Emsley, I. Fotinou, N.F. Fairweather, I.G. Charles, C. Watts,
E. Hewitt, N.W. Isaacs, The structures of the H_C fragment of
tetanus toxin with carbohydrate subunit complexes provide
insight into ganglioside binding, *J. Biol. Chem.* 275 (12) (2000)
8889–8894.
- [27] D.B. Lacy, W. Tepp, A.C. Cohen, B.R. Dasgupta, R.C. Stevens,
Crystal structure of botulinum neurotoxin type A and implica-
tions for toxicity, *Nat. Struct. Biol.* 5 (1998) 898–902.
- [28] C.P. Ponting, R.B. Russell, Identification of distant homologues
of FGFs suggests a common ancestor for all beta-trefoil proteins,
J. Mol. Biol. 302 (2000) 1041–1047.
- [29] S. Krauchenco, J.A. Silva, R.A.P. Nagem, J.R. Brandão Neto,
V.P. Forrer, R. Carmona e Ferreira, M.L.R. Macedo, J.C.
Novello, S. Marangoni, I. Polikarpov, Crystallization and
preliminary X-ray diffraction analysis of a novel trypsin inhibitor
from seeds of *Copaifera langsdorffii*, *Acta Crystallogr. D* 57 (2001)
1316–1318.
- [30] A.J. McCoy, A.A. Kortt, The 1.8 Å crystal structure of winged
bean albumin 1, the major albumin from *Psophocarpus tetrago-*
nolobus (L.) DC, *J. Mol. Biol.* 269 (1997) 881–891.
- [31] F. Vallee, A. Kadziola, Y. Bourne, M. Juy, K.W. Rodenburg, B.
Svensson, R. Haser, Barley α -amylase bound to its endogenous
protein inhibitor BASI: crystal structure of the complex at 1.9 Å
resolution, *Structure* 6 (5) (1998) 649–659.
- [32] J.A. Clements, The molecular biology of the kallikreins and their
roles in inflammation, in: S.G. Farmer (Ed.), *The Kinin System*,
vol. 15, Academic Press, San Diego, CA, 1997, pp. 71–97.
- [33] R.A. Blevins, A. Tulinsky, The refinement and the structure of the
dimmer of α -chymotrypsin at 1.67 angstroms resolution, *J.*
Biol. Chem. 260 (1985) 4264.
- [34] Z. Chen, W. Bode, Refined 2.5 angstroms X-ray crystal structure
of the complex formed by porcine kallikrein A and the bovine
pancreatic trypsin inhibitor, *J. Mol. Biol.* 164 (1983) 283.
- [35] D.W. Chung, K. Fujikawa, B.A. McMullen, E.W. Davie, Human
plasma prekallikrein, a zymogene to a serine protease that
contains four tandem repeats, *Biochemistry* 25 (1986) 2410–2417.
- [36] G.L. Shaw, B. Davis, J. Keeler, A.R. Fersht, Backbone dynamics
of chymotrypsin inhibitor 2: effect of breaking the active-site bond
and its implications for the mechanism of inhibition of serine
proteases, *Biochemistry* 34 (7) (1995) 2225–2233.
- [37] S. Ravichandran, U. Sen, C. Chakrabarti, J.K. Dattagupta,
Cryocrystallography of a Kunitz-type serine protease inhibitor:
the 90 K structure of a winged bean chymotrypsin inhibitor (WCI)
at 2.13 Å resolution, *Acta Crystallogr. D* 55 (1999) 1814–1821.



Three-dimensional structure of an unusual Kunitz (STI) type trypsin inhibitor from *Copaifera langsdorffii*

Sandra Krauchenco ^a, Ronaldo A.P. Nagem ^a, José A. da Silva ^b,
Sérgio Marangoni ^b, Igor Polikarpov ^{a,*}

^a Grupo de Cristalografia, Department of Physics and Informatics, Institute of Physics (IFSC), University of São Paulo (USP),
Avenue Trabalhador Saocarlene, 400, CEP 13560-970, São Carlos, SP, Brazil

^b Instituto de Biologia, UNICAMP, Campinas, SP, Brazil

Received 27 November 2003; received in revised form 11 March 2004; accepted 12 March 2004

Abstract

The crystallographic structure of a novel trypsin inhibitor (CTI) from *Copaifera langsdorffii* is reported. The structure was solved by MIRAS procedure and refined to a crystallographic residual of 17.3% ($R_{\text{free}} = 20.3\%$) at 1.8 Å resolution. Two isomorphous derivatives were obtained by quick cryo-soaking approach. CTI is the first structure of a member of Kunitz (STI) family formed by two noncovalently bound polypeptide chains and only one disulfide bridge. A standard Kunitz-type inhibitor has a single polypeptide chain and two disulfide bridges. Structural features granting CTI high inhibitory activity are discussed.

© 2004 Published by Elsevier SAS.

Keywords: Kunitz-type trypsin inhibitor; Quick cryo-soaking; X-ray structure; β -Trefoil fold; *Copaifera langsdorffii*

1. Introduction

Proteinase inhibitors are proteins capable to inhibit the activity of hydrolytic enzymes by forming stable stoichiometric complexes [1]. Proteinase inhibitors are widely distributed among living organisms and their major physiological function appears to be to prevent unwanted proteolysis [2]. Plant inhibitors confer a broad spectrum of resistance to pests and pathogens by inhibiting the action of the bacterial proteinases [3]. Serine proteinase inhibitors of Bowman-Birk and Kunitz (STI) families possess anticarcinogenic properties [4]. These include the ability to decrease oxygen radical

formation, suppress the growth of chemically-induced colon, breast and lung tumors and to reduce spontaneous chromosome abnormalities [5,6].

There are 18 canonical families of serine proteinase inhibitors. The Kunitz family is divided into two subfamilies: (i) Kunitz STI-like, with molecular mass of about 20 kDa and two disulfide bridges; (ii) Kunitz BPTI-like, with molecular mass of about 6.5 kDa and three disulfide bridges [7].

An unusual trypsin inhibitor (CTI) was isolated from seeds of *Copaifera langsdorffii* tree (*Leguminosae-Caesalpinioideae*) [8]. We report the three-dimensional structure of CTI solved by MIRAS using the quick cryo-soaking procedure for derivatization [9,10]. To the best of our knowledge, this is the first report of a crystallographic structure of an inhibitor of the Kunitz (STI) family with two noncovalently bound polypeptide chains and a single disulfide bridge. The structural basis of a high inhibitory activity of CTI is discussed.

2. Materials and methods

2.1. Crystallization and derivatives preparation

CTI was purified and crystallized following earlier described protocols [8,11]. The derivatives were prepared by

Abbreviations: CTI, trypsin inhibitor from *Copaifera langsdorffii*; DrTI, Kunitz (STI) type trypsin inhibitor from *Delonix regia* seeds; ESI, electrospray ionization mass spectrometry; ETI, trypsin inhibitor from *Erythrina caffra*; MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight; MIRAS, multiple isomorphous replacement with anomalous scattering; PAGE, polyacrylamide gel electrophoresis; r.m.s., root mean square; SIRAS, single isomorphous replacement with anomalous scattering; STI, trypsin inhibitor from soybean *Glycine max*; WBA, major albumin from winged bean *Psophocarpus tetragonolobus*; WCI, chymotrypsin inhibitor from winged bean *P. tetragonolobus*.

* Corresponding author. Tel.: +55-16-273-8088; fax: +55-16-273-9881.

E-mail address: ipolikarpov@ifsc.usp.br (I. Polikarpov).

quick cryo-soaking method [9,10] using sodium iodide and cesium chloride in 0.5 and 1.0 M concentrations, respectively.

2.2. Data collection and processing

The native and derivative data sets were collected on the protein crystallography beamline [12] at the Brazilian National Synchrotron Light Laboratory (Campinas, SP, Brazil), using a MAR345 image plate. The synchrotron-radiation wavelength was set to 1.54 Å to maximize both X-ray flux and the anomalous signal [13]. The diffraction data were collected with a high multiplicity and then auto-indexed and integrated with the program DENZO [14] (Table 1). The quality of each data set and the presence of the anomalous signal of cesium and iodine atoms were monitored following $|\Delta F^{\text{ANO}}|/F$ and $|\Delta F^{\text{ANO}}|/\sigma(\Delta F^{\text{ANO}})$ ratios as a function of resolution.

2.3. Phasing

Seven cesium sites were located by direct methods. Scaled intensities were input into SnB 2.1 package [15], and normalized anomalous differences E (where

$E = \Delta F^{\text{ANO}}/\sqrt{\langle (\Delta F^{\text{ANO}})^2 \rangle}$ for 151) were calculated with DREAR [16]. The bimodal distribution of the R_{min} histogram was used to identify the correct solution. The obtained heavy-atom positions were initially refined with the CNS package [17] using anomalous and isomorphous differences, and were then used as an input to SHARP [18] for phase calculations. Cs-SIRAS derived phases were submitted to density modification protocols. The resulting phases were used to calculate anomalous and isomorphous difference Fourier maps. In addition to seven cesium cations, nine iodine anions were found in the asymmetric unit of I derivative. All 16 heavy-atom sites were included in MIRAS phase calculation with SHARP. An initial electron density map was finally obtained after density modification with SOLOMON [19].

2.4. Model building and refinement

The MIRAS electron density map calculated with SHARP/SOLOMON [18,19] was used as an input to ARP/wARP [20] to build a hybrid model of the inhibitor. A total of 20 automatic building cycles were performed, tracing up to 92% of the structure. The constructed model consisted of 148 residues distributed in six chains with a few main-chain discontinuities between them. Discontinuities were

Table 1

Details of crystallographic data collection and refinement statistics. Values for the highest resolution shell are shown in parentheses

	Native	I-derivative	Cs-derivative
Space group	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2
Unit-cell parameters (Å)	$a = b = 58.71$ $c = 93.75$	$a = b = 58.42$ $c = 93.91$	$a = b = 58.33$ $c = 93.80$
Resolution range (Å)	25.3 – 1.83 (1.87 – 1.83)	25.3 – 1.92 (1.96 – 1.92)	22.9 – 2.00 (2.05 – 2.00)
Number of observations	123 604	2 85 176	3 12 518
Number of unique observations ^a	15 024	23 500	20 853
$\langle I/\sigma(I) \rangle$	23.6 (3.3)	33.7 (13.9)	34.1 (13.1)
Mosaicity (°)	0.57	0.42	0.53
Multiplicity ^a	8.2 (5.4)	12.1 (12.0)	15.0 (14.7)
Data collected (°)	130	292	360
Completeness (%)	99.7 (97.6)	99.5 (93.9)	99.8 (99.5)
R_{merge} (%) ^b	8.9 (50.8)	7.3 (17.0)	9.3 (23.8)
Refinement statistics			
Number of protein atoms	1449		
Number of waters molecules	183		
Number of reflections used in refinement (working/test data sets)	13 661/1363		
R_{factor} (%) ^c	17.3		
R_{free} (%) ^d	20.3		
Average B_{factor}			
Main-chain (Å ²)	20.1		
Side-chains (Å ²)	22.4		
Waters (Å ²)	39.4		
R.m.s. deviations from ideal geometry			
Bonds (Å)	0.019		
Bonds angles (°)	1.6		

^a Multiplicities of derivative data sets were calculated with Friedel-related reflections treated separately. Multiplicity of the native data set was calculated with Friedel-related reflections treated as equivalent.

^b $R_{\text{merge}} = \sum_{hkl} |I - \langle I \rangle| / \sum_{hkl} I$.

^c $R_{\text{factor}} = \sum (|F_{\text{obs}}| - |F_{\text{calc}}|) / \sum F_{\text{obs}}$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

^d R_{free} was calculated using 1363 reflections randomly selected in a test data set.

filled with water molecules clearly showing amino acid residues structures. The model was further improved using the program O [21]. The amino acid residues identity was deduced from the experimental electron density maps. The presence of cysteines and methionines residues was checked using anomalous difference Fourier maps (ΔF^{ANO} , $\phi_{\text{calc}} - 90^\circ$). Two methionines, A72 and A90, and two half-cystines, A40 and A84, forming the only disulfide bridge of CTI, were found. Furthermore, CTI does not contain free cysteines.

Simultaneously and independently, the amino acids sequence was determined using 2D-PAGE, MALDI-TOF and ESI mass spectrometry (S. Marangoni, personal communication). Comparison of the amino acid sequence and the putative sequence deduced from the electron density maps showed that three-fourths of the total sequence was correctly identified from visual inspection of the maps.

The model was refined with CNS package [17] and REFMAC5 [19] using standard protocols. The final crystallographic R_{factor} is 17.3% for 13 661 unique reflections in the working data set between 25.3 and 1.83 Å and R_{free} is 20.3% for 1363 unique reflections in the test data set within the same resolution range.

3. Results

3.1. Quality of the refined model

The data in Table 1 show a well refined structure with excellent stereochemistry. The CTI model is very well defined by electron density. All main-chain dihedral angles are within the allowed regions of the Ramachandran plot [22]. The loop regions, including the reactive loop, display elevated B-values as compared to the core of the molecule.

3.2. Overall architecture

CTI is all beta protein with a β -trefoil fold (Fig. 1a) formed by 12 antiparallels β -strands with long interconnecting loops [23,24]. The protein could be divided into three approximately equal subdomains related by a pseudo-threefold axis roughly parallel to the barrel axis of the trefoil (Fig. 1b). Each of the subdomains consists of about 60 amino acids, half of which form four β -strands (β_m , $m=1,2,3,4$) and the other half form the loops (L^m) that connect them. Each subdomain is structurally organized as $L_n^1\beta_n^1L_n^2\beta_n^2L_n^3\beta_n^3L_n^4\beta_n^4$. The superposition performed using the 30 topologically equivalent C^α atoms that compose the β -strands of each repetition unit result in the r.m.s. deviation of 1.5 Å between subdomains T_1 (residues A1–A60) and T_2 (residues A61–B120); 1.9 Å between subdomains T_1 and T_3 (residues B121–B167) and 1.7 Å between subdomains T_2 and T_3 . Although the subdomains do not demonstrate any discernible amino acid similarity, it was suggested that the β -trefoil fold arose from a homotrimer-forming progenitor repeat, which underwent successive gene duplication events [25].

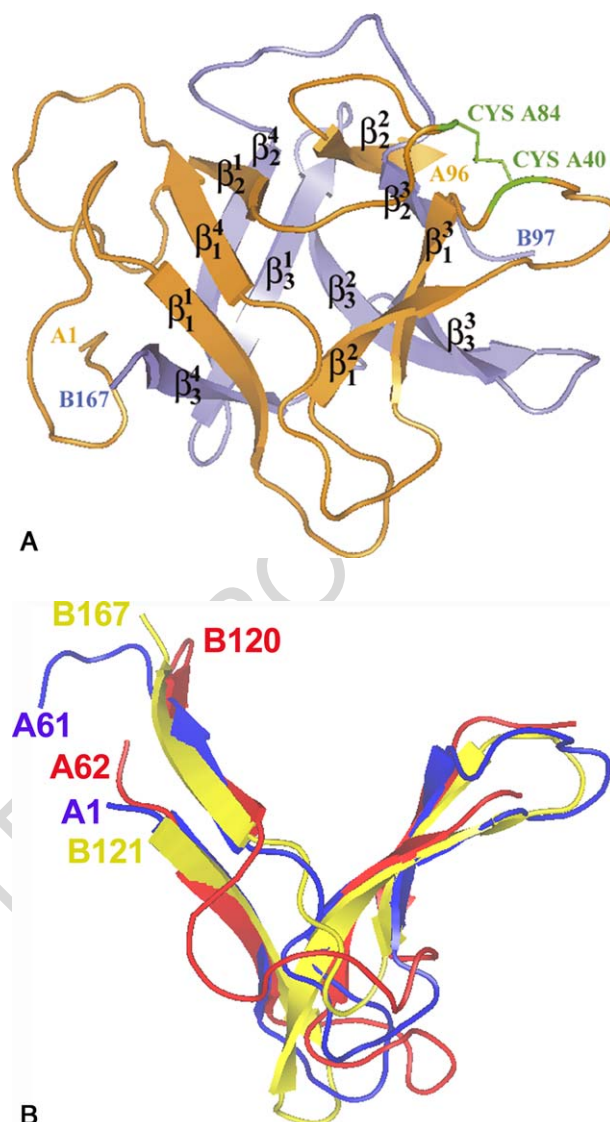


Fig. 1. (a) The overall β -trefoil fold of CTI. The chain A (residues A1–A96) painted in orange and the chain B (residues B97–B167) painted in blue. A single disulfide bridge between residues A40 and A84 is shown in green. All β -strands are numbered. (b) Superposition of the subdomains T_1 – T_3 , painted in blue, red and yellow, respectively.

3.3. CTI is composed of two polypeptide chains

The CTI crystallographic model consists of two noncovalently bound polypeptide chains; A (residues A1–A96) and B (residues B97–B167) (Fig. 1). The break between chains A and B is clearly defined by electron density. The existence of the two noncovalently bound polypeptide chains explains the presence of two bands in the SDS-PAGE gel in denaturing conditions [11]. The N-terminus of the 11 kDa band shows high homology with the members of Kunitz (STI) family, but not the N-terminus 9 kDa band that led to erroneous identification of CTI as two separate inhibitors [8]. It also explains why the attempts to crystallize each part of the inhibitor separately were unsuccessful.

3.4. There is only one disulfide bridge in CTI

One of the distinct features of the Kunitz (STI) family is the presence of two disulfide bridges that are conserved in almost all members of the family. In contrast, CTI contains only one disulfide bridge located between the residues A40 and A84. The existence of a single disulfide bridge in the structure of CTI was confirmed by anomalous difference Fourier maps (results not shown).

Amino acid sequence comparisons of CTI with homologous proteins performed by BLASTp (<http://www.ncbi.nlm.nih.gov>) reveal that the first disulfide bond is strictly conserved in all active protein inhibitors. It is located between the end of the β -strand of T2 repeat and the beginning of the first β -strand of T3 repeat. This disulfide bridge fastens the reactive site loop and contributes to the maintenance of the canonical conformation necessary to the inhibitory activity.

The second disulfide bond, absent in CTI, is conserved in the majority of the Kunitz-type inhibitors. However, there are several Kunitz-type inhibitors with either none or two disulfide bonds in the corresponding region of the polypeptide chain. These observations indicate that the second disulfide bond appears not to be essential for inhibitory activity. Moreover, biochemical studies have demonstrated that STI retains the inhibitory activity after the reduction of the second disulfide bridge but it becomes inactive in its fully reduced state [26]. Reduction of two disulfide bonds of ETI markedly decreases its inhibitory activity [27]. These facts corroborate with the hypothesis that the presence of the first disulfide bond is essential for the inhibitory activity whereas the presence of a second disulfide bond might be not.

3.5. The reactive loop

The reactive site loop of CTI is located in the chain A and is formed by the residues Phe59–Lys60–Ala61–Ser62–Pro63–Arg64–Ser65–Lys66–Tyr67–Ile68–Ser69 (from P6 to P5' position; Fig. 2a, b). Arg64 occupies P1 position. CTI reactive loop adopts canonical conformation similar to one of the standard Kunitz-type inhibitors like STI [28], ETI [29] and WCI [30].

3.6. Comparison with Kunitz (STI) family members

The three-dimensional structure of CTI exhibits the same structural fold that was found for other Kunitz (STI) family members. Most of the deviations in the conformation of their polypeptide chains are confined to the surface loop regions.

ETI is the Kunitz (STI) family member that presents the highest sequence identity, 31%, with CTI. The r.m.s. deviation between 153 superimposed C α atoms of their polypeptide chains is 1.57 Å. STI and WCI have a sequence identity with CTI of 21% and 23% and the r.m.s. deviations of 1.15 Å (for 147 C α atoms) and 1.51 Å (for 156 C α atoms), respectively. Although CTI reactive loop adopts a conformation

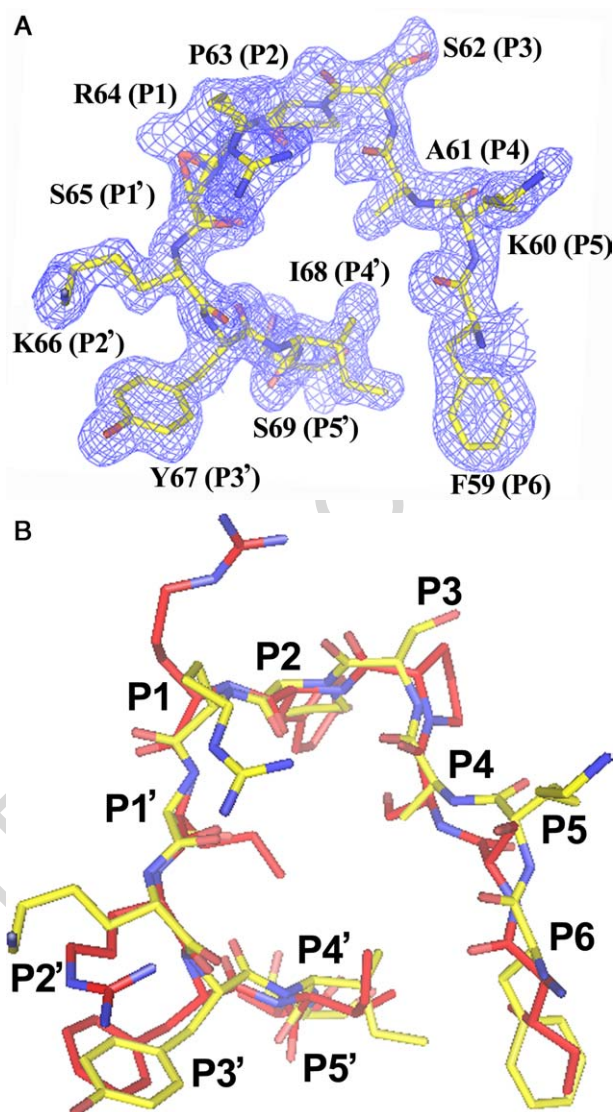


Fig. 2. (a) The CTI reactive site loop (residues PheA59–SerA69) superimposed with the final ($2mF_{\text{obs}} - DF_{\text{calc}}$, ϕ_{calc}) electron density map contoured at 1.2σ (given as a blue mesh). (b) The superposition of CTI and a standard Kunitz type trypsin inhibitor, STI reactive loops (in yellow and red, respectively).

very similar to the canonical, the global comparison shows that the sequential and structural agreement is better within STI, ETI and WCI group, than between these proteins and CTI. STI has a sequence identity of 38% with ETI, with a r.m.s. deviation of 1.21 Å for 149 C α atoms, and of 40% with WCI, with a r.m.s. deviation of 1.38 Å for 151 C α atoms superimposed; ETI and WCI have the best agreement with a sequence identity of 53% and a r.m.s. deviation of 0.93 Å for 158 C α atoms superimposed.

DrTI is a new member of the Kunitz (STI) family which has an extra amino acid residue in the reactive loop between residues P1 and P2 [31]. Although this insertion distorts the reactive loop causing it to lose the canonical conformation, a r.m.s. deviation of the superposition of CTI and DrTI is 1.57 Å for 146 C α atoms superimposed. Amino acid sequence identity of these two proteins is 25%.

The largest deviations are observed in comparison of CTI with another family member, WBA [32], which has 19% of sequence identity with CTI. The r.m.s. deviation between WBA and CTI is 1.69 Å for 137 best superimposed C α atoms.

4. Discussion

Trypsin is capable of cleaving peptides at nearly every lysine-X and arginine-X position (with the exception of proline at X). Therefore, potentially about 5–10% of the peptide bonds in proteins are susceptible to its proteolytic attack. If this is so, why is it unable to efficiently cleave trypsin inhibitors?

Structural and biochemical studies show that its attempts to cleave inhibitors result in a formation of trypsin-inhibitor complex, in which the inhibitor reactive loop binds to the enzyme active site in a “perfect substrate” conformation and is cleaved, but not released from the active site of the enzyme. Binding of a cleaved inhibitor in the pocket of enzyme is tight and oriented; this prevents acyl-enzyme hydrolysis and favors the inverse reaction with the relegation of the leaving group, that dramatically slows down (up to 10⁷ times) the whole reaction [33]. The canonical conformation of the reactive loop, thus, is very important for a suitable positioning and an effective adaptation to a proteolytic enzyme active site.

Like the canonical inhibitors from the Kunitz (STI) family, CTI has a high inhibitory activity against trypsin ($K_i = 1.2$ nM) [8]. In order to understand the structural basis of this phenomenon, we superimposed the structure of CTI with the structures of STI (PDB ID 1AVU) and STI complexed with porcine trypsin (PDB ID 1AVW). These comparisons demonstrate that CTI reaction loop adopts canonical conformation very similar to the one of STI (Fig. 2b). In addition, most of the trypsin-inhibitor contacts are conserved in the putative CTI:trypsin complex (results not shown).

It is tempting to suggest that the high inhibitory activity of CTI against trypsin is related to a conservation of the canonical conformation of its reactive loop. Furthermore, potential negative effects of a break in a polypeptide chain and the lack of one disulfide bridge seems to be offset by a structural integrity of the protein molecule locked in a canonical conformation by a dense network of hydrogen bonds.

5. Protein data bank

Atomic coordinates have been deposited in the Protein Data Bank (<http://www.rcsb.org/pdb/>) for release upon publication. The PDB entry corresponding to the structure is 1R8O.

Acknowledgements

We thank Professor Richard C. Garrat, Valéria Peyrl Forrer, Renata Carmona e Ferreira and José Ribeiro Brandão

Neto for their scientific assistance. Financial support from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP grants 99/03387-4, 99/04497-8 and 00/02317-1) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) are acknowledged.

References

- [1] W. Bode, R. Huber, Proteinase–protein inhibitor interaction, *Biomed. Biochim. Acta* 50 (1991) 437–446.
- [2] M. Laskowsky Jr, I. Kato, Protein inhibitors of proteinases, *Annu. Rev. Biochem.* 49 (1980) 593–626.
- [3] R.M. Broadway, S.S. Duffey, The effect of plant protein quality on insect digestive physiology and the toxicity of plant proteinase inhibitors, *J. Insect. Physiol.* 34 (1988) 1111–1117.
- [4] J. Yavelow, T.H. Finlay, A.R. Kennedy, W. Troll, Bowman-Birk soybean protease inhibitor as an anticarcinogen, *Cancer Res.* 43 (1983) 2454s–2459s.
- [5] P.C. Billings, P.M. Newberne, A.R. Kennedy, Protease inhibitor suppression of colon and anal gland carcinogenesis induced by dimethylhydrazine, *Carcinogenesis* 11 (1990) 1083–1086.
- [6] P.C. Billings, A.R. Morrow, C.A. Ryan, A.R. Kennedy, Inhibition of radiation-induced transformation of C3H/10T1/2 cells by carboxypeptidase inhibitor I and inhibitor II from potatoes, *Carcinogenesis* 10 (1989) 687–691.
- [7] A. Wlodawer, J. Deisenhofer, R. Huber, Comparison of two highly refined structure of bovine pancreatic trypsin inhibitor, *J. Mol. Biol.* 193 (1987) 145–156.
- [8] J.A. Silva, M.L.R. Macedo, J.C. Novello, S. Marangoni, Biochemical characterization and N-terminal sequences of two new trypsin inhibitors from *Copaifera langsdorffii* seeds, *J. Prot. Chem.* 20 (2001) 1–7.
- [9] Z. Dauter, M. Dauter, K.R. Rajashankar, Novel approach to phasing proteins: derivatization by short cryo-soaking with halides, *Acta Crystallogr D* 56 (2000) 232–237.
- [10] R.A.P. Nagem, Z. Dauter, I. Polikarpov, Protein crystal structure solution by fast incorporation of negatively and positively charged anomalous scatterers, *Acta Crystallogr D* 57 (2001) 996–1002.
- [11] S. Krauchenco, J.A. Silva, R.A.P. Nagem, J.R. Brandão Neto, V.P. Forrer, R. Carmona e Ferreira, M.L.R. Macedo, J.C. Novello, S. Marangoni, I. Polikarpov, Crystallization and preliminary X-ray diffraction analysis of a novel trypsin inhibitor from seeds of *Copaifera langsdorffii*, *Acta Crystallogr D* 57 (2001) 1316–1318.
- [12] I. Polikarpov, L.A. Perles, R.T. de Oliveira, G. Oliva, E.E. Castellano, R. Garrat, A. Craievich, Set-up and experimental parameters of the protein crystallography beamline at the Brazilian National Synchrotron Laboratory, *J. Synchrotron Rad.* 5 (1998) 72–76.
- [13] I. Polikarpov, A. Teplyakov, G. Oliva, The ultimate wavelength for protein crystallography? *Acta Crystallogr D* 53 (1997) 734–737.
- [14] Z. Otwinowski, W. Minor, Processing of X-ray data collected in oscillation mode, *Methods Enzymol.* 276 (1997) 307–326.
- [15] C.M. Weeks, R. Miller, The design and implementation of SnB v2.0, *J. Appl. Crystallogr.* 32 (1999) 120–124.
- [16] R.H. Blessing, G.D. Smith, Difference structure factor normalization for heavy-atom or anomalous-scattering substructure determinations, *J. Appl. Crystallogr.* 32 (1999) 664–670.
- [17] A.T. Brünger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, Crystallography and NMR system: a new software suite for macromolecular structure determination, *Acta Crystallogr D* 54 (1998) 905–921.
- [18] E. de La Fortelle, G. Bricogne, Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods, *Methods Enzymol.* 276 (1997) 472–494.

- 337 [19] Collaborative Computational Project, Number 4, the CCP4 suite:
338 programs for protein crystallography, *Acta Crystallogr D50* (1994)
339 760–763.
- 340 [20] A. Perrakis, T.K. Sixma, K.S. Wilson, V.S. Lamzin, wARP: improve-
341 ment and extension of crystallographic phases by weighted averaging
342 of multiple refined dummy atomic models, *Acta Crystallogr D53*
343 (1997) 448–455.
- 344 [21] T.A. Jones, M. Bergdoll, M. Kjeldgaard, in: C. Bugg, S. Ealick (Eds.),
345 *Crystallographic and Modelling Methods in Molecular Design*,
346 Springer-Verlag Press, 1990, pp. 189–195.
- 347 [22] G.N. Ramachandran, V. Sasisekharan, Stereochemistry of polypep-
348 tide chain configurations, *Adv. Protein Chem* 23 (1968) 283–437.
- 349 [23] A.G. Murzin, A.M. Lesk, C. Chothia, Beta-trefoil fold, *J. Mol. Biol*
350 223 (1992) 531–543.
- 351 [24] A.D. McLachlan, Three-fold structural pattern in the soybean trypsin
352 inhibitor (Kunitz), *J. Mol. Biol* 133 (1979) 557–563.
- 353 [25] D. Mukhopadhyay, The molecular evolutionary history of a winged
354 bean alpha-chymotrypsin inhibitor and modelling of its mutations
355 through structural analyses, *J. Mol. Evol* 50 (2000) 214–223.
- 356 [26] F.P. DiBella, I.E. Liener, Soybean trypsin inhibitor: cleavage and
357 identification of a disulfide bridge not essential for activity, *J. Biol.*
358 *Chem* 244 (1969) 2824–2829.
- [27] K. Lehle, A. Wrba, R. Jaenicke, Erythrina trypsin inhibitor retains its
native structure and function after reducing its disulfide groups, *J.*
Mol. Biol 239 (1994) 276–284.
- [28] H.K. Song, S.W. Suh, Kunitz-type soybean trypsin inhibitor revisited:
refined structure of its complex with porcine trypsin reveals an insight
into the interaction between a homologous inhibitor from *Erythrina*
caffra and tissue-type plasminogen activator, *J. Mol. Biol* 275 (1998)
347–363.
- [29] S. Onesti, P. Brick, D.M. Blow, Crystal structure of a Kunitz-type
trypsin inhibitor from *Erythrina caffra* seeds, *J. Mol. Biol* 217 (1990)
153–176.
- [30] S. Ravichandran, U. Sen, C. Chakrabarti, J.K. Dattagupta, Cryocrys-
tallography of a Kunitz-type serine protease inhibitor: the 90 K struc-
ture of a winged bean chymotrypsin inhibitor (WCI) at 2.13 Å reso-
lution, *Acta Crystallogr D55* (1999) 1814–1821.
- [31] S. Krauchenco, S.C. Pando, S. Marangoni, I. Polikarpov, Crystal
structure of the Kunitz (STI) type inhibitor from *Delonix regia* seeds,
Biochem. Biophys. Res. Commun 312 (2003) 1303–1308.
- [32] A.J. McCoy, A.A. Kortt, The 1.8 Å crystal structure of winged bean
albumin 1, the major albumin from *Psophocarpus tetragonolobus*
(*L.*) *DC*, *J. Mol. Biol* 269 (1997) 881–891.
- [33] E.S. Radisky, D.E. Koshland Jr, A clogged gutter mechanism for
protease inhibitors, *Proc. Natl. Acad. Sci. USA* 99 (2002) 10316–
10321.

359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382