Estrutura cristalográfica do inibidor de tripsina
purificado de sementes de *Enterolobium contortisiliquum*
e modelagem molecular de seus complexos
com tripsina, trombina e fator Xa

Marcos Roberto Bonfadini

Tese apresentada ao Instituto de Física
de São Carlos, da Universidade de São Paulo
para a obtenção do Título de Doutor
em Ciências: Física Aplicada

Orientador: Prof. Dr. Richard Charles Garratt

São Carlos - 2003

Bonfadini, Marcos Roberto

“Estrutura cristalográfica do inibidor de tripsina purificado de sementes de *Enterolobium contortisiliquum* e modelagem molecular do inibidor complexado com tripsina, trombina e fatorXa”

Marcos Roberto Bonfadini - São Carlos, 2003.

Tese (Doutorado) -- Instituto de Física de São Carlos,
2003 – Páginas: 130

Orientador: Prof. Dr. Richard Charles Garratt

1. cristalografia de raios X. 2. modelagem molecular
3. inibidores de tripsina 4. motivo β -*trefoil*.

I. Título

FOLHA DE APROVAÇÃO

Membros da Comissão Julgadora da Tese de Doutorado de _____, apresentada ao Instituto de Física de São Carlos, da Universidade de São Paulo, em ____/____/____.

Comissão Julgadora:

(Assinatura)

(Nome/Instituição)

(Assinatura)

(Nome/Instituição)

(Assinatura)

(Nome/Instituição)

(Assinatura)

(Nome/Instituição)

(Assinatura)

(Nome/Instituição)

DEDICATÓRIA

Dedico este trabalho a Tatiana.

She
May be the face I can't forget
A trace of pleasure or regret
May be my treasure or the price I have to pay
She may be the song that summer sings
May be the chill that autumn brings
May be a hundred different things
Within the measure of a day.
She
May be the beauty or the beast
May be the famine or the feast
May turn each day into a heaven or a hell
She may be the mirror of my dreams
A smile reflected in a stream
She may not be what she may seem
Inside her shell
She who always seems so happy in a crowd
Whose eyes can be so private and so proud
No one's allowed to see them when they cry
She may be the love that cannot hope to last
May come to me from shadows of the past
That I'll remember till the day I die
She
May be the reason I survive
The why and wherefore I'm alive
The one I'll care for through the rough and many years
I'll take her laughter and her tears
And make them all my souvenirs
For where she goes I've got to be
The meaning of my life is
She, she, she

Dedico também aos meus pais Demir e Cleide.

AGRADECIMENTOS

Agradeço a todos que de uma forma ou de outra, dentro ou fora de minha vida profissional e ao longo dos anos, contribuíram para a concretização deste trabalho. Em especial à orientação do Prof. Dr. Richard Charles Garratt.

SUMÁRIO

LISTA DE FIGURAS	iii
LISTA DE TABELAS	vii
RESUMO	ix
ABSTRACT	x
1 Introdução.....	1
1.1 Enzimas.....	3
1.1.1 Serinoproteases	6
1.2 Inibidores de Serinoproteases	17
1.2.1 Família SBTI.....	24
1.3 Objetivos.....	28
2 Materiais e Métodos	32
2.1 Cristalização e Coleta de Dados Cristalográficos.....	32
2.2 Resolução da Estrutura	37
2.3 Refinamento da Estrutura Cristalina.....	38
2.4 Interpretação Automática da Densidade Eletrônica: ARP/wARP ..	42
2.5 Modelagem Molecular por Homologia	44
2.5.1 Modelagem dos Complexos	46
2.6 Critérios de Qualidade	48
2.6.1 PROCHECK	48
2.6.2 WHATCHECK	50
2.6.3 VERIFY-3D.....	51
3 Resultados e Discussão	54

3.1	Cristalização, Coleta e Processamentos de Dados.....	54
3.2	Resolução da Estrutura	57
3.3	Refinamento da Estrutura.....	58
3.4	Avaliação da Qualidade da Estrutura com Seqüência Modificada	65
3.5	Descrição da Estrutura	69
3.5.1	O Barril	74
3.5.2	Os Grampos de Cabelo.....	78
3.5.3	Aspectos Inéditos do Enovelamento	79
3.5.4	Os Arcos do Motivo <i>Trefoil</i>	84
3.5.5	Águas	89
3.5.6	Conformação do <i>Loop</i> Reativo	90
3.5.7	Pontes de Dissulfeto.....	96
3.6	Modelagem da Seqüência Proposta.....	96
3.7	Modelagem dos Complexos Entre EcTI e Serinoproteases	100
3.7.1	Complexo EcTI/Tripsina	100
3.7.2	Complexo EcTI/Trombina.....	103
3.7.3	Complexo EcTI/Fator Xa	108
3.7.4	Complexo EcTI/Quimotripsina.....	109
3.7.5	tPA e Fator XIIa	110
4	Conclusões	113
5	Referências Bibliográficas	119

LISTA DE FIGURAS

Figura 1: Perfil de reação enzimática.	3
Figura 2: Passos do mecanismo de ação enzimática, acilação (em cima) e desacilação (embaixo).....	8
Figura 3: Representação em vista estéreo da ligação éster, entre o C da carbonila da Ile7 (posição P1 do substrato) e O γ da Ser195.....	9
Figura 4: Tripsina porcina (verde) complexada com inibidor SBTI (azul) (PDB 1AVW).....	11
Figura 5: Tripsina representada como superfície sólida complexada com SBTI (em amarelo)	12
Figura 6: Sítio ativo da tripsina em complexo com substrato (PDB 2PTC)	12
Figura 7: Cascata de coagulação sangüínea	14
Figura 8: Molécula de trombina, com suas cadeias leve (vermelho) e pesada (verde) complexada com inibidor (azul)	14
Figura 9: Fator de coagulação Xa, cadeia leve (vermelho) e pesada (verde) complexado com inibidor não protéico (azul), (PDB 1EZQ)	16
Figura 10 : Esquema ilustrando a inibição não competitiva.....	18
Figura 11 : Esquema ilustrando a inibição competitiva.....	19
Figura 12: Aspectos do <i>Enterolobium contortisiliquum</i>	26
Figura 13: Ilustração do arranjo experimental da técnica de acupuntura em gel.....	34

Figura 14: Ilustração do arranjo experimental montado para o uso da técnica de acupuntura em gel	35
Figura 15: Fluxograma do procedimento warpNtrace do programa ARP/wARP.	43
Figura 16: Cristal de EcTI crescido com a técnica de acupuntura em gel, na condição 15% PEG8000, 50mM LiSO ₄	54
Figura 17: Algumas divergências evidentes no mapa de densidade eletrônica encontradas durante a construção do modelo de EcTI.	61
Figura 18: Espacialmente próximo ao Trp94 (fortemente conservado em outros inibidores) está a Phe117.	62
Figura 19: Diagrama de Ramachandran para EcTI	65
Figura 20: a) Posição relativa ao <i>loop</i> reativo (vermelho) dos resíduos Ala81, Lys124, Asp126 presentes em regiões generosamente permitidas no diagrama de Ramachandran.....	67
Figura 21: Parâmetros de cadeia principal para EcTI fornecidos por PROCHECK.	68
Figura 22: Alinhamento estrutural com C α entre ETI (verde), SBTI (vermelho) e EcTI (azul)	70
Figura 23: O arranjo β - <i>trefoil</i> (direita) é formado a partir de três unidades <i>trefoil</i> (esquerda)	71
Figura 24: a) Sobreposição dos três motivos <i>trefoil</i> de EcTI	72
Figura 25: Vistas dos resíduos que compõem as camadas do barril do EcTI.....	77
Figura 26: As duas fitas de cada grampo de cabelo se espiralam uma em torno da outra formando o <i>coiled-coil</i>	78

Figura 27: Esquema mostrando o arranjo dos grampos de cabelo	79
Figura 28: O número de resíduos por volta (n)	80
Figura 29: Distribuição dos ângulos ϕ e ψ dos resíduos das fitas dos grampos de cabelo e fitas do barril num diagrama de Ramachandran.....	82
Figura 30: a) Localização relativa do arco no motivo <i>trefoil</i>	86
Figura 31: Alinhamento múltiplo de seqüências de domínios β - <i>trefoil</i> mostrando as três repetições internas.....	87
Figura 32 : Vista estéreo do "arco" do segundo grampo de cabelo. A cadeia lateral da Pro93 ocupa o vão entre a fita e o arco.	87
Figura 33: Vista estéreo do "arco" do terceiro motivo	88
Figura 34: <i>Loop</i> do sítio reativo de EcTI com o mapa de densidade eletrônica.....	91
Figura 35: “Malha” de ligações de hidrogênio realizadas pela Asn12...92	
Figura 36: Diagrama HERA (Hutchinson & Thornton, 1990b) das estruturas secundárias encontradas em EcTI	93
Figura 37: Distribuição do fator de temperatura médio para os átomos da cadeia principal.....	95
Figura 38: Faixas dos ângulos ϕ e ψ adotados pela cadeia principal na região P3-P3' para inibidores livres e inibidores complexados com proteinases.....	96
Figura 39: Distribuição das energias dos 50 modelos gerados na modelagem molecular por MODELLER	98
Figura 40: Diagrama de Ramachandran global do modelo 1	99
Figura 41: Perfil da verificação feita pelo VERIFY 3D em mod1	99

Figura 42: Perfil VERIFY-3D para o complexo tripsina/mod1	101
Figura 43: O perfil VERIFY-3D para o complexo trombina/EcTI.....	104
Figura 44: Sobreposição das estruturas de trombina (laranja) e tripsina (verde)	105
Figura 45: Superfície de potencial eletrostático gerada pelo programa GRASP em a) e b) Ilustrando a complementaridade de carga na interface do complexo EcTI/trombina	107
Figura 46: Perfil VERIFY-3D do complexo fator Xa/EcTI.....	108
Figura 47: A Arg63 do EcTI (em azul) poderia assumir uma conformação equivalente à Lys em BPTI (em amarelo).....	110

LISTA DE TABELAS

Tabela 1: Algumas serinoproteases, fonte e função.....	7
Tabela 2 : Famílias de inibidores protéicos que inibem serinoproteases e obedecem ao mecanismo padrão (Bode & Huber, 1992).....	20
Tabela 3: Medidas experimentais da constante de inibição K.....	29
Tabela 4: Alinhamento entre os inibidores de <i>E.caffra</i> (1TIE), de soja (1BA7) e EcTI.....	38
Tabela 5: Estatísticas da coleta de dados e processamento do cristal do EcTI.....	55
Tabela 6: Estatística da coleta de dados de difração de raios-X por faixa de resolução para EcTI.	56
Tabela 7: Contraste entre os coeficientes de correlação das soluções encontradas por AMoRE	58
Tabela 8: Alinhamento entre as seqüências esperada (seq.propos) e a efetivamente encontrada (EcTI)	60
Tabela 9: Alinhamento entre sete diferentes inibidores de protease da família Kunitz.....	73
Tabela 10: Alinhamento BLAST contendo inibidores Kunitz.....	83
Tabela 11: Alinhamento seqüencial dos <i>loops</i> reativos entre três inibidores de tripsina da família Kunitz de estrutura conhecida.....	91
Tabela 12: Fator de temperatura médio por resíduo encontrado nos <i>loops</i> de EcTI.....	93
Tabela 13: Cinco melhores modelos em 50 com <i>G-factor</i> calculado por PROCHECK a 2,0Å.....	98

Tabela 14: As ligações de hidrogênio efetuadas pelo <i>loop</i> reativo no complexo tripsina/EcTI	103
Tabela 15: Inserção na região do resíduo 60 em tPA e fator XIIa	111

RESUMO

Esse trabalho teve como objetivo a determinação da estrutura tridimensional do inibidor de tripsina purificado de sementes de *Enterolobium contortisiliquum* (EcTI) e a modelagem molecular dos complexos EcTI-tripsina, EcTI-quimotripsina, EcTI-trombina, EcTI-fator Xa, e suas análises. O EcTI possui 19kDa de peso molecular inibe tripsina, quimotripsina, calicreína plasmática humana, plasmina humana e fator XIIa, mas não inibe fator Xa e ativador de plasminogênio. Monocristais apropriados à coleta de dados de difração de raios-X foram obtidos na condição 50 do fatorial da Hampton (PEG8000 – 0,5M, LiSO₄ - 0,5M) através da técnica de acupuntura em gel à temperatura constante de 15°C após três semanas. O grupo espacial encontrado foi o P2₁ com uma molécula na unidade assimétrica. Os dados de difração foram coletados num detetor de placas de imagem MAR345 na linha de Cristalografia de Proteínas (PCr) no Laboratório Nacional de Luz Síncrotron (LNLS) em Campinas até uma resolução de 1,96 Å, completude de 87% e R_{sym}=10,3%. A estrutura foi resolvida por Substituição Molecular tendo-se como molde a estrutura do inibidor de tripsina proveniente de *Erythrina caffra* que apresenta 39% de identidade seqüencial com a seqüência primária do EcTI. Divergências entre a seqüência primária esperada e a observada no mapa de densidade eletrônica levaram à suposição da existência de isoformas para este inibidor. Acatando esta hipótese e supondo que a isoforma cristalizada era diferente da isoforma cuja seqüência foi publicada, mudanças na seqüência guiadas pelo mapa de densidades foram realizadas. Um total de 33 substituições e 1 inserção foram feitos. Diferentemente das tentativas de refinamento com a seqüência original, após as substituições o refinamento convergiu para valores aceitáveis de R_{factor}=17% e R_{free}=25%. A estrutura apresenta o enovelamento β -trefoil com uma pseudo simetria de ordem 3. Os grampos de cabelo presentes na estrutura são espiralados mas isto não é resultado de repetições sistemáticas nos ângulo ϕ/ψ como esperado. Baseado nesta estrutura cristalográfica, um modelo da seqüência original foi construído por modelagem por homologia. A avaliação do modelo em termos de estereoquímica e compatibilidade estrutura-seqüência mostra uma boa qualidade, mesmo existindo a formação de uma ponte salina enterrada no núcleo hidrofóbico da molécula. Estruturas cristalográficas de complexos entre tripsina, quimotripsina, fator Xa e trombina foram usados para gerar modelos de complexos entre as serinoproteases correspondentes e o EcTI. A falta de atividade de EcTI contra fator Xa e trombina pode ser atribuída à incompatibilidade química e estrutural na interface do complexo.

ABSTRACT

The principal objective of this work was the 3D structural determination of a Trypsin inhibitor from *Enterolobium contortisiliquum* (EcTI) seeds as well as the modeling of the complexes between EcTI and trypsin, chymotrypsin, thrombin and factor Xa and their subsequent analysis. The inhibitor has a molecular weight of 19.5 kDa and inhibits trypsin, chymotrypsin, human plasma kallikrein, human plasmin and factor XIIa but not factor Xa and tissue type plasminogen activator. A single crystal appropriate for X-ray data collection was grown using condition 50 of Hampton's Research factorial (PEG8000 - 15%, LiSO₄ - 0.5mM) by the gel acupuncture technique at 15 °C. The crystal grew in three weeks in the space group P2₁ with one molecule in the asymmetric unit. Crystallographic data were acquired using an image plate detector MAR345 on the Protein Crystallography beam line at the Laboratório Nacional de Luz Síncrotron (LNLS), Campinas, out to 1.96 Å resolution, with a completeness of 87% and R_{sym} = 10.3%. The structure was solved by the Molecular Replacement method using the previous determined structure of the trypsin inhibitor from *Erythrina caffra* which presented 39% sequence identity to EcTI. Discrepancies in the primary structure, as observed in the electron density map led to the supposition of the existence of isoforms for this particular inhibitor. In total 33 substitutions and 1 insertion were made. In contrast with the attempts made with the original sequence, after the amino acids were substituted the refinement converged to acceptable values for R_{factor}=17% and R_{free}=25%. The 3D fold of EcTI is a β -*trefoil* as expected. The hairpins present in this fold are coiled coils, but not as a result of the systematic alternation of the ϕ/ψ angles as expected. Using the crystallographic structure as a template a homology model for the original sequence was built. The model evaluation in terms of stereochemistry and structure-sequence compatibility showed it to be of generally good quality, despite the presence of a salt bridge buried in the hydrophobic core. Docking experiments using crystallographic structures of trypsin, chymotrypsin, factor Xa and thrombin complexed with their respective inhibitors were undertaken. The lack of EcTI activity against factor Xa and thrombin is predicted to be due to chemical and structural incompatibility at the complex interface.

Introdução

Capítulo Um

1 Introdução

O controle da atuação de enzimas em sistemas biológicos pode ocorrer através de vários mecanismos. Dois deles são a desativação por degradação proteolítica e o bloqueio por inibidores. Este último possui um importante papel em sistemas biológicos, especialmente naqueles que dependam de proteólise controlada para a manutenção de suas atividades (Batista *et al.*, 1996, Borges *et al.*, 1986, Travis & Salvenssen, 1983).

Inibidores protéicos de proteases estão envolvidos em processos fisiológicos como digestão, coagulação, manutenção da hemostasia, distúrbios hemorrágicos, etc., e seu estudo estrutural pode fornecer bons modelos para a compreensão do seu papel e dos seus mecanismos de atuação nestes processos (Laskowski & Qasim, 2000).

As leguminosas são ricas fontes de inibidores de tripsina e de outras serinoproteases (Onesti *et al.*, 1991) e representam uma importante fonte de proteína na alimentação humana. Inibidores enzimáticos protéicos

encontrados nas sementes destas leguminosas podem causar desordens nutricionais e efeitos tóxicos quando ingeridas e, desta forma, o seu conhecimento estrutural, nos estados nativo e complexado, pode levar à melhor compreensão de sua toxicidade.

Por outro lado, os inibidores enzimáticos protéicos também exercem papel regulador importante nas células, o que lhes atribui um possível uso farmacológico. Um destes usos é o tratamento de desordens metabólicas associadas a atividade proteolítica aumentada, como por exemplo pancreatite, enfisema, inflamação e certos tipos de câncer (Richardson, 1991). Testes em animais têm mostrado que certos inibidores de protease têm atividades anticarcinogênicas (Liener *et al.*, 1995, Troll *et al.*, 1987, Yavelow *et al.*, 1983) como supressão do crescimento induzido quimicamente de tumores nas glândulas anal e cólon em ratos (Billings *et al.*, 1990), supressão de tumor de mama em ratos e humanos (Troll *et al.*, 1980), tumores de pulmão em ratos (Witschi & Kennedy, 1989), supressão de transformações celulares induzidas quimicamente ou por radiação (Billings *et al.*, 1989, Billings *et al.*, 1987).

O inibidor de tripsina isolado de sementes de *Enterolobium contortisiliquum* (EcTI) apresenta inibição de enzimas de coagulação sangüínea e enzimas fibrinolíticas tais como o fator XIIa e plasmina, propriedades que não foram observadas em outros membros da mesma família de estrutura conhecida. Mas, por outro lado, não inibe trombina, calicreína tecidual e fator Xa (Batista, 1997).

1.1 Enzimas

Enzimas são catalisadores empregados no catabolismo de nutrientes e síntese das moléculas necessárias à sobrevivência em sistemas vivos. Essas reações químicas de síntese e metabolização devem ser altamente específicas e controladas, devendo ocorrer a velocidades adequadas à fisiologia celular. As enzimas são definidas como catalisadores por aumentarem em várias ordens de grandeza a velocidade das reações que catalisam sem alterar o equilíbrio termodinâmico e são caracterizadas também pela sua alta seletividade (Marzzoco & Torres, 1999, Moran & Scrimgeur, 1994). A velocidade de uma dada reação é aumentada quando sua energia de ativação (E_a) é diminuída e consequentemente a energia do estado de transição também é diminuída (Figura 1).

Figura 1: Perfil de reação enzimática mostrando a diminuição da energia de ativação (E_a) relativamente à mesma reação na ausência de enzima (pico mais alto).

A equação de Michaelis-Menten fornece a velocidade de formação do produto (ou velocidade inicial da reação total) supondo o estado estacionário considerando-se a formação do complexo de Michaelis (ES) reversível (Voet & Voet, 1995).

As velocidades das reações são diretamente proporcionais às concentrações dos reagentes envolvidos. Por exemplo, a velocidade de decomposição do complexo ES em E + P é $v_3 = k_3[ES]$, a velocidade de formação do complexo ES é $v_1 = k_1[E][S]$ e a velocidade de decomposição do complexo ES em E + S é $v_2 = k_2[ES]$. Na condição de estado estacionário, onde supõe-se a concentração do complexo [ES] constante, a velocidade de decomposição do complexo ES em E + P ou em E + S ($v_3 + v_2$), é igual a sua velocidade de formação (v_1). A equação de Michaelis-Menten é consequência desta condição e é escrita como

$$v_0 = \frac{V_{\max}[S]}{K_M + [S]} \quad (1)$$

onde v_0 é a velocidade inicial da reação no estado estacionário e K_M é a constante de Michaelis definida como $(k_2 + k_3)/k_1$, representando a concentração de substrato na qual a velocidade de reação é metade da velocidade máxima de reação (para uma dada concentração de enzima), $V_{\max}$ (obtida quando toda a enzima E_T se encontra sobre a forma ES, condição de saturação da enzima pelo substrato). Da definição de K_M pode-

se ainda obter a constante de dissociação K_S do complexo de Michaelis a partir da primeira etapa da reação enzimática

$$K_M = \frac{k_2}{k_1} + \frac{k_3}{k_1} = K_S + \frac{k_3}{k_1} \quad (2)$$

Se K_S diminui, a afinidade enzimática pelo substrato aumenta. Por esse prisma, K_M também mede afinidade da enzima pelo seu substrato dado que a razão k_3/k_1 é pequena se comparada com K_S , ou seja, $k_3 < k_2$.

Para a verificação da eficiência de uma catálise enzimática utiliza-se a constante catalítica, k_{cat} , que mede o número máximo de moléculas de substrato convertido em produto por sítio ativo da enzima por unidade de tempo. k_{cat} também é conhecido como número de *turnover* de uma enzima pois remete ao número de reações que cada sítio ativo cataliza por unidade de tempo. Ela pode ser escrita como

$$k_{cat} = \frac{V_{m\acute{a}x}}{[E_T]} \quad (3)$$

No estado estacionário $k_{cat} = k_3$ e portanto $V_{m\acute{a}x} = k_3[E_T]$.

Pode-se ainda definir a quantidade

$$\frac{k_{cat}}{K_M} = \frac{k_3}{K_M} = \frac{k_1 k_3}{k_2 + k_3} \quad (4)$$

k_{cat}/K_M é chamada de constante de especificidade por descrever a especificidade de uma enzima em competir pelos seus substratos e mede a eficiência catalítica de uma dada enzima (Marzzoco & Torres, 1999, Moran & Scrimgeur, 1994). A partir da equação de Michaelis-Menten e substituindo

$V_{\max}$ por $k_{\text{cat}}[E_T]$ se $[S] \ll K_m$ e portanto $[E] \approx [E_T]$, (pois pouca enzima se encontra na forma $[ES]$) então $v_0 \approx \frac{k_{\text{cat}}}{K_M} [E][S]$, e $\frac{k_{\text{cat}}}{K_M}$ é a constante de velocidade aparente de segunda ordem. Entretanto esta eficiência catalítica é limitada pela taxa máxima de colisões aleatórias entre as moléculas de enzima e substrato.

1.1.1 Serinoproteases

As proteases são divididas em quatro famílias funcionais de acordo com o grupo funcional mais proeminente no sítio ativo: serinoproteases, cisteinoproteases, aspartilproteases e metaloproteases. São caracterizadas pela capacidade de hidrolisar ligações peptídicas presentes em substratos protéicos ou peptídicos. Esta hidrólise envolve a adição de uma molécula de água ao substrato resultando na sua clivagem e a geração de dois produtos.

As serinoproteases estão presentes em vários tecidos de animais e plantas, assim como em microorganismos (Tabela 1) e são as enzimas proteolíticas mais intensamente estudadas e melhor caracterizadas.

Tabela 1: Algumas serinoproteases, fonte e função.

Enzima	Fonte	Função
Tripsina	Pâncreas	Digestão de proteínas
Quimotripsina	Pâncreas	Digestão de proteínas
Elastase	Pâncreas	Digestão de proteínas
Trombina	Soro de vertebrados	Coagulação sangüínea
Calicreína	Sangue e tecidos	Controle da coagulação
Subtilisina	<i>Bacillus subtilis</i>	Possivelmente Digestão
Diversos fatores do sistema de coagulação	Soro sangüíneo	Coagulação
Toxinas epidermolíticas	<i>Staphylococcus aureus</i>	Causadores da síndrome da pele escaldada
Proteosoma	Intracelular	Processamento e degradação de proteínas
NS3	Flavivírus	Processamento da poliproteína viral

Estudos do mecanismo catalítico e de inibição em quimotripsina, subtilisina e carboxipeptidase II (Robertus *et al.*, 1972) culminaram no modelo vigente (Wilmouth *et al.*, 2001) da polarização e transferência de prótons entre os resíduos que compõem a *tríade catalítica* conservada no sítio ativo, formada pelos resíduos Asp_c, His_c e Ser_c (c=catalítico). Esta transferência se dá do resíduo Ser195 para o resíduo His57. A carga positiva adquirida por este último em consequência da transferência é estabilizada pelo resíduo Asp102 (numeração da quimotripsinogênio) durante a catálise (Figura 2).

Figura 2: Passos do mecanismo de ação enzimática, acilação (em cima) e desacilação (embaixo).

A catálise acontece em dois passos: acilação e em seguida desacilação da enzima. Durante o primeiro passo um intermediário tetraédrico se forma e é estabilizado por ligações de hidrogênio com resíduos que pertencem ao bolsão do oxiânion (Figura 3).

Em enzimas da família quimotripsina esse bolsão é formado pela amida da Ser catalítica do sítio ativo e da Gly193 estruturalmente vizinha. A transferência do próton do resíduo His57 leva ao colapso deste intermediário formando uma acil-enzima (éster) e à liberação do fragmento C-terminal. Numa segunda reação (desacilação), o ataque de uma molécula de água à carbonila do éster leva ao segundo intermediário tetraédrico, que colapsa originando o fragmento N-terminal.

Figura 3: Representação em vista estéreo da ligação éster, entre o C da carbonila da Ile7 (posição P1 do substrato) e O γ da Ser195 (enzima), em complexo de elastase pancreática porcina (código PDB 1GVK) (Katona *et al.*, 2002). A figura representa a forma da acil-enzima formada durante o primeiro passo da catálise.

Foram utilizadas na última fase do nosso trabalho quatro diferentes serinoproteases para fins de modelagem, e desta forma, faz-se necessário um breve relato sobre cada uma delas.

1.1.1.1 Tripsina

A tripsina é uma serinoprotease caracterizada por clivar especificamente a ligação peptídica do lado C-terminal dos resíduos Lys ou Arg. Assim como outras serinoproteases, a tripsina tem sido extensivamente estudada tanto em termos da sua bioquímica quanto em termos da sua estrutura. O seu sítio de ligação foi caracterizado usando inibidores protéicos (Bode & Huber, 1992) que interagem de acordo com o “mecanismo padrão” (Laskowski & Kato, 1980) descrito na Seção 1.1.1 acima, e em maiores detalhes na seção 1.2 a seguir.

A tripsina é uma das três principais serinoproteases envolvidas no processo digestivo, juntamente com a pepsina e a quimotripsina. Ela estende o processo de digestão começado no estômago até o intestino delgado onde o ambiente ligeiramente mais alcalino, com pH em torno de 8, promove sua atividade enzimática máxima. Ela é produzida na forma inativa (tripsinogênio) pelo pâncreas e é homóloga à quimotripsina, também produzida no pâncreas (como quimotripsinogênio). Ambos os mecanismos de ação, são similares envolvendo os resíduos da tríade catalítica e o bolsão oxianion, que são seqüencialmente e estruturalmente conservados. A diferença essencial entre as duas está em suas especificidades com cada uma sendo ativa preferencialmente sobre ligações peptídicas que possuam grupos carboxílicos doados por aminoácidos específicos. No caso da tripsina estes aminoácidos são os positivamente carregados: arginina e lisina, e na quimotripsina são as cadeias aromáticas: tirosina, fenilalanina e triptofano. Segundo a nomenclatura de Schechter & Berger, (1967) este resíduo ocupa a posição denominada P1 no substrato, sendo a ligação peptídica hidrolisada entre os resíduos P1 e P1'. Resíduos antes de P1 na cadeia polipeptídica são denominados P2, P3, P4, etc. e os do lado C terminal são chamados de P1', P2', P3', etc. No caso da tripsina e quimotripsina, a especificidade no sítio de clivagem se dá quase que exclusivamente pela natureza do resíduo na posição P1 da proteína a ser clivada, como descrito acima. Este resíduo interage com um bolsão na enzima denominado S1. Na tripsina, a presença do resíduo Asp189 neste bolsão é responsável pela seletividade da enzima. Uma Arg na posição P1 do substrato interage diretamente com o Asp189 através de ligações de hidrogênio e no caso de

uma Lys em P1, devido a sua cadeia menor, esta interação é mediada por uma molécula de água (Huber & Bode, 1978). Em quimotripsina a substituição do Asp189 por Ser explica sua especificidade por aromáticos.

Como ilustração da atuação de um inibidor sobre tripsina apresentamos na Figura 4 uma estrutura de tripsina (verde) proveniente de pâncreas porcino complexada com inibidor proveniente de soja (SBTI) depositada no Protein Data Bank - PDB (Berman *et al.*, 2000) sob o código 1AVW (Song & Suh, 1998).

Figura 4: Tripsina porcina (verde) complexada com inibidor SBTI (azul) (PDB 1AVW). O *loop* do sítio reativo do inibidor penetra na fenda entre os dois domínios da enzima.

A Figura 5 e a Figura 6 mostram a especificidade e complementaridade espacial no reconhecimento do resíduo P1 pelo bolsão S1 e o sítio ativo da protease. A cadeia lateral da Arg (mostrada em “ball and stick” na Figura 5) na posição P1 penetra uma cavidade na enzima, o bolsão S1. A Asp189 no fundo deste bolsão é o principal fator responsável pela especificidade da tripsina (Figura 6).

Figura 5: Tripsina representada como superfície sólida complexada com SBTI (em amarelo), representado pelos C α e a Arg na posição P1 do inibidor representada em “ball and stick”.

Figura 6: Sítio ativo da tripsina em complexo com substrato (PDB 2PTC) (Marquart *et al.*, 1983). Os resíduos da tríade catalítica (Ser195, His97, Asp102) são representados em “ball and stick” junto com a lisina P1 do substrato. Este último é estabilizado pelo Asp189 no bolsão S1 (lado direito inferior). A ligação peptídica atacada pela Ser195 está indicada.

1.1.1.2 Trombina

Para a sobrevivência de um organismo superior, é essencial que cortes ou feridas parem de sangrar, ou seja, há a necessidade de um mecanismo adequado para a manutenção das condições normais do fluxo sanguíneo (hemostasia). Esse mecanismo deve manter o fluxo adequado de sangue

nos vasos sangüíneos em condições normais e estar preparado para estancar imediatamente vasos danificados^{*,†}.

A trombina, que é uma serinoprotease homóloga à tripsina e à quimotripsina, está no centro deste processo de coagulação sangüínea (Figura 7). Na via extrínseca de coagulação, após um corte tecidual vasos sangüíneos se rompem e a cascata de coagulação tem início com a sinalização por parte de moléculas do fator tecidual presente na superfície celular, que normalmente não está em contato com sangue. O fator tecidual ativa algumas moléculas de fator VII que por sua vez ativam um número maior de fator X que, num processo de amplificação, ativa uma quantidade maior de trombina. A trombina, então, cliva o fibrinogênio (solúvel) convertendo-o em monômeros de fibrina que polimerizam para formar um coágulo de fibrina insolúvel (van de Locht *et al.*, 1997). A trombina também pode ser ativada na via intrínseca envolvendo os fatores XI, IX e X, além de ativar outros fatores de coagulação como os fatores V, VIII, XI e XIII, através de proteólise. Após a ligação com trombomodulina, a trombina sofre modificações para ativar a proteína C, responsável pela inativação dos fatores Va e VIIa (Esmon, 1989).

^{*} <http://www.geocities.com/gepsaude/textos/hemostase.html> – 18/07/2003.

[†] <http://www.thrombin.com> – 18/07/03.

Figura 7: Cascata de coagulação sanguínea (via extrínseca e via intrínseca). Pontos de mediação positiva e negativa são indicados pelos sinais “+” e “-”, respectivamente.

Figura 8: Molécula de trombina, com suas cadeias leve (vermelho) e pesada (verde) complexada com inibidor (azul) proveniente de pâncreas bovino, (PDB 1BTH) (van de Locht *et al.*, 1997).

A molécula de trombina possui duas cadeias ligadas por pontes de dissulfeto, a cadeia L(α) com 26 resíduos e a cadeia H(β) com 256 resíduos. Suas dimensões aproximadas são 45x45x50 Å e é formada por dois barris β de seis fitas como as demais serinoproteases da família quimotripsina. Os resíduos da tríade catalítica se encontram na junção entre os dois barris. O *loop* à direita na Figura 8 forma uma “tampa” hidrofóbica sobre o sítio ativo restringindo aproximação da maioria dos possíveis substratos protéicos

(Stubbs & Bode, 1995). Trombina é muito mais específica em relação a seqüência de aminoácidos do substrato, comparativamente à tripsina e quimotripsina. A seqüência canônica reconhecida por trombina é $\phi\phi$ PRGS (ϕ =hidrofóbico), sendo que a clivagem ocorre entre a Arg e a Gly (Smith & Johnson, 1988, Doi *et al.*, 1986). Esta especificidade é fundamental para um controle adequado da atividade hidrolítica da trombina no soro sangüíneo. Além do sítio ativo, a trombina também interage com moléculas efetoras através de dois exossítios: exossítio 1, responsável pela interação com fibrinogênio, e o exossítio 2, responsável pela interação com heparina.

1.1.1.3 Fator Xa

O fator Xa (Figura 9) é forma ativa do fator X e tem papel importante no processo de coagulação e ativação plaquetária (Figura 7). Ele é um componente essencial no processo de formação da trombina (Walenga *et al.*, 2003). A compreensão de sua inibição e ativação representa uma possibilidade estratégica no desenvolvimento de novos fármacos antitrombose (Brandstetter *et al.*, 1996). O fator X humano é secretado na corrente sangüínea como uma proteína dimérica (M=59 kDa) glicosilada onde cada monômero consiste de duas cadeias unidas por uma ponte dissulfeto. A cadeia leve possui 52 resíduos e três pontes de dissulfeto intracadeia. A cadeia pesada possui 228 resíduos e quatro pontes de dissulfeto intracadeia. Os resíduos Cys44 da cadeia leve e Cys122 da cadeia pesada formam uma ponte de dissulfeto intercadeia (Maignan *et al.*, 2000). A ativação do fator X ocorre na superfície da membrana fosfolipídica

na presença de Ca^{++} através de sua clivagem pelo complexo fator VIIa/fator tecidual (via extrínseca). Na via intrínseca o fator IXa realiza esta mesma clivagem na presença do fator VIIIa. A estrutura secundária do fator Xa é muito similar à da trombina, mas seu sítio ativo não possui a mesma “tampa” que se observa na trombina (Stubbs & Bode, 1995). Mesmo assim, o fator Xa é altamente específico em relação ao substrato, reconhecendo preferencialmente a sequência IEGR onde a hidrólise acontece após a Arg P1 (Nagai & Thogersen, 1984).

Figura 9: Fator de coagulação Xa, cadeia leve (vermelho) e pesada (verde) complexado com inibidor não protéico (azul), (PDB 1EZQ) (Maignan *et al.*, 2000). Os resíduos Cys44 da cadeia leve e Cys122 da cadeia pesada formam uma ponte de dissulfeto intercadeia.

1.1.1.4 Quimotripsina

A forma inativa da quimotripsina é o quimotripsinogênio com 245 resíduos, que no processo de ativação sofre clivagem entre os resíduos 14-15 e 147-148. Como as demais serinoproteases discutidas até agora, possui dois domínios com ~120 resíduos, cada domínio contendo um barril β de seis fitas. Seu sítio ativo situa-se entre os dois domínios com a His57 e Asp102 estando no primeiro domínio e a Ser 195 estando no segundo domínio

(Scheidig *et al.*, 1997). A origem de sua seletividade por resíduos aromáticos na posição P1 já foi anteriormente aqui descrita (seção 1.1.1.1).

1.2 Inibidores de Serinoproteases

Grande parte da compreensão que se tem do mecanismo catalítico das proteases provém dos estudos estruturais, utilizando-se cristalografia de raios X ou RMN, não só das enzimas, mas também de seus complexos com substratos ou com seus inibidores (Hubbard *et al.*, 1998).

Os inibidores enzimáticos protéicos exercem papel regulador importante nas células (Marzzoco & Torres, 1999, Lehninger, 1993, Brandén & Tooze, 1991). Pode-se agrupá-los em duas grandes categorias, segundo a estabilidade de sua ligação com a enzima:

- Inibidores irreversíveis;
- Inibidores reversíveis.

Os inibidores irreversíveis interagem com as enzimas levando a uma inativação praticamente definitiva devido a ligações covalentes formadas entre a enzima e o inibidor. Após essa inativação a enzima original não pode ser regenerada ou recuperada. Geralmente estes inibidores são muito tóxicos e de baixa especificidade.

Os inibidores reversíveis são classicamente divididos entre: competitivos e não competitivos. Essa classificação se refere ao

estabelecimento de competição entre o inibidor e o substrato pelo centro ativo da enzima.

Os inibidores não competitivos, I_{NC} , não possuem semelhança conformacional com o substrato, ligando-se muitas vezes, inespecificamente a resíduos que não pertencem ao sítio ativo da enzima induzindo-lhe uma alteração estrutural que acaba impedindo a catálise. Esta ligação está sujeita a um equilíbrio do tipo:

onde K_{INC}

$$K_{INC} = \frac{[E][I_{NC}]}{[EI_{NC}]} = \frac{k_{-1}}{k_1} \quad (6)$$

Por outro lado, o inibidor não competitivo pode também muitas vezes formar um complexo ternário ESI_{NC} incapaz de gerar produto (Figura 10).

Figura 10 : Esquema ilustrando a inibição não competitiva.

Por possuírem semelhança conformacional com o substrato, os inibidores competitivos I_c , são capazes de ligarem-se ao sítio ativo da

enzima produzindo um complexo enzima-inibidor, mas sem levar a cabo nenhuma catálise

onde K_{IC} é chamada de constante de inibição e mede a afinidade da enzima ao inibidor

$$K_{IC} = \frac{[E][I_C]}{[EI_C]} = \frac{k_4}{k_1} \quad (8)$$

A Figura 11 exemplifica a situação esquematicamente.

Figura 11 : Esquema ilustrando a inibição competitiva.

As serinoproteases e seus inibidores protéicos naturais estão entre os modelos de reconhecimento proteína-proteína mais intensamente estudados (Apostoluk & Otlewski, 1998, Jones & Thornton, 1996, Janin & Chothia, 1990). Os inibidores protéicos de proteases não formam uma classe única podendo ser agrupados em famílias de acordo com sua massa molecular, composição de aminoácidos, localização do sítio reativo, especificidade inibitória, número de pontes de dissulfeto, etc. Apesar das divergências existentes quanto às possíveis classificações (Laskowski & Qasim, 2000)

pode-se adotar uma classificação mais conservadora ou clássica (Bode & Huber, 1992) (Tabela 2):

Tabela 2 : Famílias de inibidores protéicos que inibem serinoproteases e obedecem ao mecanismo padrão (Bode & Huber, 1992).

Família	Kunitz de animais (BpTI – B ovine P ancreatic T rypsin I nhibitor)
	Kunitz de vegetais (SbTI – S oy b ean T rypsin I nhibitor)
	Kazal
	SSI (S trep t omices S ubtilisin I nhibitor)
	Batata tipo I
	Batata tipo II
	Bowman-Birk
	Abóbora
	Hirudina
	Serpina

A interação serinoprotease - inibidor protéico pode ser representada (Otlewski & Zbyryt, 1994) pelo mecanismo padrão

onde I representa o inibidor intacto, I* o inibidor modificado - clivado na ligação peptídica P1-P1', C é o complexo inibidor-protease estável do tipo Michaelis que é rapidamente formado mas dissocia-se muito lentamente. k_{on} e k_{on}^* são as constantes de formação do complexo protease-inibidor intacto e protease-inibidor clivado, respectivamente. k_{off} e k_{off}^* são as constantes de dissociação do complexo C. Em termos das concentrações dos componentes do sistema a constante de associação, K_a , considerando apenas o equilíbrio entre E, I e C é dada por

$$K_a = \frac{[C]}{[E][I]} \quad (10)$$

se a presença do inibidor clivado, I^* , for levada em consideração obtém-se uma constante mais global, K_{assoc} :

$$K_{assoc} = \frac{[C]}{[E]([I] + [I^*])} \quad (11)$$

A determinação estrutural de alguns inibidores e complexos inibidor-protease levou à elucidação de um mecanismo de ação comum (mecanismo padrão) aos inibidores e a uma compreensão geral das interações específicas proteína-proteína envolvendo os inibidores e suas proteases cognatas em nível molecular. Com base nestas estruturas, um *loop* presente na superfície das moléculas de inibidores, chamado sítio reativo, loop reativo ou *loop* canônico, pôde ser identificado como um motivo estrutural de reconhecimento que interage forte e especificamente com o sítio ativo da enzima resultando num complexo enzima-inibidor.

Outros aspectos do mecanismo padrão incluem (Otlewski & Zbyryt, 1994, Bode & Huber, 1992): 1) o complexo C é mais estável que o complexo de Michaelis ES para o caso de substratos, 2) a constante de velocidade de hidrólise do sítio reativo é pequena, 3) a hidrólise é reversível, ou seja, a forma clivada continua ativa (inibitória) formando o mesmo complexo com a enzima, permitindo a reformação da ligação peptídica, 4) o valor de

$k_{hid} = \frac{[I]}{[I^*]}$ quando em equilíbrio é geralmente próximo de 1 em pH neutro.

Bode & Huber (1992) definiram o *loop* canônico como sendo constituído por seis resíduos P_3 - P_2 - P_1 - P_1' - P_2' - P_3' (conforme nomenclatura de Schechter & Berger) - onde a ligação peptídica P_1 - P_1' representa o sítio de clivagem proteolítica. Sugerem também que este *loop* apresenta motivo conformacional de cadeia principal que imita a conformação de uma ligação produtiva de um substrato (ver também Laskowski & Kato, 1980) e que suas características conformacionais são mantidas inclusive quando complexados. Em complexos entre enzima e inibidor, tipicamente a distância entre a carbonila P_1 e o $O\gamma$ da Ser195 está entre 2,6-2,8 Å e a ligação P_1 - P_1' está intacta.

Estas propriedades conformacionais da cadeia principal do *loop* reativo são intrínsecas aos inibidores protéicos, embora se admita que esse motivo não seja exclusivo dos *loops* de inibidores (Jackson & Russel, 2000, Apostoluk & Otlewski, 1998) estando presentes também em regiões de *loop* em proteínas que não apresentam estas características. A similaridade entre os *loops* de ligação das diferentes famílias de inibidores sugere que o *loop* de ligação tenha aparecido independentemente em diferentes famílias de proteínas, e pode ser considerado um exemplo de evolução convergente.

O mecanismo de ação ou forma de reconhecimento remete à ligação de um substrato ideal (Hubbard *et al.*, 1991) constituindo-se de uma seção curta de folha- β antiparalela entre o segmento 214-216 (numeração da quimotripsina) da enzima e a seção N-terminal do *loop* de ligação do inibidor

(resíduos P3-P1), uma ligação de hidrogênio entre o oxigênio da carbonila do resíduo P1 e os grupos amida da cadeia principal da Ser195 e Gly193 (bolsão oxianion) e uma aproximação do O γ da Ser195 em direção do carbono da carbonila de P1.

A diferença essencial entre inibidores protéicos e substratos refere-se à rápida dissociação do substrato do sítio ativo da enzima após a hidrólise, enquanto que os inibidores protéicos, via de regra, permanecem complexados com a enzima. Essa característica representa uma questão importante, uma vez que os inibidores protéicos apresentam seqüências peptídicas que se ligam a proteases específicas de um modo muito semelhante à de um substrato e desta forma, espera-se que sofra hidrólise e rápida dissociação. Entretanto, os inibidores protéicos ligam-se mais fortemente comparativamente aos melhores substratos destas enzimas, com constantes de associação maiores que 10^{14}M^{-1} e são hidrolisados lentamente com constantes de dissociação da ordem de $10^6\text{-}10^{10}\text{ M}^{-1}$ (Radisky & Koshland, 2002).

Algumas hipóteses foram levantadas na literatura para a explicação da baixa reatividade entre inibidor e enzima: 1) a grande rigidez do complexo frustra o ataque nucleofílico produtivo (Peräkylä, & Kollman, 2000, Takeuchi, *et al.*, 1991, Fujinaga *et al.*, 1982); 2) a pobre orientação dos grupos reativos resulta num complexo não produtivo (Coombs *et al.*, 1999, Martin *et al.*, 1996); 3) o posicionamento do grupo de partida $\text{H}_2\text{N-R}_2$ no complexo acil-enzima favorece a reação de volta ao complexo de Michaelis (Peräkylä, &

Kollman, 2000, Longstaff *et al.*, 1990). Recentemente, Radisky & Koshland, (2002) 4) a partir de estudos com subtilisina e o inibidor Cl2 chegaram à conclusão que o favorecimento da reação reversa é a contribuinte principal à inibição.

1.2.1 Família SBTI

O primeiro inibidor de protease proveniente de planta a ser bem caracterizado foi o inibidor de tripsina de soja (SBTI – Soybean Trypsin Inhibitor) da família Kunitz (de Meester *et al.*, 1998). Tendo sido utilizado como o principal substrato nos trabalhos bioquímicos e cinéticos que levaram à definição do mecanismo de ação padrão de inibição dos inibidores de protease. O papel fisiológico dos inibidores desta família não está completamente compreendido, mas tem sido proposto que eles tenham uma função defensiva contra as proteases presentes no aparelho digestivo de microorganismos, vertebrados e invertebrados, e que estejam envolvidos no controle das proteases endógenas durante o período de dormência da semente (Laskowski & Kato, 1980).

Os inibidores da família Kunitz são encontrados em todas as sementes de leguminosas (Sampaio *et al.*, 1991), possuem massa molecular por volta de 20kDa e geralmente têm duas pontes de dissulfeto estabilizando a sua estrutura. Porém, recentemente foram descritos inibidores possuindo apenas uma ponte de dissulfeto (do Socorro *at al.*, 2002) ou nenhuma ponte (Oliva, 2003). Os inibidores purificados de sementes das subfamílias *Caesalpinoideae* ou *Papilonoideae* apresentam apenas uma cadeia

polipeptídica, porém aqueles purificados das sementes da subfamília *Mimosoideae* são proteínas compostas por duas cadeias polipeptídicas (Batista *et al.*, 1996) que são oriundas do processamento pós-traducional de uma única cadeia homóloga aos demais inibidores Kunitz.

O inibidor de tripsina de soja do tipo Kunitz foi inicialmente isolado e caracterizado por Kunitz (Kunitz, 1947a e Kunitz, 1947b) consistindo de 181 resíduos, podendo ser hidrolisados pela tripsina com o sítio de clivagem situado entre a Arg63 e a Ile64 (Osawa & Laskowski, 1966) tornando-se uma referência para esta família de proteínas.

1.2.1.1 *Enterolobium contortisiliquum*

O *Enterolobium contortisiliquum* é uma leguminosa que pertence à família *Leguminosae*, subfamília *Mimosoideae* (Lorenzi, 1992). A etimologia de sua denominação fornece: do grego, *entero* de intestino e *lobium* de lóbulo, ou volta, em alusão à forma de seus frutos; do latim, *contortis* de retorcido e *liquum* de líquido. Seus nomes mais comuns são: orelha-de-negro, timburi, pau-de-sabão, tamboril-do-campo, timbó, timbaú. Florescendo de Setembro a Novembro, tem ocorrência no Brasil (do Pará até o Rio Grande do Sul), na Bolívia, Paraguai e Argentina. A sua árvore possui cerca de 15 a 30 m de altura e uma copa ampla. Sua casca é pardo-acinzentada e suas flores são brancas, o fruto é duro tendo de 6 a 8 cm de diâmetro e é fortemente encurvado, e suas sementes possuem cerca de 10 mm de comprimento. O fruto e as folhas são nocivos ao gado, devido a saponina neles existente.

Sua madeira avermelhada, macia e resistente a umidade é usada na fabricação de canoas de tronco inteiro, além de embalagens, brinquedos e móveis, e devido ao seu rápido crescimento, possui um grande potencial para reflorestamentos e recuperação de áreas agredidas. A Figura 12 ilustra alguns aspectos do *Enterolobium contortisiliquum*.

Figura 12: Aspectos do *Enterolobium contortisiliquum*. No sentido horário: porte, folhas, tronco e casca, fruto e sementes^{‡,§}.

Batista *et al.* (1996) descreveram o isolamento de um inibidor de tripsina proveniente de *Enterolobium contortisiliquum* (EcTI). O inibidor foi

[‡] http://www.garden.com.br/gcenter2/materia_garden_novo.asp?noticia1=560 - 14/07/2003

[§] <http://www.esalq.usp.br/trilhas/uteis/ut24.htm> - 14/07/2003

caracterizado com um peso molecular de 19,6 kDa, possuindo duas cadeias, identificadas em gel de SDS/PAGE. São no total 174 resíduos, sendo que 134 na cadeia pesada (α) de 15,1 kDa e 40 resíduos na cadeia leve (β) com 4,5 kDa. Sua atividade inibitória se mantém após incubação até 60 °C por 10 min, mas não a 100 °C. A sua pré-incubação em pH na faixa 2,0-12,0 por 10 min também não afeta sua atividade inibitória mostrando que o EcTI é uma molécula bastante estável com relação a sua atividade em ambientes bastante distintos.

Diversas medidas experimentais da constante de inibição K_i do inibidor EcTI foram realizadas frente diversas serinoproteases (Batista, 1997). Entre elas damos destaque para os resultados de tripsina bovina, quimotripsina bovina, calicreína plasmática humana, plasmina humana, fator Xa humano, fator XIIa humano e ativador de plasminogênio humano na Tabela 3 que também compara a atividade inibitória da EcTI com outros inibidores Kunitz. Neste trabalho ainda foi constatado que nenhum destes inibidores inativa a trombina, sendo sugerido não serem conhecidos quaisquer inibidores vegetais do tipo SBTI da família Kunitz que sejam ativos sobre esta enzima. A Tabela 3 também mostra a alta afinidade do EcTI pela tripsina, quimotripsina, calicreína plasmática, plasmina e fator XIIa, e uma ausência de atividade contra fator Xa e tPA (ativador tecidual de plasminogênio). Um aspecto curioso sobre o EcTI é a sua capacidade de inibir Fator XIIa mas não tPA. Como pode ser vista na Tabela 3, para todos os demais inibidores há uma correlação entre a capacidade de um inibidor inibir estas duas enzimas no sentido de que ou inibir as duas ou nenhuma

delas. A capacidade de distinguir entre fator Xlla e tPA parece ser uma propriedade única do EcTI.

1.3 Objetivos

Esse trabalho teve como objetivo a determinação da estrutura tridimensional do inibidor de tripsina de sementes de *Enterolobium contortisiliquum*, EcTI, e a modelagem molecular de complexos de EcTI com serino proteases de estrutura conhecida (EcTI-tripsina, EcTI-quimotripsina , EcTI-trombina, EcTI-fator Xa), e suas análises. Faz parte também dos objetivos o estudo das características do enovelamento β -*trefoil* de EcTI, enovelamento este comum aos membros da família Kunitz até hoje conhecidos além de outras famílias de proteínas não relacionadas. O objetivo final era a elaboração de uma tentativa de explicar em termos estruturais as propriedades inibitórias do EcTI relatadas na literatura.

Tabela 3: Medidas experimentais da constante de inibição K, medida em nM, para diversas serinoproteases e inibidores. SBTI – Soybean trypsin inhibitor, BuTI – *Bauhinia unguolata* trypsin inhibitor, BvTI – *Bauhinia variegata* trypsin inhibitor, EcTI – *E. contortisiliquum* trypsin inhibitor, LiTI – *Leucaena leucocephala*, SpTI – *Swartzia pickelli* trypsin inhibitor. Valores entre parênteses correspondem ao desvio padrão dos valores determinados, N.I. – não inibe.

Inibidor	Serinoproteases						
	Tripsina Bovina	Quimotripsina Bovina	Calicreína Plasmática Humana	Plasmina Humana	Fator Xa Humano	Fator XIIa Humano	Ativador de Plasminogênio Humano
SBTI	1,29(±0,73)	70,29(±31,67)	5,70(±3,54)	192,10(±78,37)	1367,59(±218,20)	N.I.	N.I.
BuTI	20,68(±4,35)	27,28(±10,76)	6,77(±1,52)	76,88(±16,79)	18,54(±2,10)	10,22(±0,86)	502(±568)
BvTI	1,52(±0,93)	12,23(±3,08)	22,65(±3,35)	2,96(±0,54)	93,07(±9,38)	21,25(±1,32)	31,93(±13,71)
EcTI	0,88(±0,25)	1,11(±0,67)	6,15(±0,98)	9,36(±0,92)	N.I.	81,81(±14,89)	N.I.
LiTI	3,61(±0,79)	14,59(±3,85)	14,08(±2,43)	2,12(±1,01)	N.I.	N.I.	N.I.
SpTI	47,40(±15,09)	123,40(±29,56)	20,25(±2,14)	2,52(±0,45)	N.I.	N.I.	N.I.

Materiais e Métodos

Capítulo Dois

2 Materiais e Métodos

Este capítulo está estruturado com uma breve introdução teórica antes da descrição dos passos do processo de determinação e refinamento da estrutura cristalográfica do EcTI, quando julgado relevante.

2.1 Cristalização e Coleta de Dados Cristalográficos

A extração do EcTI a partir das sementes de *E.contortisiliquum* e a determinação de sua estrutura primária estão descritos na referência (Batista *et al.*, 1996). O EcTI foi dializado (Batista *et al.*, 2001) contra o tampão 0,02M TrisHCl, pH 8,0 e concentrado usando-se membranas Centricon YM-3 (Millipore) até 7,5 mg/ml. A concentração de proteína foi determinada medindo-se a absorbância a 280nm de uma amostra usando-se o coeficiente de extinção de $1,0 \text{ cm}^{-1}$ para 1,0 mg/ml de solução de EcTI. A solução obtida foi usada para se efetuar ensaios de cristalização com os *kits* da Hampton *Research Crystal Screen* (Fatorial I) empregando-se o método da difusão de vapor com gota pendurada. Como os cristais crescidos por

este método apresentavam hábito de aglomerado na condição número 50 (15% PEG8000, 0,5M LiSO₄), tentou-se a técnica de acupuntura em gel (Moreno & Soriano-García, 1999) (Figura 13) a fim de otimizar as condições iniciais de cristalização, visando um mono-cristal adequado à difração de raios X.

Uma das características desta técnica é que os cristais são obtidos dentro de capilares apropriados à coleta de dados de difração de raios X, e permite a formação na parte inferior do capilar de um gradiente de difusão que transporta o agente precipitante em direção ao seu topo. Moreno *et al.*, (1999) mostraram que: 1) a densidade de nucleação diminui ao longo do capilar, 2) existe uma relação direta entre a altura do capilar e o tamanho dos cristais obtidos, 3) a taxa de crescimento dos cristais varia de acordo com sua posição no capilar, sendo mais lenta para cristais crescendo em localizações mais próximas do topo do capilar. A supersaturação σ neste sistema é definida como a razão entre a concentração atual $[P]$ da proteína e sua concentração no equilíbrio $[P_e]$, lembrando que $[P_e]$ é a concentração de equilíbrio da proteína para uma dada concentração de precipitante:

$$\sigma = \frac{[P]}{[P_e]} \quad (12)$$

e a taxa na qual a supersaturação aumenta é

$$\Theta = \frac{\delta\sigma}{\delta t} \quad (13)$$

A variável Θ mede a variação no fluxo de unidades de crescimento para a superfície do cristal. Nestas condições para um dado tempo t , σ

diminui como a proximidade do topo do capilar e para uma dada altura h , σ aumenta em função do tempo.

Figura 13: Ilustração do arranjo experimental da técnica de acupuntura em gel.

Para a cristalização do EcTI foi utilizado um cassete de vidro consistindo de duas placas de vidro retangulares de 7 cm X 10 cm mantidas separadas por um espaçador em forma de "U" cobrindo os dois lados maiores e um dos lados menores da placa (Figura 14). Após prender as placas com garras, o espaço entre elas foi de aproximadamente 4mm. O cassete foi então preenchido, até uma altura de 4cm (11,2 cm³) com Agarose 1% previamente aquecida. Uma vez a Agarose tendo atingido o estado gel, um capilar cilíndrico de 0,5 mm foi preenchido, por capilaridade, com solução de proteína (solução mãe a 10 mg/ml) e uma de suas extremidades selada com cera. O capilar foi então "espetado" no gel com uma penetração aproximada de 7 mm. Logo em seguida, o cassete foi preenchido com solução precipitante (15% PEG8000 - 0,5M LiSO₄) até uma

altura aproximada de 8,0 cm ($11,2 \text{ cm}^3$) e finalmente, a parte aberta do cassete foi fechada com fita adesiva.

Figura 14: Ilustração do arranjo experimental montado para o uso da técnica de acupuntura em gel. As bordas pretas do desenho representam tiras de borracha colocadas entre as placas de vidro.

Nesta técnica, o controle do processo de crescimento é feito ao microscópio ótico com o cassete colocado na posição horizontal e após a verificação, o cassete é novamente colocado na vertical (García-Ruiz & Moreno, 1994). O nosso experimento foi mantido à temperatura constante de 292K e após três semanas, monocristais apropriados à coleta de dados foram observados no capilar.

Dado que o mono-cristal foi crescido na ausência de agente crioprotetor, foi necessário transferi-lo do capilar para uma solução contendo o líquido mãe acrescido de 25% (v/v) de glicerol e este teve sua temperatura abaixada para 100K utilizando-se fluxo de nitrogênio gasoso. Os dados foram coletados num detetor de placas de imagem MAR345 na

linha de Cristalografia de Proteínas (PCr) no Laboratório Nacional de Luz Síncrotron (LNLS) (Polikarpov *et al.*, 1998) pelo responsável técnico José Ribeiro Brandão e pela Dra. Isabel Batista, assim como a redução dos dados e seu escalonamento.

O detetor foi ajustado para leitura até 300 mm de diâmetro e a distância cristal detetor ajustada para 90 mm. Os dados observáveis se estenderam até 1,7 Å de resolução, no entanto, o surgimento de anéis relacionados a cristais de gelo nas imagens dos padrões de difração forçou-nos a aceitar o limite de 1,96 Å. A redução dos dados e seu escalonamento foram realizados utilizando os programas DENZO (auto-indexação, refinamento e integração) e SCALEPACK (escalonamento dos dados) (Otwinowski & Minor, 1997).

O cristal pertence ao grupo espacial $P2_1$ encontrado através das extinções sistemáticas observadas ao longo do eixo b^* e com parâmetros de rede $a=37,12$ Å, $b=38,42$ Å e $c=54,08$ Å e $\beta=98,08^\circ$ possuindo uma molécula por unidade assimétrica. O número de moléculas por unidade assimétrica foi calculado através do conteúdo de solvente no cristal que foi, por sua vez, estimado a partir do número de Matthews, V_M (Matthews, 1968). O número de moléculas por unidade assimétrica encontrado foi um, uma vez que, assumindo duas moléculas por unidade assimétrica, obtém-se $V_M = 0,98$ Å³Da⁻¹ que corresponde a um conteúdo negativo de solvente no cristal, obviamente, um dado não realístico.

2.2 Resolução da Estrutura

A técnica de substituição molecular foi utilizada para a resolução da estrutura e já foi extensivamente descrita na literatura (Navaza, 1997, Rossmann, 1972). Basicamente, seu conceito envolve calcular as fases a partir de um modelo (molde) que seja estruturalmente muito parecido (homólogo) à estrutura cristalina procurada. Esse conjunto de fases iniciais será utilizado para se dar início à construção do modelo cristalográfico e subsequente refinamento.

A estrutura do EcTI foi determinada por substituição molecular utilizando-se o programa AMoRE (**A**utomated **M**olecular **R**eplacement) (Navaza & Saludjian, 1997) do pacote CCP4 (CCP4, 1994). A função de rotação foi calculada entre 10 e 3 Å e as potenciais soluções foram encontradas utilizando-se as coordenadas atômicas do inibidor de tripsina de *Erythrina caffra* resolvido a 2,5 Å de resolução (PDB 1TIE) (Onesti *et al.*, 1991) como molde, apresentando 39,29% de identidade com EcTI (Tabela 4). O 1TIE teve seus resíduos não conservados reduzidos a alaninas para a entrada em AMoRE.

Tabela 4: Alinhamento entre os inibidores de *E.caffra* (1TIE), de soja (1BA7) e EcTI mostrando 39,29% entre 1TIE e EcTI, 36,99% entre EcTI e 1BA7 e 42,35% entre 1TIE e 1BA7 de porcentagem de identidade sequencial.

```

Erythrina 1TIE      1  VLLDGNGEVVQNGGTYLLPQVWAQGGGVQLAKTGEETCPLTVVQSPNELSDGKPIRIE
EcTI (propos)      1  KELLDSDGDILRNNGTTYIILPALRGKGGGLELAKTGDETCPLNVVQARGETKRGRPAIIW
SBTI               1BA7 1  DFVLDNEGNPLENGGTYIILSDI  TAFGGIRAAPTGNERCPLTVVQSRNELDKGIGTIIS

Erythrina 1TIE      61  SRLRSAPFIPDDDKVRIGFA YA  PKCAPSPW WTVVEDEQEGLSVKLSEDESTQFDYPF
EcTI (propos)      61  TPPRIAILTPAFYLNIEFQ TKDLPACLREYS RLPREEEQHS EVKLAPKEEAA  AF
SBTI               1BA7 61  SPYRIRFIAEGHPLSLKFDSFAVIMLCVGIPTEWSVVEDLPEGPAVKIGENKDAMDGW F

Erythrina 1TIE      121 KFEQVSDQ LHSYKLLYCEG  KHEKCASIGINRD QKGYRRLVVTEDYPLTVVLKKDES
EcTI (propos)      121 GXEKLKPY RDDYKIVYCEGSDDDSCDKLGISID DENNRRLVVKDGDPLAVRFVKAHR
SBTI               1BA7 121 RLERVSDDEFNNYKLVFCPQQAEDDKCGDIGISIDHDDGTRRLVVSKNKPLVVQFQKLDK

Erythrina 1TIE      181 S
EcTI (propos)      181 RG
SBTI               1BA7 181 ESL

```

2.3 Refinamento da Estrutura Cristalina

O que se quer com o refinamento da estrutura cristalina é obter um modelo da molécula sob investigação que concorde o mais possível com os dados cristalográficos observados. Em cristalografia de moléculas pequenas o número de reflexões observadas geralmente excede em várias vezes (~20) o número de parâmetros a serem refinados. Este não é o caso para proteínas, de forma que vínculos e restrições geométricas devem ser impostos ao modelo para se melhorar a razão dados observados/número de parâmetros. O refinamento de estruturas protéicas visa a obtenção de um modelo que esteja em concordância tanto com as intensidades observadas quanto com as restrições estereoquímicas impostas. Portanto, há a necessidade de critérios de acompanhamento tanto dos termos cristalográficos quanto dos termos estereoquímicos (Kleywegt, 2000).

Historicamente, a estatística mais usada para avaliar quão bem um modelo cristalográfico se adequa aos dados experimentais é o R_{factor} ou $R_{\text{cristalográfico}}$,

$$R_{\text{factor}} = \frac{\sum w |F_o - k|F_c||}{\sum |F_o|} \quad (14)$$

Entretanto, os valores de R_{factor} só têm significado se o número de observações experimentais e o número de restrições impostas ao modelo excedem o seu número de parâmetros (Kleywegt, 2000). Em 1992, Brünger introduziu o R_{free} (Brünger, 1992) cuja definição é idêntica à do R_{factor} exceto que o R_{free} é calculado usando um pequeno subconjunto de reflexões retiradas aleatoriamente (geralmente em torno de 5%) que não são usadas no refinamento do modelo. É sugerido (Kleywegt & Jones, 1997) que a diferença $R_{\text{free}} - R_{\text{factor}}$ deva ser pequena ou alternativamente, a razão $R_{\text{free}} / R_{\text{factor}}$ deva estar próxima à unidade. Bons conjuntos de dados podem fornecer diferenças da ordem de 0,02 e dados com qualidade inferior ou baixa resolução, a diferença pode chegar a 0,08. Um outro fator que pode também alterar esta diferença é redundância na unidade assimétrica, ou simetria não cristalográfica.

A verificação da qualidade do modelo por resíduo, individualmente, no espaço real também pode ser feita através do índice R de espaço real sugerido por (Jones *et al.*, 1991). Este índice mede a similaridade entre um mapa calculado (ρ_c) a partir do modelo e outro (ρ_o) a partir dos dados experimentais e os relaciona da seguinte forma

$$R = \frac{\sum |\rho_o - \rho_c|}{\sum |\rho_o + \rho_c|} \quad (15)$$

com a soma sendo realizada sobre todos os pontos do *grid* do mapa que cobre o conjuntos de átomos selecionados. Outros critérios de avaliação do modelo cristalográfico serão apresentados na Seção 2.6.

Após a obtenção da melhor solução de substituição molecular procedeu-se com um refinamento de corpo rígido com os resíduos não conservados ainda reduzidos a alaninas e utilizou-se a resolução de 3,0Å.

Inicialmente, o modelo foi construído e manualmente modificado com o programa gráfico O (Jones *et al.*, 1991) examinando tanto mapas do tipo $2F_o - F_c$ quanto $F_o - F_c$ e subsequêntemente refinado usando verossimilhança máxima no programa REFMAC (Murshudov *et al.*, 1997) do pacote cristalográfico CCP4 (CCP4, 1994). Com o aparecimento de densidade eletrônica para as cadeias laterais previstas pela seqüência de aminoácidos as mesmas foram incluídas no modelo. Após o refinamento do novo modelo, novos mapas foram calculados e o processo iterado. Após 30 ciclos de refinamento os fatores R e R_{free} estavam em 0,27 e 0,33 faltando ainda incluir algumas regiões do modelo.

As modificações possíveis foram feitas, mas não foi observada a convergência do R_{factor} após 20 ciclos de refinamento. Acreditando-se ter atingido com esse refinamento uma situação na qual o refinamento não convergia para valores aceitáveis de R e R_{free} , foi tentado o procedimento de

simulated annealing (SA) com mapas de omissão e refinamento das cadeias laterais utilizando o espaço de ângulos torcionais do programa CNS (Brünger *et al.*, 1998). Um ciclo típico de minimização consistia de 40 ciclos internos de minimização dos termos estereoquímicos e cristalográficos, seguido de simulação de dinâmica molecular na qual a molécula era arrefecida lentamente a partir da temperatura inicial de 10.000K até uma temperatura final de 300K, com redução de 25K na temperatura do banho térmico a cada 25fsec. Após 13 ciclos de refinamento usando SA também não foi observada convergência satisfatória para os fatores R e R_{free} .

Numa inspeção criteriosa de toda a cadeia, resíduo por resíduo, levantou-se a suspeita da incompatibilidade para alguns resíduos entre o mapa densidade observado e o resíduo a ele atribuído. Desconfiando tratar-se da cristalização de uma isoforma de EcTI, diferente daquela usado para sequenciamento ou de eventuais erros na seqüência, decidiu-se reinterpretar o mapa por completo.

Realizou-se uma construção automática do modelo utilizando o programa ARP/wARP, uma vez que nossos dados de difração se estendem até 1,96 Å, resolução maior do que a mínima requerida para a utilização do programa (2,3 Å).

2.4 Interpretação Automática da Densidade Eletrônica: ARP/wARP

A partir de um conjunto de amplitudes de fatores de estrutura e estimativas das suas fases um mapa de densidade eletrônica pode ser calculado e conseqüentemente interpretado. Cada vez mais, métodos automatizados estão sendo empregados para esta finalidade. O programa ARP/wARP um destes métodos.

O programa ARP/wARP** (Perrakis *et al.*, 1999) carrega consigo o conceito de unificação dos passos iterativos de construção do modelo protéico e seu refinamento, em associação com o programa REFMAC (Murshudov *et al.*, 1997). Como informação inicial são utilizados mapas de densidade eletrônica e modelos parciais. Dentro desse conceito, a construção do modelo é vista como um processo integral, ou seja, que alterna ajustes ao modelo no espaço real e refinamento dos seus parâmetros no espaço recíproco.

Inicialmente o programa atribui átomos livres aos picos de densidade eletrônica do mapa fornecido, e partes deste modelo de átomos livres são automaticamente reconhecidas como elementos de uma proteína. Dessa forma, um modelo atômico protéico pode inicialmente ser construído, existindo nesta etapa um modelo “híbrido” de átomos livres e partes da

** ARP-Automated Refinement Procedure, wARP-weighted ARP

proteína construída. Esse modelo parcial fornece informações estereoquímicas sendo que o modelo de átomos livres descreve os picos de densidade eletrônica ainda não modelados. Iterativamente, através da construção do modelo e seu subsequente refinamento, obtêm-se um modelo híbrido com cada vez menos átomos livres e maiores partes da proteína construídas (Figura 15). Esse procedimento é chamado de warpNtrace.

Figura 15: Fluxograma do procedimento warpNtrace do programa ARP/wARP.

O procedimento warpNtrace foi usado para automaticamente construir um modelo protéico a partir do último mapa de densidade eletrônica encontrado utilizando o protocolo padrão em 20 ciclos envolvendo também o refinamento iterativo de cada novo modelo. Mais oito ciclos de refinamento com verossimilhança máxima no programa REFMAC foram usados para terminar o refinamento quando os fatores R e R_{free} pararam de cair. Durante este oito ciclos foram incluídas 147 moléculas de água, cada qual atribuída aos picos maiores que 3σ no mapa $F_{\text{obs}} - F_{\text{calc}}$, e um íon sulfato.

2.5 Modelagem Molecular por Homologia

O conceito de modelagem por homologia se baseia na conservação do enovelamento tridimensional entre membros de uma mesma família de proteínas (Blundell *et al.*, 1987, Greer, 1990, Sanchez & Sali, 1997).

O programa MODELLER 6a (Sali & Blundell, 1993) é freqüentemente utilizado em modelagem estrutural tridimensional de proteínas por homologia. Neste programa o usuário fornece o alinhamento da seqüência a ser modelada contra as estruturas relacionadas conhecidas e um modelo atômico será automaticamente calculado.

MODELLER modela a estrutura 3D da proteína pela satisfação de restrições espaciais. Em princípio, as restrições podem vir de diferentes tipos de fontes como: estruturas homólogas, refinamento por RMN, regras de empacotamento da estrutura secundária, espectroscopia de fluorescência, reconstrução por imagem em microscopia eletrônica, mutagênese sítio dirigida, intuição, potenciais de campo de força médio de átomo-átomo e resíduo-resíduo, etc. O MODELLER consegue obter automaticamente as restrições espaciais apenas de estruturas conhecidas homólogas sendo as outras restrições fornecidas pelo usuário.

No caso mais simples apenas as restrições por homologia juntamente com as restrições que descrevem a estereoquímica padrão de proteínas são utilizadas. Cada restrição, por exemplo uma distância entre dois C α da

estrutura homóloga conhecida, será utilizada na otimização do modelo através do alinhamento entre as seqüências. Cada restrição é codificada na forma de uma função de densidade de probabilidade (*Probability Density Function* - *pdf*) condicional, cuja forma pode depender de vários fatores, entre eles o tipo de aminoácido, sua acessibilidade ao solvente, identidade seqüencial entre as duas proteínas, etc. As *pdf*s individuais para cada restrição são usadas na construção de uma *pdf* molecular. Um modelo 3D é obtido pela otimização desta função de densidade de probabilidade molecular. Esta otimização ocorre através do procedimento de minimização da função alvo que emprega os métodos de gradiente conjugado (Press *et al.*, 1986) usando o método da função alvo variável (Brawn & Go, 1985) e dinâmica molecular com anelamento simulado usando o campo de força CHARMM (Brooks *et al.*, 1983).

Dentre as restrições utilizadas pelo programa estão: distância C α - C α , distâncias entre nitrogênios e oxigênios, existência de ligações de hidrogênio, ângulos ϕ , ψ e ω , ângulos diédricos e restrições quanto a estereoquímica padrão.

Um alinhamento entre as seqüências do inibidor inicialmente fornecida e a efetivamente encontrada através do mapa diferença foi realizado utilizando-se o programa MULTALIGN do pacote AMPS (Barton, 1990) usando a matriz PAM250 e uma penalidade a gaps de 8 e posteriormente ajustado manualmente. Cinquenta modelos foram construídos envolvendo um protocolo que faz uma cópia das coordenadas

atômicas comuns às duas seqüências e usa a biblioteca de topologias do programa CHARMM para obter as coordenadas restantes. Uma randomização de 4 Å nas coordenadas foi aplicada a cada uma das cinquenta estruturas antes da otimização que utilizou até 300 ciclos do método da função alvo variável empregando o método do gradiente conjugado. O refinamento foi realizado no espaço cartesiano utilizando um algoritmo de dinâmica molecular padrão com anelamento simulado como implementado no MODELLER6a. Limites nas distâncias para restrições inter-atômicas foram 14 Å (para distância C α -C α), 11 Å (para distâncias N-O), 5,5 Å (para distâncias entre átomos da cadeia principal com cadeia lateral) e 5,0 Å (cadeia lateral com cadeia lateral). Os cinquenta modelos foram inicialmente avaliados usando a função energia (ou melhor pseudo-energia) do programa MODELLER, que mede quão bem foram satisfeitas as restrições espaciais para um dado modelo. Subseqüentemente outras avaliações foram empregadas como descritas a seguir.

2.5.1 Modelagem dos Complexos

No caso da construção dos modelos de complexo, uma busca no PDB (Berman *et al.*, 2000) foi realizada por estruturas resolvidas de serinoproteases complexadas com inibidores do tipo Kunitz. O complexo com estrutura resolvida a mais alta resolução foi escolhido com base na sua atividade apresentada nos ensaios de (Batista, 1997). Uma vez encontrado o candidato foi feito um alinhamento estrutural entre o EcTI e o inibidor da

estrutura do complexo depositado no banco PDB através de seus C α utilizando-se o programa gráfico O (Jones *et al.*, 1991). No caso do complexo ser de protease com inibidor não protéico o alinhamento foi realizado tendo os C α do sítio reativo, P3 a P3', como referência. O modelo para o complexo entre EcTI e tripsina de porco foi feito através da sobreposição do modelo de EcTI encontrado por homologia e o inibidor da tripsina porcina (PDB 1AVW) formando um modelo inicial do complexo que foi minimizado, como será descrito adiante. O modelo para o complexo entre EcTI e trombina foi feito através da sobreposição do modelo de EcTI e o inibidor de tripsina de pâncreas bovino (PDB 1BTH) tendo como referência os resíduos P3 a P3'. Da mesma forma procedeu-se na sobreposição do EcTI e o inibidor não protéico de fator Xa (PDB 1EZQ) para o modelo inicial do complexo entre EcTI e fator Xa. Numa tentativa de explicar as capacidades de EcTI de inibir quimotripsina além da tripsina foi usado o complexo entre a quimotripsina e BPTI. O EcTI foi sobreposto em BPTI usando os resíduos P3 a P3' como descrito acima. O modelo não foi minimizado. A modelagem manual da Arg63 na conformação da Lys15 de BPTI foi realizada com o programa O com o intuito de examinar a viabilidade desta Arg assumir a mesma conformação sem gerar impedimentos estéricos.

A minimização de energia dos complexos foi realizada com 4000 passos do algoritmo *Steepest descent* e campo de força AMBER, considerando o complexo em água (implícita) (constante dielétrica $\eta=80$), no

programa DISCOVER do pacote InsightII e os seus fatores de qualidade foram calculados como descritos a seguir.

2.6 Critérios de Qualidade

Diversas metodologias existem para a checagem e avaliação da qualidade de uma estrutura, seja ela determinada experimentalmente ou por modelagem por homologia. Segue uma breve descrição de algumas delas.

2.6.1 PROCHECK

O PROCHECK^{††} (Laskowski *et al.*, 1993) analisa tanto a estereoquímica de cada resíduo individualmente quanto a estereoquímica global da estrutura tendo como parâmetros informações derivadas de estruturas resolvidas a alta resolução (Morris *et al.*, 1992) que constituem sua base de dados. As informações estereoquímicas checadas pelo programa são: ligações covalentes, planaridade de grupos planares (aromáticos, ligações peptídicas, etc.), ângulos diédricos, quiralidade, interações não covalentes, ligações de hidrogênio da cadeia principal e pontes de dissulfeto.

Dentre as várias análises fornecidas pelo PROCHECK estão: análise de estrutura secundária, acessibilidade estimada por resíduo, gráfico de Ramachandran e *G-factors* para comprimentos de ligação da cadeia

^{††} <http://biotech.embl-heidelberg.de:8400/chk/ProCheck/prochecks.html>

principal, ângulos da cadeia principal, distribuição dos ângulos ϕ/ψ , distribuição dos ângulos χ_1/χ_2 e ω .

O *G-factor* fornece uma medida do desvio de uma dada propriedade estereoquímica (ângulos torcionais, distâncias de ligações covalentes e ângulos da cadeia principal, por exemplo), baseada nas distribuições observadas destes parâmetros estereoquímicos em sua base de dados. Para um dado resíduo, um baixo valor do *G-factor* indica uma baixa probabilidade de ocorrência do parâmetro que se observa. Pode-se calcular o *G-factor* global que envolve os diversos *G-factors* para toda a estrutura, num único número. O *G-factor* global é a melhor medida da qualidade total da estereoquímica de uma molécula.

O diagrama de Ramachandran (Ramakrishnan & Ramachandran, 1965) indica a distribuição dos ângulos ϕ e ψ dos resíduos pertencentes a uma determinada estrutura. Seus indicadores de qualidade baseiam-se na análise de 118 estruturas de resolução até pelo menos 2,0 Å de resolução e R_{factor} menor 20%. Nos seus critérios uma estrutura de boa qualidade deve ter 90% ou mais de seus resíduos nas regiões mais favoráveis.

2.6.2 WHATCHECK

O programa WHATCHECK realiza várias checagens diferentes numa estrutura, entre elas a ocupação atômica, a consistência e validade dos operadores de simetria e a qualidade global do modelo em termos de contatos atômicos^{††}. Esta última chamada de *QUALITY check* (Vriend & Sander, 1993) que compara o ambiente químico de cada resíduo com o ambiente químico médio de todos os resíduos do mesmo tipo em uma base de dados composta de arquivos PDB de boa qualidade. Mais especificamente, cada fragmento rígido de cada resíduo da estrutura é individualmente sobreposto num fragmento de referência junto com sua distribuição atômica local, numa caixa de $16 \times 16 \times 16 \text{ \AA}^3$. A comparação entre a distribuição presente na estrutura e a distribuição de referência em função da estrutura secundária do resíduo e da sua acessibilidade ao solvente, permite a determinação da qualidade do resíduo, que é expressa em termos de um desvio em relação à média esperada. O valor médio para as 180 estruturas de boa qualidade do seu banco de dados é $-0,5 \pm 0,4$. Sua classificação quanto aos valores médios segue a regra:

- $-3,0$ - o modelo é considerado ruim;
- $-2,0$ - o modelo é considerado muito pobre;
- $-1,5$ - o modelo é considerado bom;
- $-1,0$ até $-0,5$: o modelo é de muito boa qualidade;
- $> -0,5$ - o modelo é considerado “perfeito”.

2.6.3 VERIFY-3D

O VERIFY-3D^{§§} (Lüthy *et al.*, 1992) mede a compatibilidade entre a seqüência de aminoácidos de uma proteína e o modelo da sua estrutura tridimensional. Para tanto, o programa utiliza a seguinte metodologia: 1) reduz a estrutura 3D a uma seqüência de ambientes químicos (um para cada resíduo) categorizados entre 18 possibilidades de acordo com a área do resíduo exposta ao solvente, com a fração de contatos feitos com átomos polares e a estrutura secundária local (perfil 3D), 2) utilizando uma matriz (chamada matriz 3D-1D (Bowie *et al.*, 1991)) que descreve a probabilidade de se encontrar cada um dos 20 aminoácidos em cada uma das 18 classes de ambientes químicos, o programa determina a probabilidade, por resíduo, de acordo com o tipo de aminoácido e a natureza do ambiente químico calculado no passo anterior. A matriz 3D-1D é previamente conhecida e derivada de uma análise de estruturas conhecidas de boa qualidade. O resultado é uma medida da compatibilidade entre a seqüência e sua estrutura 3D descrita pelo seu perfil tridimensional. Lüthy *et al.* (1992) determinaram empiricamente o índice global esperado de $S_{\text{calc}} = \exp(-0,83 + 1,008 \times \ln(L))$ de compatibilidade entre a seqüência e a estrutura 3D onde L é o comprimento da seqüência e S_{calc} representando, então, a soma das probabilidades individuais dos resíduos. Lüthy *et al.* também sugerem um limite inferior de $0,45 \times S_{\text{calc}}$ para a confiabilidade da compatibilidade seqüência/estrutura. Valores menores que $0,45 \times S_{\text{calc}}$

^{##} <http://biotech.ebi.ac.uk:8400/chk/whatif/>

^{§§} Este programa pode ser rodado on-line ✓: www.doe-mbi.ucla.edu/Services/Verify_3D/

indicam uma estrutura incorreta, valores em torno de $0,45 \times S_{\text{calc}}$ podem estar corretas mas possuem qualidade questionável e valores próximos ou acima de S_{calc} indicam uma estrutura confiável.

Além da qualidade global, a variação da qualidade ao longo da estrutura pode ser avaliada. Essa avaliação é realizada utilizando-se a média das probabilidades individuais promediadas dentro de uma janela deslissante (tipicamente 25 resíduos) no sentido do N para o C-terminal, produzindo um perfil de qualidade ao longo da molécula. Isto pode ser útil na identificação de regiões problemáticas na estrutura.

A qualidade da estrutura cristalográfica final da EcTI, após a convergência do fator R e R_{free} , foi avaliada com os programas acima mencionados. Em PROCHECK as informações estereoquímicas foram checadas considerando 2,0 Å de resolução como *cutoff*. A compatibilidade entre a sequência de aminoácidos de EcTI e o modelo de sua estrutura foi averigüada através da versão on-line de VERIFY-3D. A qualidade global do modelo em termos de contatos atômicos de cada resíduo com o ambiente estereoquímico médio e seu desvio em relação à média esperada foi calculada através do programa QUALTY de WHATCHECK.

Resultados e Discussão

Capítulo Três

3 Resultados e Discussão

3.1 Cristalização, Coleta e Processamentos de Dados

Do procedimento descrito na Seção 2.1, e após três semanas, sob a mesma condição de cristalização (15% PEG8000, 0,5M LiSO₄) que levou anteriormente aos cristais de hábito de aglomerado, monocristais apropriados à coleta de dados foram observados no capilar (Figura 16, abaixo).

Figura 16: Cristal de EcTI crescido com a técnica de acupuntura em gel, na condição 15% PEG8000, 50mM LiSO₄.

A redução dos dados e seu escalonamento foram realizados utilizando os programas DENZO (auto-indexação, refinamento e integração) e SCALEPACK (escalonamento dos dados) (Otwinowski & Minor, 1997) pelo operador da linha PCr no LNLS e pela Dra. Batista. Foram obtidas 11.610 reflexões únicas de um total de 23.475 observações com um R_{merge} de 10,3%, correspondendo à aproximadamente 87% do número de reflexões esperadas a 1,96 Å de resolução (Tabela 5).

O valor observado para a completeza está um pouco abaixo do ideal devido, em parte, à formação de anéis de gelo. O processamento das imagens para a retirada destes anéis induziu a baixa completeza para algumas faixas de resolução (2,83-2,66, 2,24-2,17, por exemplo), como pode ser visto na Tabela 6.

Tabela 5: Estatísticas da coleta de dados e processamento do cristal do EcTI. Os valores entre parênteses referem-se à última faixa de resolução (2,01-1,96Å).

Estatísticas da coleta de dados e processamento	
Grupo espacial	P2 ₁
Parâmetros da cela unitária (Å); (°)	a=37,12; b=38,42; c=54,08; β=98,08
Faixa de resolução (Å)	20,00 – 1,96
Número de reflexões coletadas	23.475
Número de reflexões únicas	11.610
Redundância	2,1
$R_{\text{sym}}^{\dagger}$ (%)	10,3 (31,1)
Completeza (%)	87,0 (87,0)
$(I/\sigma(I)) < 2$ (%)	20,1 (51,5)

$$^{\dagger} R_{\text{sym}} = \frac{\sum_h \sum_i |I(h)_i - \langle I(h) \rangle|}{\sum_h \sum_i I(h)_i}$$

Tabela 6: Estatística da coleta de dados de difração de raios-X por faixa de resolução para EcTI.

Faixa de resolução Å	%Completeza	% Refl. com $(I/\sigma(I)) <$		R_{merge}
		2	3	
20,00 – 4,81	88,5	2,7	3,9	0,059
4,81 – 3,83	92,0	1,3	3,3	0,076
3,83 – 3,35	93,8	3,6	5,0	0,096
3,35 – 3,04	92,8	8,3	12,8	0,128
3,04 – 2,83	95,1	9,7	13,5	0,116
2,83 – 2,66	71,3	9,5	15,2	0,136
2,66 – 2,53	95,6	17,3	25,7	0,145
2,53 – 2,42	79,3	18,5	26,1	0,268
2,42 – 2,32	94,0	23,5	34,4	0,191
2,32 – 2,24	94,4	25,8	34,4	0,348
2,24 – 2,17	69,8	21,0	30,9	0,206
2,17 – 2,11	94,7	37,7	49,7	0,237
2,11 – 2,06	95,1	42,2	55,3	0,262
2,06 – 2,01	60,4	31,6	40,1	0,282
2,01 – 1,96	87,0	51,5	61,5	0,311
Todos hkl	87,0	20,1	27,2	0,103

O cristal pertence ao grupo espacial $P2_1$ com parâmetros de rede $a=37,12$ Å, $b=38,42$ Å e $c=54,08$ Å e $\beta=98,08^\circ$ e uma molécula por unidade assimétrica. O número de Matthews, V_M (Matthews, 1968) calculado foi de $1,96$ Å³Da⁻¹ estimando o conteúdo de solvente no cristal de 36,82% (v/v) e uma densidade cristalina (Drenth, 1999) de $1,34$ gcm⁻³. O valor de V_M está próximo do limite inferior esperado ($1,7$ Å³Da⁻¹) correspondendo a um cristal com baixo conteúdo de solvente, compatível com o fato de o cristal difratar a relativamente alta resolução ($1,7$ Å). O número de moléculas por unidade assimétrica encontrado foi um, uma vez que assumindo duas moléculas por unidade assimétrica obtém-se $V_M = 0,98$ Å³Da⁻¹ que corresponde a um

conteúdo negativo de solvente no cristal, indicando ser este um dado não realístico.

3.2 Resolução da Estrutura

A estrutura do EcTI foi determinada por substituição molecular utilizando-se o programa AMoRE (Navaza & Saludjian, 1997) do pacote CCP4 (CCP4, 1994) utilizando-se as coordenadas atômicas do inibidor de tripsina de *Erythrina caffra* resolvido a 2,5Å de resolução (PDB 1TIE) (Onesti *et al.*, 1991) como molde que apresenta 39,29% de identidade com EcTI (Tabela 4, pág.: 37). O coeficiente de correlação encontrado foi de 33,3%. Os ângulos Eulerianos da solução da função roto-translação foram $\alpha=63,2^\circ$, $\beta=9,5^\circ$, $\gamma=141,0^\circ$ e as translações em coordenadas fracionárias $x=0,3999$, $y=0,0000$, $z=0,2396$.

O contraste entre as potenciais soluções fornecidas por AMoRE pode ser visto na Tabela 7. O valor relativamente baixo do coeficiente de correlação observado para a primeira solução provavelmente se deve à baixa identidade seqüencial entre 1TIE e a seqüência inicialmente proposta do EcTI (Batista *et al.*, 1996) e à clivagem interna presente em sua cadeia polipeptídica que poderia levar a uma diferença marcante na estrutura local deste *loop* (veja mais adiante).

Tabela 7: Contraste entre os coeficientes de correlação das soluções encontradas por AMoRE. A melhor solução apresentou um coeficiente de correlação de 33,3%, enquanto que as outras potenciais soluções se concentraram em torno de 29,2%. α , β , γ são ângulos Eulerianos em graus e x, y, z correspondem a translações em coordenadas fracionárias.

α	β	γ	x	y	z	Cc
63,2	9,5	141,0	0,3999	0,0000	0,2396	33,3
59,9	9,3	143,9	0,4786	0,0000	0,2301	29,2
64,8	10,6	138,4	0,0402	0,0000	0,2086	29,4
63,6	9,7	140,7	0,3246	0,0000	0,4819	29,4
62,6	9,2	141,4	0,3420	0,0000	0,3273	28,7
68,3	9,6	136,4	0,4496	0,0000	0,3636	29,2
64,5	9,4	140,0	0,0621	0,0000	0,0618	28,6
67,0	9,6	136,9	0,3449	0,0000	0,0722	28,8
70,2	9,3	134,6	0,4214	0,0000	0,4092	28,9
68,0	10,2	136,0	0,2822	0,0000	0,2819	28,7

3.3 Refinamento da Estrutura

Procedeu-se com um refinamento de corpo rígido com os resíduos não conservados ainda reduzidos a alaninas e utilizou-se a resolução de 3,0Å, com o R_{factor} mantendo-se essencialmente constante (de 0,523 para 0,520) e o R_{free} caindo de 0,51 para 0,50.

Inicialmente, o modelo foi construído e manualmente modificado com o programa gráfico O (Jones *et al.*, 1991) examinando-se tanto mapas do tipo $2F_o - F_c$ quanto $F_o - F_c$ e subsequente refinado utilizando verossimilhança máxima no programa REFMAC (Murshudov *et al.*, 1997) do pacote cristalográfico CCP4 (CCP4, 1994). Com o aparecimento de densidade eletrônica para as cadeias laterais, as mesmas foram incluídas no modelo. Após o refinamento do novo modelo, novos mapas foram calculados e o processo iterado. Após 10 ciclos de refinamento moléculas de água

foram incluídas em regiões de densidade compatível, mas sem sucesso na melhora dos fatores R. Após 20 ciclos de refinamento os fatores R e R_{free} estavam em 0,27 e 0,33 faltando ainda incluir algumas regiões do modelo. Acreditando-se ter atingido com esse refinamento uma situação na qual não convergiria para valores aceitáveis de R e R_{free} , foi tentado o procedimento de *simulated annealing* que também não foi capaz de reduzir os valores dos fatores R.

No entanto, ao longo destas tentativas puderam ser observadas divergências em algumas posições da estrutura quanto ao mapa de densidade eletrônica observado e o resíduo a ele atribuído naquela posição da seqüência. O uso de mapas de omissão nestes casos juntamente com o alinhamento com outros inibidores homólogos levou à reconstrução do modelo usando resíduos cujas cadeias laterais eram compatíveis com a densidade experimental. Desta forma, a molécula teve sua seqüência publicada reavaliada através do mapa de densidade eletrônica.

Utilizou-se o programa ARP/wARP (Seção 2.4) para a construção automática das cadeias principal e lateral. Após as mutações, aliando mapa de densidade e sugestão do programa, terem sido efetuadas foram realizados ainda outros oito ciclos de refinamento com REFMAC. Os valores de R_{free} e R_{factor} no último ciclo de refinamento estavam em 0,25 e 0,17, respectivamente, o que foi considerado uma convergência satisfatória. Apesar da diferença entre R e R_{free} ser relativamente alta ela está dentro dos padrões sugeridos por Kleywegt & Jones, (1997) os quais indicam que bons

conjuntos de dados podem fornecer diferenças da ordem de 0,02 e dados com qualidade inferior ou baixa resolução, a diferença pode chegar a 0,08. O R_{free} foi calculado através de 5% dos dados retirados aleatoriamente do conjunto de dados inicial.

A Tabela 8 mostra o alinhamento entre as seqüências publicada (Batista *et al.*, 1996) e a efetivamente encontrada através do mapa de densidade eletrônica. Os resíduos marcados em cinza fazem parte do segmento dos resíduos 91-117 que não contaram com *overlap* de peptídeos e fragmentos no sequenciamento, tendo tido suas posições definidas pela identidade seqüencial com as estrutura primária de outros inibidores da família Kunitz já determinadas.

Tabela 8: Alinhamento entre as seqüências esperada (seq.propos) e a efetivamente encontrada (EctI). Os resíduos em vermelho representam as divergências encontradas, salientado em azul está o resíduo P1 do sítio reativo e em preto um resíduo não definido na determinação da seqüência primária. Na região cinza estão marcados resíduos que tiveram suas posições definidas pela identidade seqüencial com as seqüências de outros inibidores da família Kunitz já determinadas.

	1	10	20	30	40	50	60
seq.propos ecti	KELLDSDGDIL	RNGGTYIILPA	LRGKGGGLELA	KTGDETCPLNVVQAR	GETKRG	RPAIIW	
ecti Enterolobi	ELLDSDGDIL	SNGGTYIILPA	GR KGGGLELA	ASGDETCPLNVVQAR	SETKSG	KPAIIW	
	61	70	80	90	101	110	120
seq.propos ecti	TPP	RIALTPAFYLNIEF	QTKDLPA	CLREYSRLPREE	QHSEVKLAP	KEEAAFG	X EKL
ecti Enterolobi	TPP	SVATLTPAFYLNIEF	AGAALAGCL	GGYSRLPWKVE	SEVKIAP	SADESAS	SGGSFKI
	121	130	140	150	160	170	
seq.propos ecti	KPYRDDYK	IVYCEGGSDDDS	CKDLGISIDDENNR	RLVVKDGDPLAVRFV	KAHRRG		
ecti Enterolobi	KPYKDDYK	LVYC	CS	DLGISIDDENNR	RLVVKDGDPLAVRFV	A	

Cadeia leve

A Figura 17 mostra algumas divergências e substituições realizadas com correspondente mapa de densidade. Todas as figuras que envolvam mapa de densidade eletrônica e cadeias laterais foram feitas com o programa PyMOL (DeLano, 1998).

Figura 17: Algumas divergências evidentes no mapa de densidade eletrônica encontradas durante a construção do modelo de EcTl. Em a) a substituição da Arg pelo Trp94 é reforçada pela conservação deste Trp em outras seqüências de inibidores do tipo Kunitz de vegetais (Tabela 10, pág.: 82). b) Substituição de Ile por Thr66, c) substituição de Gly por Ser48. d) Substituição de Thr por Ser34, e) substituição de Ser por Pro93 e f) substituição de Arg por Lys54. Em verde esta representada a seqüência esperada. O mapa $2F_o - F_c$ está mostrado em cada caso.

O fator de correlação entre mapa e a cadeia lateral atribuída a uma posição espacial é medida pelo RS_fit do programa gráfico O (Jones *et al.*, 1991). Para todas as substituições realizadas, este parâmetro mostrou uma melhor correlação com o mapa diferença para os resíduos substituídos. Como exemplo, as substituições mostradas na Figura 17 mostram a seguintes mudanças com relação ao mapa final (entre parênteses vem o valor de RS_fit para o resíduo antes da substituição): Ser34 – 0,96 (0,83), Ser48 – 0,96 (0,88), Lys54 – 0,94 (0,53), Thr66 – 0,91 (0,64), Pro93 – 0,96 (0,85), Trp94 - 0,94 (0,71). Foram realizadas um total de 33 substituições e

uma inserção em cujo caso a seqüência final está mostrada na tabela (Tabela 8).

Algumas das modificações feitas merecem comentário. É provável que a densidade compatível com a serina encontrada na posição 63 (P1 do sítio reativo) seja puramente uma consequência da desordem na cadeia lateral e o resíduo presente seja de fato uma arginina, como esperado para um inibidor de tripsina. A maioria das outras substituições e inserções são provavelmente sem consequências importantes, mas pelo menos duas merecem consideração. A Figura 18 mostra o ambiente do Trp95 (Figura 17a) modelado no lugar da arginina prevista por Batista *et al.* (1996). Claramente uma arginina nesta posição é quimicamente desfavorável por causa do ambiente hidrofóbico em volta. Além disso este triptofano é fortemente conservado nesta posição em outros inibidores. Especialmente próximo, na posição 117, um ácido glutâmico previsto por Batista *et al.* (1996) (Tabela 8) mostrou densidade compatível com uma fenilalanina (Figura 18).

Figura 18: Especialmente próximo ao Trp94 (fortemente conservado em outros inibidores) está a Phe117, com densidade compatível, diferentemente do Glu (em amarelo na mesma posição que a Phe117) previsto por Batista *et al.* (1996).

Estes resultados levantam uma questão: foi cristalizada uma isoforma diferente do inibidor comparado-se com a amostra sequenciada ou existem erros na seqüência de aminoácidos? No caso dos dois aminoácidos citados acima, devido ao ambiente hidrofóbico onde se situam, parece provável que foi um erro de sequenciamento. É interessante que ambos os casos Arg95Trp e Glu117Phe pertençam a uma região da seqüência onde não houveram sobreposições durante o sequenciamento (Batista *et al.*, 1996), sendo o sequenciamento feito por homologia. Porém, mesmo neste caso, há dúvida pois a formação de uma ponte salina entre Arg95 e Glu117 (na seqüência original) é uma possibilidade que não pode ser descartada. Pontes salinas enterradas em proteínas são raras mas existem. Por exemplo a ponte salina entre o N-terminal e Asp194 da quimotripsina após ativação do quimotripsinogênio. Portanto, ainda restam dúvidas se a amostra cristalizada era uma isoforma da molécula sequenciada.

A dificuldade e precisão com que se pode identificar cadeias laterais de aminoácidos num mapa de densidade eletrônica depende da resolução dos dados coletados, do grau de desordem estrutural local no cristal e qualidade da estimativa das fases. As desordens na rede cristalina eliminam as intensidades difratadas a mais alta resolução e o movimento de muitas cadeias laterais na superfície da molécula torna estes grupos parcialmente ou totalmente não visíveis (Anderson *et al.*, 1978a). Vários trabalhos (Watenpaugh *et al.*, 1973, Lipscomb *et al.*, 1969; Kendrew *et al.*, 1961) tentaram encontrar a seqüência de aminoácidos a partir apenas do mapa de

densidade eletrônica, nestes casos os mapas possuíam resolução entre 2,3 a 1,5 Å. A comparação entre as seqüências produzidas desta forma e as seqüências primárias obtidas posteriormente mostrou que as seqüências derivadas do mapa estavam entre 50 a 60% corretas. Anderson *et al.* (1978a) afirmam que mesmo com os erros na identificação do resíduo, o modelo pode ser acurado o suficiente para se proceder com o refinamento (Anderson *et al.*, 1978b) das coordenadas cristalográficas.

Há vários métodos ou procedimentos (Kleywegt & Jones, 1997, Kleywegt & Jones, 1995) que podem ser usados para prevenir ou evitar que o modelo não seja afetado por subjetividades como experiência acumulada, pré-concepções, expectativas e práticas locais (Kleywegt, 2000). Mesmo com dados a alta resolução e boas fases, e 95% do modelo sendo consequência dos dados, ainda resta neste caso a subjetividade no sentido de: 1) que programa de refinamento utilizar; 2) modelar ou não explicitamente os átomos de hidrogênio; 3) como modelar os fatores de temperatura; 4) quais picos no mapa interpretar como moléculas de solvente, etc. As práticas de prevenção citadas acima são: o uso de bases de dados de estruturas bem resolvidas na construção do modelo (Kleywegt & Jones, 1998), o uso de indicadores de qualidade local e o uso de indicadores de qualidade global. É necessário notar que erros são quase que inevitáveis e que estas práticas ajudam a detectar e a consertar esses erros para que se obtenha o melhor modelo possível, que ainda será interpretado em termos do papel biológico da proteína que ele representa.

Porém, o sucesso do refinamento do modelo, sua compatibilidade com o mapa de densidade eletrônica e com as seqüências de outros inibidores Kunitz levam a crer que a seqüência gerada por esta abordagem está mais próxima à real, para a amostra cristalizada.

3.4 Avaliação da Qualidade da Estrutura com Seqüência Modificada

O diagrama de Ramachandran (Figura 19) mostra que 97,6% dos resíduos estão em regiões permitidas (regiões A, B, L, a, b, l, p no gráfico). Os resíduos Ala81, Lys124 e Asp126 representam 2,4% dos resíduos não Pro e não Gly de EcTI e estão situados nas regiões generosamente permitidas do gráfico.

Figura 19: Diagrama de Ramachandran para EcTI. As 17 Gly presentes na estrutura são mostradas como triângulos.

A Figura 20 ilustra estes resíduos e suas densidades eletrônicas. Os resíduos Lys124 e Asp126 fazem parte do *loop* que aparece na parte

superior e o resíduo Ala81 aparece na parte inferior da Figura 20a. Ao lado direito da Figura 20a está mostrado o *loop* reativo e sua densidade eletrônica, observando-se a distante posição relativa destes resíduos em relação ao sítio reativo.

A qualidade do mapa final ($2F_o - F_c$) para Asp126 e Ala81 é boa, não deixando dúvida de que foram bem modelados. Tanto Lys124 quanto Asp126 fazem parte de um *loop* entre os motivos 2 e 3 da estrutura. A conformação deste *loop* varia entre membros da família Kunitz, sendo desordenado na estrutura de SBTI (PDB 1BA7) (de Meester *et al.*, 1998). Ala81 faz parte do *loop* do motivo 2 (descrito mais adiante), que também demonstra grande variabilidade de conformação para a cadeia principal entre SBTI, o inibidor de *E.caffra* (1TIE) e EcTI. O mapa de densidade eletrônica final neste caso não deixa dúvida que o percurso da cadeia principal foi traçado corretamente. O fator de correlação entre o mapa e a cadeia lateral atribuída a uma posição espacial é medida pelo RS_fit do programa gráfico O (Jones *et al.*, 1991). Para estes resíduos temos os fatores de correlação 0,90 para Ala81, 0,96 para Asp126, e 0,92 para Lys124.

Um resumo dos principais parâmetros estereoquímicos da cadeia principal estão mostrados na Figura 21, indicando que todos estão dentro da faixa de valores esperados, dada a resolução da estrutura.

Figura 20: a) Posição relativa ao *loop* reativo (vermelho) dos resíduos Ala81, Lys124, Asp126 presentes em regiões generosamente permitidas no diagrama de Ramachandran. b), c), d) mostram suas densidades eletrônicas, no mapa final $2F_o - F_c$.

Figura 21: Parâmetros de cadeia principal para EcTI fornecidos por PROCHECK.

O mapa de densidade eletrônica para a cadeia principal é contínuo exceto para o resíduo 24, para a região do *loop* 97-101 e para a região entre os resíduos 132 e 141, que não foram incluídos no modelo. As três regiões correspondem a *loops* na estrutura. A última região corresponde ao ponto onde o EcTI imaturo é processado para gerar as cadeias pesada (α) e leve (β) do EcTI maduro. Este processamento acontece dentro da semente e leva à formação de uma cadeia pesada (resíduos 1-134) e leve (135-170). O último resíduo observável da cadeia pesada é a Cys132 e o primeiro resíduo observável da cadeia leve é a Cys141. A cadeia lateral do resíduo 63 foi modelada como Ser por não haver mapa de densidade eletrônica compatível com a cadeia lateral da Arg esperada. Supõe-se que muito provavelmente os átomos além do C γ , nas diversas moléculas ao longo do cristal, apresentem vibrações térmicas de grande amplitude, ou desordem estática.

A presença de *loops*, desordenados nas estruturas de inibidores da família Kunitz de plantas é comum. O *loop* dos resíduos 97-101 também se encontra desordenado no inibidor de *E.caffra* e o *loop* em torno do sítio de processamento de EcTI (132-141) é igualmente desordenado no SBTI, apesar deste não sofrer o mesmo processamento proteolítico. A presença de *loops* grandes e desordenados em membros da família portanto, parece ser uma característica que não interfere na atividade inibitória. Porém, não há nenhum padrão claro de conservação neste sentido e a ordenação ou não de um determinado *loop* pode ser em maior parte uma função do empacotamento cristalino do que uma propriedade inerente da estrutura.

A estrutura final consiste de 160 resíduos de aminoácidos (1165 átomos), 147 moléculas de água e um íon sulfato.

3.5 Descrição da Estrutura

A molécula EcTI possui estrutura similar à de outros inibidores de serinoproteases da família Kunitz. Uma sobreposição de três estruturas homólogas (ETI, SBTI e EcTI) utilizando-se os C α é mostrada na Figura 22 abaixo. As moléculas desta família apresentam uma topologia que consiste de doze fitas- β anti-paralelas conectadas por *loops* longos observados na superfície da molécula. Os *loops* externos, incluindo a região do sítio reativo, apresentam desvios de uma estrutura para outra embora o *core* da estrutura seja estruturalmente conservado. A variação estrutural do *loop* pode ser

vista na Figura 22 onde fica evidente que as regiões mais variáveis estão dispostas, relativamente ao sítio reativo, do lado oposto da molécula. O sítio reativo, apesar de variar um pouco de uma estrutura para outra, é muito mais conservado em termos de conformação, como será discutido adiante. Os rmsd de comparação entre EcTI e SBTI e entre EcTI e ETI foram de 1,41 e 1,18 Å, respectivamente.

Figura 22: Alinhamento estrutural com C α entre ETI (verde), SBTI (vermelho) e EcTI (azul). O destaque em “ball and stick” é para o resíduo P1. O *loop* da molécula que inclui o sítio reativo é mais conservado estruturalmente entre inibidores Kunitz, comparativamente aos outros *loops*.

Seis das doze fitas em todos os membros da família Kunitz formam um barril- β e as outras seis formam uma “tampa” para o barril composta de três grampos- β . O barril presente nestas estruturas é uma consequência do motivo estrutural conhecido como motivo *trefoil* descrito por Murzin *et al.* (1992) e McLachlan (1979) e ilustrado na Figura 23. Cada motivo é constituído de duas fitas antiparalelas que participam do barril e duas fitas que formam um dos grampos de cabelo. O conjunto de três motivos *trefoil*

forma um β -*trefoil* (Figura 23b) levando assim ao barril de seis fitas e à “tampa” de três grampos.

Figura 23: O arranjo β -*trefoil* (direita) é formado a partir de três unidades *trefoil* (esquerda). As duas fitas superiores dos três motivos formam o barril de 6 fitas e os 3 grampos de cabelo um arranjo triangular, ou “tampa” do barril.

Devido possuir três motivos *trefoil* a estrutura como um todo apresenta um eixo de pseudo simetria de ordem três que passa pelo centro do barril e entre os três grampos de cabelo. As unidades *trefoil* não são detectáveis por análise de seqüência, mas são facilmente observáveis a partir da sua sobreposição e alinhamento estrutural. A Figura 24a mostra os motivos *trefoil* do EcTI sobrepostos ficando evidente que são topologicamente equivalentes. No caso do EcTI, o desvio médio quadrático da sobreposição, em média é 0,8 Å para os resíduos pertencentes às fitas.

O enovelamento β -*trefoil* é um motivo estrutural também encontrado em proteínas muitas vezes não relacionadas entre si (Chothia & Janin, 1982). Como por exemplo, interleucina-1 α (Graves *et al.*, 1990), interleucina-1 β (Finzel *et al.*, 1989, Priestle *et al.*, 1989), fatores de crescimento de

fibroblasto ácido e básico (Zhu *et al.*, 1991), toxinas do tipo ricina e proteínas de empacotamento de actina tipo histactofilinas, além dos inibidores de proteases da família Kunitz. As funcionalidades das diferentes famílias são atribuídas a aspectos estruturais variáveis na superfície da molécula. O enovelamento β -*trefoil* provavelmente caracteriza um arcabouço sobre o qual funcionalidades diversas podem ser embutidas através de variações, particularmente nos *loops* na superfície. Foi proposto por McLachlan (1979) que este tipo de enovelamento talvez possa ser o resultado da triplicação de um mesmo gene e Ponting & Russel (2000) apresentam evidências estatísticas de que todas as proteínas com enovelamento β -*trefoil* sejam homólogas, reforçando a idéia de um ancestral comum.

Figura 24: a) Sobreposição dos três motivos *trefoil* de EcTI; com o primeiro motivo em vermelho, o segundo em azul e o terceiro em amarelo. Em rosa está a representação em fita do motivo vermelho. b) Os três motivos *trefoil* em conjunto formando o β -*trefoil* de EcTI.

A Tabela 9 mostra o alinhamento de diferentes inibidores de protease com os motivos estruturais acima citados em destaque.

Tabela 9: Alinhamento entre sete diferentes inibidores de protease da família Kunitz. HP indica *HairPin* (grampo de cabelo) e FiB indica a *i*-ésima Fita do Barril. EcTI = *Enterolobium Contortisiliquum trypsin inhibitor* (sequência encontrada pelo mapa de densidade), 1TIE = *Erythrina caffra trypsin inhibitor*, 1BA7 = *soybean trypsin inhibitor*, 2WBC = *winged bean chymotrypsin inhibitor*, 1WBA = *winged bean albumin*, 1AVA = *barley α -amylase inhibitor*. A Tabela 8 mostra a sequência inicialmente proposta para o EcTI e a sequência aqui encontrada. Em magenta está marcado o resíduo P1 do *loop* reativo situado entre o primeiro e o segundo motivo *trefoil*. As fitas estão marcadas em cinza e os resíduos dessas fitas que compõem o núcleo hidrofóbico estão marcados em amarelo. Os resíduos dos grampos de cabelo que também fazem parte do núcleo hidrofóbico estão marcados em azul (resíduos das fitas internas do grampo de cabelo) e em verde (resíduos das fitas externas).

			---F1B---		-----HP1-----		-----F2B----					
EcTI	Enterolobi	1	ELLDSGDILSN	GTYIILPAG	GKGGG	ELAASGDET	CPLNVVQARSETKSG	KPAIIWT	PPSVAT			
1TIE	Erythrina	1	VLLDGNGEVVQNG	GTYIILPQ	VWAQGGG	QLAKTGEET	CPLTVVQSPNELSDG	KPIRIES	RLRSAF			
1BA7	Soybean Tr	1	DFVLDNEGNPLENG	GTYIILSDI	TAFGG	RAAPTGNER	CPLTVVQSRNELDKG	IGTIISS	PYRIRF			
2WBC	Winged Bea	1	DDDLVDAEAGNLVENG	GTYIILPHI	IAWHGGG	ETAKTGNEP	CPLTVVRSPNEVSKG	EPIRISS	QFLSLF			
1WBA	Winged Bea	1	ADDPVYDAEAGNKLVNR	GKYTIVSF	SD	GAC	DVVATGNENPEDPLS	IVKSTRNIMYA	TSISSED	KTPPQPRN		
1AVA	Barley ALP	1	ADPPPVDHTDGHELRA	DANYVLSA	NRAHGGG	IMAPGHGRH	CPLFVSQDPNGQHDG	FPVRITPYGVA	PSDKI			
			---F3B---		-----HP2-----		---F4B---					
EcTI	Enterolobi	59	LTPAF	FYLNIEFAGA	ALAA	CLGGYSRLP	KVE	SEVKIAPS	ADESASGG	S	FKIKPYKD	
1TIE	Erythrina	59	IPDD	DKVRIGFAYA	PK	C	APS PW	TVVE	GLSVKLS	EDESTQFD	Y	PFKEQVSD
1BA7	Soybean Tr	59	IAEGH	PLSLKFDS	FAVIMLC	VGI	PTE	SVVEDLPE	GPAVKIG	ENKDAMDG	W	FRLERVSD
2WBC	Winged Bea	59	IPRG	SLVALGFANP	PS	C	AAS PW	TVVD	S	PQGPAVKLS	QQKLPEKD	ILVFKFEKVSH
1WBA	Winged Bea	59	ILENM	RLKINFATD	P		HKGDV	SVVD	FQPDGQQLKLA	GRYPNQV	KGAFTIQKGSN	
1AVA	Barley ALP	59	IRLST	DVRI SFRA	YTT	CL	QS	TE	NHIDSEL	AAGR	RVITGPVKDPSPSGR	ENAFRIEKYSG
			---F5B---		-----HP3-----		---F6B---					
EcTI	Enterolobi	121		DYKLIVYC	CSD	EGISIDDQ	NNRN	IVVK	DGDP	LAVRFVA		
1TIE	Erythrina	121	QLH	SYKLLYCEGKHE	KCAS	EGINRDQK	GYRR	IVVT	EDYP	LTVVLLKKDESS		
1BA7	Soybean TR	121	DEFN	NYKLVFC	PQQAEDDKCGD	EGISIDHDDGTRRLVVS	KNKP	LVVQFQKLDKESL				
2WBC	Winged Bea	121	SNIH	VYKLLYCHDEEDVKCDQY		EGIHRDRN	GNNR	IVVT	EENP	LELVLLAKSETASSH		
1WBA	Winged Bea	121	TP	RTYKLLFC	PVGSP	CKN	EGISTDPEGKKR	LIVVS	YQSDP	LVVKKFRRHEPE		
1AVA	Barley ALP	121	AEV	HEYKLMSCGD	WCQD	EGVFRDLKGGAWFLGAT	EPYH	VVV	FKKAPPA			

3.5.1 O Barril

Barris- β diferentemente de folhas- β abertas possuem cisalhamento, ou seja, existe um grau de deslocamento relativo entre fitas adjacentes na folha, cuja consequência é a inclinação das fitas em relação ao eixo do barril. O número de cisalhamento (S) é o deslocamento total ao longo de uma fita (em resíduos) após completar uma volta no barril na direção das ligações de hidrogênio. Num barril ideal formado por folhas- β , o cisalhamento e o seu raio são dados por:

$$S = \frac{(nb \tan \alpha)}{a} \quad (16)$$

$$R = \frac{b}{[2\sin(\pi/n)\cos \alpha]} \quad (17)$$

onde a é a distância entre $C\alpha$'s ao longo das fitas, b é a distância perpendicular entre fitas vizinhas, n é o número de fitas que formam o barril, S é o número de cisalhamento da folha, α é o ângulo médio entre a direção das fitas e o eixo do barril e R é o raio médio do barril (Murzin *et al.*, 1992, McLachlan, 1979). Para EcTI, como para os demais inibidores Kunitz $n = 6$ e $S = 12$ que leva a valores de $\alpha = 56,31^\circ$ e $R = 7,93 \text{ \AA}$, supondo $a = 3,3 \text{ \AA}$ e $b = 4,4 \text{ \AA}$ (Lesk *et al.*, 1989).

Em suma, todos os barris de inibidores tipo Kunitz apresentam geometrias similares e já descritas (Murzin *et al.*, 1992). Porém um aspecto do barril ainda não comentado e curioso é a natureza do padrão de ligações

de hidrogênio. Há mais de uma maneira de gerar um barril de 6 fitas, com cisalhamento 12, uma vez que o deslocamento (em resíduos) de uma fita para a próxima não precisa ser constante, apesar da necessidade de apresentar uma média de dois deslocamentos.

Observamos que as ligações de hidrogênio ocorrem em maior número entre fitas de motivos diferentes e em menor número entre fitas de um mesmo motivo. As ligações de hidrogênio entre as fitas que formam o barril de um mesmo motivo *trefoil* são em média 2,33 ligações (4, 2, 3 para os motivos 1, 2 e 3) e entre motivos diferentes 5,67 (5,7,5 entre motivos 1 e 2, 2 e 3, 3 e 1 respectivamente).

Esta diferença no número de ligações de hidrogênio é consequência do maior deslocamento entre fitas do mesmo motivo do que entre fitas de motivos diferentes. No segundo caso, a presença de uma quantidade maior de ligações de hidrogênio, além de uma conexão relativamente curta, leva à formação de grampos- β bem estabilizados entre fitas consecutivas, entretanto, provenientes de motivos diferentes. O alinhamento na Tabela 9 (pág.: 73) mostra que as conexões entre F2B e F3B e entre F4B e F5B estão entre os mais conservados, sendo a primeira conexão, o sítio reativo. Esta conservação leva a uma assimetria na variabilidade estrutural de um inibidor para outro como já comentado e mostrado na Figura 22 onde o lado da molécula que inclui o sítio reativo é mais conservado estruturalmente.

Este padrão de ligações de hidrogênio está coerente com a hipótese do enovelamento β -*trefoil* ter surgido a partir de uma triplicação gênica de um gene que codificava um motivo *trefoil* (McLachlan, 1979). Neste caso, supõe-se a necessidade de um número maior de interações entre fitas de motivos diferentes, para fortalecer a estrutura trimérica que teria existido antes da triplicação proposta.

Os resíduos no interior do barril formam um arranjo regular com três camadas compostas pelas cadeias laterais provenientes de cada uma das fitas. A camada superior, formada pelos resíduos Gly14, Thr60, Phe71, Pro122, e Ala170, aponta para o solvente. As outras duas camadas, a camada do meio sendo formada pelos resíduos Tyr16, Ile58, Leu73, Ile120, Tyr127, Phe168 e a camada inferior formada pelos resíduos Ile18, Ala56, Ile75, Phe118, Leu129 e Val166, são formadas por resíduos grandes ou de tamanho médio. Em inibidores Kunitz de uma forma geral a camada do meio e a camada inferior são formadas por resíduos hidrofóbicos.

Figura 25: Vistas dos resíduos que compõem as camadas do barril do EcTI. a) Vista lateral, em verde: resíduos que compõem a camada mais baixa, em laranja: os resíduos que compõem a camada do meio e em magenta: os resíduos compondo a camada superior. b) Vista superior (ao longo do pseudo eixo de ordem três) com densidade eletrônica do mapa $2F_o - F_c$ final.

A predominância de resíduos com cadeias laterais volumosas é esperada para um barril de $n=6$ e $S=12$, pois o raio do barril é relativamente grande quando comparado, por exemplo, com um barril tipo TIM de oito fitas. Murzin *et al.* (1992) descreve que o núcleo hidrofóbico do enovelamento, que é preservado entre todas as estruturas tipo β -*trefoil* é composto por nove aminoácidos hidrofóbicos críticos. O *core* se situa na interface entre o barril e a “tampa” formada pelos grampos de cabelo. Três dos nove resíduos vem da camada mais baixa do barril (das fitas F2B, F4B e F6B), três vem das fitas externas e três das fitas internas dos grampos de cabelo. Como é de se esperar, estas nove posições em EcTI são ocupadas por resíduos hidrofóbicos e estão destacados na Tabela 9 (pág.: 73).

3.5.2 Os Grampos de Cabelo

Os três grampos- β não estão apenas torcidos, como em folhas- β de uma forma geral, mas estão também espiralados, ou seja, o eixo que passa pelo “centro” de cada fita é helicoidal de mão direita. Desta forma, as folhas- β dos grampos de cabelo formam um *coiled-coil*, ou super-hélice (Figura 26).

Figura 26: As duas fitas de cada grampo de cabelo se espiralam uma em torno da outra formando o *coiled-coil*. Na figura vemos o “espiralamento” entre duas fitas do grampo- β do segundo motivo.

Este “espiralamento” permite que as fitas internas, próximas ao pseudo eixo de ordem três, formem ligações de hidrogênio entre si (Figura 27) aproximando as cadeias laterais hidrofóbicas para formarem o núcleo hidrofóbico descrito acima e mostrado na Figura 30. Os círculos na Figura 27 representam os resíduos marcados com as mesmas cores no alinhamento da Tabela 9 (pág.: 73).

Figura 27: Esquema mostrando o arranjo dos grampos de cabelo. A torção dos grampos de cabelo permite às fitas internas dos três grampos de cabelo realizarem ligações de hidrogênio entre si. Os resíduos internos e externos dos grampos- β estão indicados na Tabela 9 com as mesmas cores.

3.5.3 Aspectos Inéditos do Enovelamento

Mesmo o enovelamento β -*trefoil* já ter sido analisado e descrito em profundidade e extensão (Murzin *et al.*, 1992, McLachlan, 1979), entretanto, serão detalhadas algumas observações inéditas.

3.5.3.1 Os Grampos e Seu Espiralamento

Para fitas- β ideais, pode-se obter o grau de torção de uma fita, θ , da seguinte relação envolvendo os ângulos ϕ e ψ

$$\cos\left(\frac{\theta}{2}\right) = -0,817 \sin\left(\frac{\phi + \psi}{2}\right) + 0,045 \sin\left(\frac{\phi - \psi}{2}\right) \quad (18)$$

e o número de resíduos esperados por volta nas fitas- β sendo dado por:

$$n = \frac{360}{\theta} \quad (19)$$

com $180^\circ \geq \theta \geq -180^\circ$.

Por exemplo, um valor de $n = -2,3$ representa 2,3 resíduos por volta, com cada resíduo girando -156° em torno do eixo da fita de mão esquerda, de eixo reto. Assim, a quiralidade das fitas que compõem cada grampo de cabelo pode ser calculada ao longo de cada fita e quantificado em termos de n , a partir dos ângulos torcionais ϕ e ψ . Para uma fita- β reta o parâmetro n deve ser aproximadamente constante.

Figura 28: O número de resíduos por volta (n), calculado a partir dos ângulos (ϕ, ψ) dos resíduos de cada fita dos três grampos de cabelo.

Diferentemente de uma fita- β reta, num *coiled-coil* ideal de duas fitas a combinação de ângulos (ϕ, ψ) são diferentes para resíduos pares e resíduos ímpares e se repetem sistematicamente ao longo da fita, de modo que o eixo da fita é espiralado (Salemme & Weatherford, 1981a e b). Este fenômeno também é observado, por exemplo, em colágeno onde os ângulos (ϕ, ψ) se repetem de três em três resíduos.

Apesar de apresentar a forma de um *coiled coil* e tendo sido descrito assim, não foi observado este fenômeno no caso dos grampos- β de EcTI.

Em termos do parâmetro n observamos um valor de $n \approx -2,3$ no centro de cada fita e valores cada vez mais elevados nas extremidades (Figura 28). Este padrão leva a uma distorção no eixo de cada fita imitando o *coiled coil* para um trecho curto de fita. Talvez não seja observada uma repetição dos ângulos ϕ e ψ (e por consequência n) devido às fitas serem relativamente curtas.

A Figura 29 mostra uma tendência para os resíduos nestas fitas espiraladas de fugirem dos valores ideais de uma fita de eixo reto ($\phi \approx -120$, $\psi \approx +140$, $n = -2,3$). Esses resíduos tendem a ocuparem a região P (poliprolina) do diagrama de Ramachandran (Wilmot & Thornton, 1990) o que indica uma estrutura incomum. Na Figura 29 os ângulos ϕ e ψ dos resíduos das fitas dos grampos de cabelo são comparados com os ângulos das fitas do barril de EcTI, destacando a diferença na distribuição.

Finalizando, as fitas dos grampos de cabelo em EcTI não são fitas retas e nem apresentam o espiralamento esperado de um *coiled coil*. São estruturas distorcidas onde o parâmetro n é próximo ao valor clássico de $-2,3$ apenas no centro, em função de uma distribuição atípica de ângulos ao longo da fita. O resultado é uma estrutura característica que apenas superficialmente aparenta um *coiled coil*.

Figura 29: Distribuição dos ângulos ϕ e ψ dos resíduos das fitas dos grampes de cabelo e fitas do barril num diagrama de Ramachandran. Os círculos azuis indicam os ângulos ϕ e ψ dos resíduos pertencentes às fitas do barril e os triângulos vermelhos representam os ângulos ϕ e ψ dos resíduos das fitas dos grampes de cabelo. Esses resíduos tendem a ocuparem a região P (poliprolina), aproximadamente a região do círculo, diferentemente dos resíduos presentes nas fitas do barril. A reta $\psi = -\phi$ representa aproximadamente a região para quiralidade $n = -2$.

Tabela 10: Alinhamento BLAST contendo inibidores Kunitz. Marcados em amarelo estão os resíduos hidrofóbicos das fitas do barril, marcados em azul estão os resíduos internos do grampo de cabelo e em verde os resíduos externos do grampo de cabelo, em marrom está marcada a Gly53, em vermelho resíduos hidrofóbicos críticos que fazem parte dos grampos. Os resíduos hidrofóbicos das fitas do barril estão marcados em amarelo, as Gly's conservadas no início da primeira fita do grampo de cabelo 1 e do grampo de cabelo 3 estão em vermelho, o Trp conservado no início da primeira fita do grampo de cabelo 2 esta marcado também em vermelho. 1EYL = inibidor de quimotripsina.

		10	20	30	40	50	60	70	80	90	100	110	120	
consensus	1	PVLDTDGNELRNGG	SYLLPAIRGAGG	LEAAKTGNETCPLT	VQSPNELD	NGLFVRISS	PLRSPFIPVVRVS	TLNLSFAAAATDCACVPT	---VWKVVEESQAGGQ	---LTK				
ECTI (propos)	2	ELLDSDDGDIRLNGG	TYLLPAIRKGKGG	LELAKTGDGECPLN	VQARGETK	RGRLPFISS	PPRIAILTP	---AFYLNIEFQTKLPA	LCRLREYSRLP	PEEEQHSE	---VKL		108	
1BA7_A	3	ELVNDNGPLENGG	TYLLSDITAFGG	LRAPTGTNERCLPT	VQSPNELD	RGIGTIISS	PRDIFAEGLHSLK	DSFAVIMLCVGLPT	---PWSVEDLPE	GP-AKKI			107	
1AVA_C	5	PVHDDTDGHELADANN	YVLSANRAHGg	LTMAPGHCGRHCLPT	VQDPNGQH	DCFFVVRIT	PYGVAPSDKIIRLS	TDVIRISFRAYT	---TCLQSt	---EWHIDSELAAGR	RR-HWIT		111	
1EYL_A	6	DLVDAEGNLVNGG	TYLLPHIWAHGG	LEATKTGNEPCPLT	VRSNPVEVS	RGEPIRISS	QFLSLFIPRG	---SLVALGIANPPSCAASP	---NMTTVVDSQG	---P-AKKL			103	
1TIE	1	ELVNDGEVVGNGG	TYLLPQWAAQGG	VLQAKTGEETCPLT	VQSPNELS	DKGPIRIE	SLRASFIPGD	---DKVRIGFAYAPKCAASP	---NMTVVEDEQE	GL-SWKL			103	
1WBA	4	PVYDAENGLNVNRKY	TYVFSDDGAGTD	LVATNGENPEDPLS	VKSTRNIM	---YATSISSD	KTPPOQRN	ILNMYRLINFATDPHKGD	---VWSVVDQ	QDPQGG	---LKL		106	
gi 134841	42	PVLNDINGDEVRRAGN	YVMSAITWAGG	g1RLAHLDMMSK	---CATDVIVSPNLD	NGDPIIT	IPATADPESTVILAS	TYQTFRNIAITNKLGVnn	---vNMQIGQ	---HDSAsGQyFKA			152	
gi 20141344	47	EVDYDQGNPLRIGSG	TYINNPLLGAGAV	---LYLNVN	---LQCPNA	LOHMSIFqfpg	RGPTFVV	VVRKSESDDYGVVRVMVT	---VYIKFFVTKITL	CDVqt	---VWKVNDE	---Q-LVVT	152	
gi 266323	4	PVVDTEGHSLRAGE	SYVYLPSAPSGHG	LTMAPRNVIPCLLL	QAQTDERR	---KCFPVIR	TVGGGAASPTLR	LVRS	TDVIRISFRAYT	---LCVQSt	---EWHVGD	DEPLTGAR	---RWT	110
gi 416662	30	PVLDTDGDDELQTGV	YVYVLLSSISGAGG	glalCRATGQS	---CPEIIVQRRSDLD	NGTFVIFISN	---ADSKDDVVRVS	TDVNI	IEFVPIRDLRCSTs	---tvWRLDN	DYNSAGKwWIT		138	
gi 6166552	35	PVLDTIDGKELRTGTY	NYIVFVRLDRLGL	---tvsATPNTGTFV	CPVCRVYTKRVED	---HGRPLAFIP	---ENPKEDVVRVS	TDNLQINISAFM	---PCRWtsst	WRRLDKYDSE	STGQyFT		145	
gi 124012	39	PVLNDINGENKLNRG	SYYSIRISIGRGALGG	dvlyLKSPNSAD	---CPDGFVYNSDVGp	---SDRPLRIFIP	---LSGGIFED	DLNLTQNFITVKLVCSy	---tHKVGNL	NLNAyfrTmLET			147	
gi 125722	27	PVLDTDDDLQVNGG	TYVYMLPVRMRKGKGG	LEVDSTGKEICPLT	VQSPNELD	RGIGTLFV	SLPHALFIAERYPLSIK	GSFAVITLCAGMPT	---DPAIVERE	---GL-QAVK			131	
gi 124940	2	ELVDVEGKTQNGG	TYVYLVPLRGGGG	MEAAKVGNEDCPLT	VKSLDENS	NGEPIRIAS	RLRSTFTIPEY	---SLNVLG	FADDPKPCASP	---PWTVVDK	GSerLP	---SKKL	105	
gi 124938	2	PLLDSEGEVLVRNG	TYVYLLPDRWALGG	LEAAATGTETCPLT	VRSNPVEVS	---VGEPLRISS	QLRSGFTIPDY	---SVVIRIGIANPPKCAASP	---NMTTVVEDQ	QGP	---SWKL		104	
gi 462388	1	PVLDTIDGELNVNGG	TYVYLLPHIWAHGG	LEAARTGKETCPLT	VQSPFEVS	---NGEPIRIAS	QFLSTFIPDG	---SPYAI	GIANPPSCAASP	---NMTTVVETSEG	---L-AKKL		103	
gi 462427	2	ELLDDAGDIIRLNG	YAYLLPAIRKGKGG	LTALKTGDESCPLT	VQAQSETK	---RGPLAVIT	TPPKIAILTPG	---PYLNIEFQ	QPRDLPA	CLQKyst	LRWKVGESE	---KKI	106	
		130	140	150	160	170	180	190	200					
consensus	109	GGVKG	---PEREDGNI	FKIEKVS	---DYVGYKL	LYCPSGEE	---DDCKCKD	GINID	---DKGYRR	VJVSDE	---NPLVVVFVK		176	
ECTI (propos)	106	APKE	---EAAAFGX	EKLKPYRD	---D	---YKIVYCEGGSD	---DDSCKD	DLGISID	---DENNRRL	VVKDGD	---DPLAVRFVK		169	
1BA7_A	108	GENKD	---AM	---DGMFLER	VSDsd	---EFNNYKL	LVFCPQAAE	---DDCKGD	DLGISIDhd	GGTRR	---VLSVSKN	---KPLVVQFQK	177	
1AVA_C	112	GPVKDpsp	SGrEn	AFRIEKYSga	---EVEHYKL	LMSCGD	---WQCDQ	Q	VFrd	LKGGA	WPLGATE	---PYHVVFQK	175	
1EYL_A	108	SQQL	---Pekd	ILVLFK	FEKVSHs	---NIHVYKL	LYQHDEED	---VKCDQY	IGIHRD	---RNGNRRL	VVTTE	---NPLELLVLLK	177	
1TIE	104	SEDES	---Tqf	---DYPFK	FEKVS	---HQKASD	---GHYSYKL	LYKCEGK	---HEKASG	GINRD	---QKGYRR	LVVDE	---YPLTVLVLK	168
1WBA	107	AGRYP	---Nqv	KGAF	TIQKGSN	---TPRTYK	LYLPCPVG	---SPCKNI	GISID	---PEGKKR	LVVSYQs	DPVLVVFKHR	171	
gi 134841	153	GEFVSd	---Nsn	QFKIEL	VDNALn	---sYKLTQ	QCFG	---SDCKYN	GRFHDHMLR	TRLALS	---SPFVFIK		176	
gi 20141344	153	Gf	---kVGen	EDIFIK	IMKTDLP	vtgpgSKY	LYLHCPSH	---LGCKNY	GNFNKngYPR	LVVDDK	---DFIPFVIK		221	
gi 266323	111	GLPIGpsp	SGrEn	AFRVEKYGG	---G	---CYKL	LVSCRD	---SCQDQ	Q	VSrd	GAGAAW	TG-ASQ	---PPHVVFVK	171
gi 416662	39	DGVKGpsp	PntLCSWFKIEKAGVl	---G	---GKYRFCPSVCDs	---cTTLCS	DLGRHSDd	QGI	---RLASD	---NEWAMWPK			178	
gi 6166552	146	GGVKGpsp	PTISISWFKIEFVSHs	---G	---GKYLVFCPTVCDS	---cKVKCGD	GIYID	---QKGRRL	RLSD	---KPPAEFNE			216	
gi 124012	148	GTIGI	---QadSSYF	KIVKLSnf	---G	---GNLYL	XCPITPPf	lcpferd	NFCAKVGVIQNGKR	---RLALNE	---NPLDVLFE		Q219	
gi 125722	132	LAARD	---TV	---DGMFL	NIERSRe	---YND	---YKLVFCPQAAE	---DNKCED	DLGIQID	---DGIIRR	VLSKN	---KPLVVQFQK	197	
gi 124940	106	GEYKD	---SEL	---DYPFK	FERVYaa	---sKMYAYKL	LYCSEDEE	---eEMMCKD	IGVYRD	---QEGYQR	VVSKH	---NPLVVGFVK	176	
gi 124938	105	SELKS	---TKf	---DYLFK	FEKVTs	---KFSSYK	LYKCAK	---DRTCKD	IGIYRD	---QKGYER	VDTE	---NPLVVFIK	168	
gi 462388	104	LEHKT	---PEed	TKTFK	FQKVSsp	---NRYYN	LYNSYCQREDDd	---LKCDQY	IGIRRD	---AKGYRR	LVVNT	---NPLELLVLV	173	
gi 462427	107	APKEK	---Eofl	VGSF	KIKPYRD	---DYK	---LVYCESDd	---ESCKD	DLGISID	---DENNRRL	VVKDG	---HPLAVRFK	171	

3.5.4 Os Arcos do Motivo *Trefoil*

O motivo *trefoil* apresenta um aspecto estrutural incomum no qual a primeira fita de cada grampo de cabelo é obrigada a passar por baixo da conexão entre a segunda fita do grampo e a segunda fita do barril, como mostrado na Figura 30a. Chamaremos esta conexão de “arco” do motivo de agora em diante. Foram examinados os três arcos na estrutura do EcTI e comparados com as demais estruturas de enovelamento tipo β -*trefoil*.

3.5.4.1 Motivo1

Foi observado um padrão nos inibidores Kunitz de vegetais (Tabela 10) de Gly's conservadas no início da primeira fita do grampo de cabelo 1 (Figura 30). Em EcTI Gly26, Gly27 e Gly28 formam o início desta fita que passa por baixo do arco. Neste motivo, o arco é curto, formado pelo resíduo Ser48-Ser52 passando “apertado” por cima da fita. Resíduos com cadeias laterais maiores do que Gly na fita entrariam em choque estérico com os resíduos que formam o arco, justificando a conservação da Gly's. Há também a necessidade de uma glicina (Gly53) no início da segunda fita do barril, logo após o arco, também por razões de espaço. A Tabela 10 mostra a conservação deste resíduo.

Os resíduos Gly27 e Gly28 participam de uma saliência- β juntamente com Ala46, contribuindo para a curvatura da primeira fita assim que passa por baixo do arco (Figura 30b). A conservação das Gly's parece ser um

aspecto peculiar dos inibidores tipo Kunitz pois as demais estruturas tipo β -*trefoil* (FGF, interleucinas, ricinas, etc.) estudadas por Ponting & Russel, (2000) são mais variáveis nestas posições em função de possuírem um arco mais elevado. Porém, mesmo assim, nestas estruturas a primeira das três glicinas (Gly26 em EcTI) geralmente se conserva (Figura 31). Esta conservação ocorre não apenas no primeiro motivo, mas nos três motivos de cada estrutura, sendo característica do motivo β -*trefoil*. Nestes casos, porém, a conservação da glicina tem outra explicação, pois ela adota uma conformação α_L ($\phi \approx 50^\circ$, $\psi \approx 50^\circ$) e termina numa pequena hélice- α , ou volta α , na conexão entre a primeira fita do barril e a primeira fita do grampo (Figura 23, pág.:71). Hélices- α são freqüentemente terminadas por Gly's ou Asn's na conformação α_L (Nagarajaram *et al.*, 1993, Richardson & Richardson, 1988) justificando sua conservação.

Figura 30: a) Localização relativa do arco no motivo *trefoil* b) Vista em estéreo do "Arco" sobre a Gly27 do grupo de cabelo 1. Ligações de hidrogênio da saliência- β envolvendo Gly27, Gly28 e Gln45 estão indicadas. O arco passa por cima na figura e a fita passa pelo centro.

3.5.4.2 Motivo 2

No segundo motivo de EcTI o arco é maior (resíduos Pro107-Gly115) e passa mais alto sobre a fita- β eliminando a necessidade de Gly's na entrada da fita (Figura 32). A Pro87, no início da fita preenche a cavidade entre a fita e o arco. No inibidor de *E.caffra* esta posição é ocupada por um Trp cuja cadeia lateral aponta para baixo do arco na direção do sítio de interação com a serinoprotease. Em outras proteínas tipo β -*trefoil* um resíduo aromático (Figura 31) nesta posição é preferido, fazendo sempre contatos de van der Waals com o arco, não somente no segundo motivo, mas nos três.

Figura 31: Alinhamento múltiplo de seqüências de domínios β -trefoil mostrando as três repetições internas (Pointing & Russel, 2000) com cada repetição separada por filas. A seta indica a glicina discutida no texto, e o asterisco o resíduo aromático.

Esta interação, portanto, parece canônica para a maioria dos arcos encontrados em β -trefoils.

Figura 32 : Vista estéreo do "arco" do segundo grampo de cabelo. A cadeia lateral da Pro93 ocupa o vão entre a fita e o arco.

3.5.4.3 Motivo 3

No terceiro motivo de EcTI o arco também passa mais alto sobre a fita- β de entrada e é composto pelos resíduos Val158 até Pro163. Uma Gly na posição 145 é estritamente conservado nos inibidores tipo Kunitz para evitar um impedimento estérico com a carbonila de Asp162 do arco (Figura 33).

Concluindo, o EcTI utiliza uma série de mecanismos estruturais para resolver a dificuldade de passar o arco por cima da fita- β de entrada ao grampo, uma necessidade do enovelamento do motivo β -*trefoil* e portanto para o β -*trefoil* como um todo. Estes mecanismos incluem o uso de glicinas em posições estratégicas, variações no comprimento do arco e o empacotamento de um resíduo aromático proveniente da fita de entrada com o lado inferior do arco.

Figura 33: Vista estéreo do "arco" do terceiro motivo. Este *loop* é o mais longo dos três grampos de cabelo. A Gly 145 (em laranja) é conservada por razões de impedimentos estéricos que seriam gerados por aminoácidos com cadeias laterais maiores.

3.5.5 Águas

Na estrutura final foram encontradas 147 moléculas de água cada qual atribuída aos picos maiores que 3σ no mapa $F_{\text{obs}} - F_{\text{calc}}$. Algumas destas moléculas de água estão localizadas em locais “estratégicos”, onde desempenham papel de mediação entre algumas interações importantes. Como é o caso da W3 (está quase sobre o pseudo-eixo de ordem três molecular) que interage com as carbonilas dos resíduos Asn42, Val103 e Leu156. Cada um destes resíduos vem de um dos motivos *trefoil* sugerindo que esta molécula de água faça parte integral da estrutura. Uma água na mesma posição é conservada em ETI (PDB 1TIE) (Onesti *et al.*, 1991) e em EcTI é uma das águas mais bem estruturadas, sendo o segundo maior pico no mapa diferença e refinada até um fator de temperatura de $26 \text{ \AA}^2$. Várias outras águas mais bem orientadas (W4, W8, W11 e W12) com fatores de temperatura comparáveis aos dos átomos da cadeia polipeptídica, são relativamente enterradas e parecem desempenhar papéis estruturais. Nenhuma delas parece ser importante para a manutenção da estrutura do *loop* reativo.

3.5.6 Conformação do *Loop* Reativo

O sítio de clivagem (sítio reativo), se localiza entre os resíduos Arg63-Val64. Em inibidores de proteínas de um modo geral a região entre P3 e P3' apresenta uma estrutura comum conhecida como *loop* canônico (Osawa, K., Laskowski, 1966) e se localiza no *loop* que conecta o primeiro motivo com o segundo motivo. A sequência de aminoácidos de P3 a P3' do *loop* reativo de EcTI é semelhante, mas não idêntico àquela encontrada no inibidor de purificado de *E.caffra* ETI (1TIE) e SBTI (1BA7) como se observa na Tabela 11 abaixo, onde as linhas verticais duplas representam a ligação peptídica que será clivada. Uma prolina é presente na posição P3 em SBTI e também é observada em dois inibidores de *Bauhinia bauhinoides* e *Bauhinia ruja* (Oliva, 2003). Por outro lado uma Pro na posição P2 é encontrada em inibidores de *Bauhinia unguolata* e *Bauhinia variegata* (Oliva, 2003). Prolinas nas posições P3 e P2 já foram relatadas na literatura em inibidores de *Adenanthera pavonina* (Negreiros *et al.*, 1991) e *Prosopis juliflora* (Richardson, 1991), entretanto, EcTI é a primeira estrutura cristalográfica de um inibidor deste tipo com prolinas em ambas as posições. Não há incompatibilidade entre o ângulo ϕ da prolina ($\phi \approx -65^\circ$) e a conformação canônica do *loop*, mesmo P3 estando no limite do normalmente encontrado (Apostoluk & Otlewski, 1998). Porém a presença de uma prolina na posição P3 limita a formação de ligações de hidrogênio com o sítio estendido da serinoprotease.

Tabela 11: Alinhamento seqüencial dos *loops* reativos entre três inibidores de tripsina da família Kunitz de estrutura conhecida. Código de cores usado (verde) aminoácidos hidrofóbicos, (vermelho) aminoácidos carregados, (cian) aminoácidos polares. 1TIE inibidor de E.caffra, 1BA7 inibidor de soja.

	P3	P2	P1	P1'	P2'	P3'
EcTI	Pro	Pro	Arg	Val	Ala	Thr
1TIE	Arg	Leu	Arg	Ser	Ala	Phe
1BA7	Pro	Tyr	Arg	Ile	Arg	Phe

Figura 34: *Loop* do sítio reativo de EcTI com o mapa de densidade eletrônica mostrando que sua estrutura é bem definida. Embora a qualidade geral do mapa de Fourier no *loop* de ligação seja boa, pôde-se apenas observar densidade eletrônica parcial para a cadeia lateral da Arg63, sendo portanto modelada como Ser.

Apesar da ausência de pontes de dissulfeto ou interações eletrostáticas fortes, o sítio reativo é mantido numa conformação ordenada por ligações de hidrogênio a Asn12 e os resíduos Thr60, Pro62, Val64 e Ala65, formando uma “malha” de ligações que mantém o *loop* em um estado de baixa flexibilidade (Tabela 12) e (Figura 35), verificado inclusive com relação aos fatores de temperatura.

Figura 35: “Malha” de ligações de hidrogênio realizadas pela Asn12 estabilizando o *loop* reativo. A Arg63 na posição P1 do EcTI foi modelada como Ser.

Tendo como critério a classificação do programa PROMOTIF (Hutchinson & Thornton, 1990a) (Figura 36) para a identificação de elementos de estrutura secundária, foi averigüado o fator de temperatura isotrópico médio por resíduo através do programa *B_average* (CCP4, 1994) para toda a estrutura. Os valores para os resíduos pertencentes aos *loops* podem ser vistos na Tabela 12.

Figura 36: Diagrama HERA (Hutchinson & Thornton, 1990b) das estruturas secundárias encontradas em EcTI. Resíduos encontrados em regiões de fita-β são mostrados em retângulos e as regiões de hélice em elipses.

Tabela 12: Fator de temperatura médio por resíduo encontrado nos *loops* de EcTI. MC_Bav = Fator de temperatura médio dos átomos da cadeia principal, SC_Bav = Fator de temperatura médio dos átomos da cadeia lateral, Bav = Fator de temperatura médio para a cadeia peptídica como um todo. Todos os valores estão em $\text{\AA}^2$.

Loop	MC_Bav	SC_Bav	Bav
22-27	33,18	43,90	33,96
33-41	27,11	27,46	27,34
47-54	25,16	29,21	26,89
60-73	28,35	29,01	28,86
88-93	25,33	25,88	26,12
112-116	32,62	20,44	32,86
124-125	30,90	37,35	34,25
150-152	25,67	29,73	27,93
160-163	22,85	20,05	24,20

O fator de temperatura médio para o EcTI foi de $25,64 \text{ \AA}^2$ para os átomos da cadeia principal, $26,72 \text{ \AA}^2$ para os átomos da cadeia lateral e $26,14 \text{ \AA}^2$ considerando-se todos os 1.170 átomos da cadeia. Os resíduos do

loop reativo fazem parte de um *loop* maior que se estende do resíduo Thr60 ao resíduo Leu73. Este *loop* apresenta fatores de temperatura para a cadeia principal ($28,35 \text{ \AA}^2$) comparável ao EcTI como um todo. Este dado, junto com a densidade eletrônica clara e não ambígua demonstra que o *loop* reativo está bem estruturado. Provavelmente a presença de prolinas nas posições P2 e P3 contribua para a sua rigidez. Porém a cadeia lateral da Arg63 não apresentou densidade eletrônica além do átomo γ , sugerindo sua flexibilidade. De uma forma geral estas observações são compatíveis com o conceito dos inibidores possuírem um *loop* pré-formado para interação com proteases. Assim a falta de perda de entropia durante a formação do complexo EI, poderia contribuir para a sua alta afinidade (Hubbard *et al.*, 1991).

De um modo geral, uma forte correlação pode ser observada entre resíduos pertencentes às fitas- β e um valor relativamente baixo para o fator de temperatura (Figura 37). Como pode ser observado, por exemplo, para as fitas do primeiro motivo (Gly14–Ala21, Gly28–Ala32, Asn42–Ala46, Pro55–Trp59). O maior valor observado foi para o resíduo Ala81 ($40,4 \text{ \AA}^2$) que está no *loop* entre as fitas 1 e 2 do segundo motivo. O segundo maior pico observado refere-se aos resíduos Gly23 ($38,9 \text{ \AA}^2$) e Lys25 ($37,6 \text{ \AA}^2$) que fazem parte do *loop* que liga as fitas 1 e 2 do motivo 1. Estes resíduos estão desconectados entre si no modelo elaborado uma vez que, devido à total ausência de densidade eletrônica neste local não foi possível construir nenhuma cadeia para o resíduo Gly24.

Figura 37: Distribuição do fator de temperatura médio para os átomos da cadeia principal. As regiões das fitas estão destacadas com os traços pretos.

As distâncias $C\alpha-C\alpha$ dos segmentos P6 a P6' encontrados na estrutura de EcTI estão dentro da variabilidade apresentada por Apostoluk & Otlewski (1998), com exceção de algumas distâncias que apresentam um valor menor comparativamente aos mínimos apresentados no artigo acima: P2-P3' ($<0,18$), P5-P3' ($<0,13$), P6-P3' ($<0,22$) e P6-P2' ($<0,09$). Valores dos ângulos ϕ e ψ para o *loop* reativo também estão dentro das faixas esperadas (Figura 38), portanto, podemos concluir que há pouco de incomum sobre o *loop* reativo do EcTI.

Figura 38: Faixas dos ângulos ϕ e ψ adotados pela cadeia principal na região P3-P3' para inibidores livres e inibidores complexados com proteinases (barras cinzas), de acordo com Apostoluk & Otlewski (1998). As estrelas marcam os ângulos encontrados no modelo de EcTI.

3.5.7 Pontes de Dissulfeto

As duas pontes de dissulfeto na estrutura apresentaram densidade clara e não ambígua. Ambas apresentam a conformação de uma espiral de mão esquerda ($\chi_3 \approx 90^\circ$). A segunda ponte (entre os resíduos 132 e 141) une a entrada e saída do loop clivado em EcTI que leva à formação de duas cadeias. Do ponto de vista funcional é difícil atribuir qualquer consequência significativa a esta clivagem.

3.6 Modelagem da Seqüência Proposta

Como foram observadas diferenças seqüenciais entre a seqüência primária esperada e a seqüência efetivamente observada no mapa de densidade

eletrônica, foi realizada uma modelagem molecular por homologia para a averiguação das possíveis influências estruturais destas “mutações” e para subsequente modelagem de complexos com diversas serinoproteases.

O molde selecionado para a construção do modelo foi a própria estrutura de EcTI encontrada cristalograficamente e definiu-se como alvo a seqüência inicialmente proposta (Batista *et al.*, 1996). Um alinhamento foi realizado utilizando-se o programa MULTALIGN (Barton, 1990) (Tabela 8), entre a seqüência proposta, que possui 174 resíduos, e a seqüência efetivamente encontrada através do mapa de densidade eletrônica, possuindo 159 resíduos. Não foram incluídos na modelagem os resíduos 132-141 por se tratar de uma região de quebra de cadeia possuindo um C e um N-terminal bastantes “flexíveis”, os resíduos AHRRG do C-terminal da cadeia leve também não foram incluídos na construção do modelo.

O programa MODELLER (Seção 2,5) foi utilizado para a realização da modelagem por homologia. Foram calculados 50 modelos para que se obtivesse uma solução estatisticamente relevante. A (Figura 39) mostra a distribuição de pseudo-energias para os 50 modelos. As 5 melhores soluções passaram por testes de validação. A melhor solução foi o modelo número 4, correspondente ao modelo de menor energia (Tabela 13). Este modelo será chamado de mod1 de agora em diante.

Figura 39: Distribuição das energias dos 50 modelos gerados na modelagem molecular por MODELLER usando a sequência encontrada de EcTI como molde para a sequência proposta. As energias variam de 1184,06 a 1298,93 kcal/mol.

Tabela 13: Cinco melhores modelos em 50 com *G-factor* calculado por PROCHECK a 2,0Å.

Ordem	# modelo	E(kcal/mol)	<i>G-factor</i>
5	50	1.298,933	-0,3
4	3	1.294,348	-0,2
3	36	1.284,831	-0,2
2	14	1.249,489	-0,3
1	4	1.184,064	-0,2

O diagrama de Ramachandran e gráfico de Verify3D estão apresentados na Figura 40 e Figura 41 demonstrando a qualidade estereoquímica e compatibilidade estrutura-sequência do modelo. O diagrama de Ramachandran mostra apenas dois resíduos em regiões generosamente permitidas. Asn152 está próxima à região α_L e a inserção Phe118 (relativamente à sequência proposta) foi provavelmente responsável pela Lys119 adotar ângulos generosamente permitidos ($\phi = -140,6^\circ$ / $\psi = 158,9^\circ$). O perfil Verify-3D demonstra que nenhuma região do modelo passa abaixo de ~0,15, indicando que o modelo não apresenta problemas gerais em termos de ambientes químicos dos resíduos.

Figura 40: Diagrama de Ramachandran global do modelo 1 obtido utilizando-se a estrutura cristalográfica como molde. 87,2% dos resíduos estão nas regiões mais favorecidas.

Figura 41: Perfil da verificação feita pelo VERIFY 3D em mod1 da Tabela 13, indicando que não há incompatibilidade entre seqüência e estrutura.

Porém, como comentado anteriormente, a estrutura modelada da sequência original apresenta Glu117 e Arg95 enterrados no núcleo hidrofóbico da molécula. Arg95 por exemplo está rodeada por Ile19, Leu30, Val44, Ile76 e Val103. Há a possibilidade de se formar uma ponte salina para entre os dois resíduos, inclusive um dos rotâmeros mais comuns para a arginina traz o seu grupo guanidino para próximo do ácido carboxílico do Glu117. Apesar de não ser absolutamente proibido, pontes salinas enterradas são incomuns, e na seqüência fornecida pela interpretação do

mapa de densidade eletrônica, Arg95 e Glu117 são substituídos por Trp e Phe, respectivamente. Como já mencionado, estes dois resíduos se encontram numa região da sequência original onde não houve sobreposição de peptídeos durante o sequenciamento, sugerindo que houve dúvidas na atribuição da sequência correta.

3.7 Modelagem dos Complexos Entre EcTI e Serinoproteases

Numa tentativa de explicar as propriedades do EcTI em termos da sua capacidade de inibir serinoproteases, modelos correspondentes aos complexos foram feitos.

3.7.1 Complexo EcTI/Tripsina

O diagrama de Ramachandran mostrou 99,1% dos resíduos em regiões mais favorecidas ou adicionalmente permitidas. Nenhuma região no modelo apresentou valores negativos no perfil Verify-3D (Figura 42) indicando boa compatibilidade entre sequência e estrutura.

Figura 42: Perfil VERIFY-3D para o complexo tripsina/mod1. O “degrau” aproximadamente no resíduo 230 corresponde ao fim da cadeia da tripsina e início da cadeia do inibidor EcTI. Nenhuma região apresenta valores negativos indicando boa compatibilidade entre seqüência e estrutura.

Um fator de correlação de complementaridade de forma entre as duas proteínas foi calculado utilizando-se o algoritmo de Lawrence e Colman (Lawrence & Colman, 1993). O cálculo é feito a partir de uma função escalar envolvendo os vetores normais da superfície de Connolly de cada molécula (Richards, 1977). Interfaces com complementaridade $S_c=1$ se complementam perfeitamente ao passo que $S_c=0$ indica interfaces não correlacionadas topograficamente. Para o complexo tripsina/mod1 a complementaridade de superfície encontrada foi de $S_c = 0,73$, que está dentro do intervalo esperado para complexos proteína/inibidor (0,70 – 0,76) coerente com o K_i experimental de 0,88 nM. A qualidade do modelo também foi verificada pelo índice de qualidade do programa WHATIF (Vriend, 1990). A qualidade global de -0,87 indica que, de uma forma, geral, os ambientes químicos da estrutura complexada, inclusive a interface, são aceitáveis.

Uma averiguação da manutenção das ligações do *loop* reativo no complexo tripsina/inibidor foi realizada e observou-se que Pro62 (P3) do inibidor mantém sua ligação com a Gly216 da tripsina e a Arg64 (P1) também mantém suas ligações com o Asp189 (bolsão S1) (Tabela 14), podendo ser comparada com as ligações realizadas pelo *loop* reativo de SBTI em complexo com tripsina (PDB 1AVW), usado para a sobreposição.

Próximo ao *loop* reativo ocorre um empilhamento de aromáticos que ficam escondidos na formação do complexo. Participam deste empilhamento os resíduos Trp60 e Phe114 do inibidor e a Tyr217 da tripsina. A interação entre Phe114 (EcTI) e Tyr217 (tripsina) é específica para a tripsina de porco e não deve representar nenhuma contribuição à interação com tripsina bovina, pois Phe114 é substituído por Ser na enzima bovina. A proximidade da Phe114 à enzima é resultado da conformação do arco do segundo motivo. O arco apresenta uma conformação bastante diferente em EcTI quando comparado à SBTI e ETI (de *E.caffra*) levando a novas interações com a protease, eventualmente envolvidas na modulação da especificidade pela enzima alvo. Isto é coerente com as especulações de de Meester *et al.*, (1998), que notaram que diferenças estruturais nesta região poderia explicar a capacidade inibitória de *E.caffra* (ETI) contra ativador tecidual de plasminogênio (tPA), enquanto que SBTI é inativo.

Tabela 14: As ligações de hidrogênio efetuadas pelo *loop* reativo no complexo tripsina/EcTI são mantidas quando comparadas com as ligações realizadas pelo *loop* reativo do inibidor no complexo tripsina porcina/SBTI (PDB 1AVW).

P3	Pro 62 O	2,8	O A216 Gly
P2	Não	faz	Ligações
P1	Arg 64 NH2	2,9	OD2 A189 Asp
	Arg 64 NH2	3,2	O A190 Ser
	Arg 64 NH2	3,1	O A219 Gly
	Arg 64 NH1	2,9	O A215 Trp
	Arg 64 NH1	3,1	O A21 Gly
	Arg 64 O	2,8	N A193 Gly
	Arg 64 O	3,2	N A194 Asp
	Arg 64 O	2,9	N A195 Ser
	Arg 64 O	3,1	OG A195 Ser
P1'	Não	faz	Ligações
P2'	Não	faz	Ligações
P3'	Não	faz	Ligações

3.7.2 Complexo EcTI/Trombina

Medidas de K_i (Batista, 1997) (Tabela 3) mostram que EcTI não inibe trombina, nem fator Xa apesar das mesmas apresentarem especificidade por aminoácidos básicos em P1. Numa tentativa de explicar tais resultados, modelos para os complexos correspondentes foram gerados. Após a minimização de energia, o modelo do complexo apresentou uma complementaridade de superfície de $S_c = 0,71$, dentro dos limites para um complexo enzima inibidor. Porém apesar de uma aparente compatibilidade de forma, outras medidas de qualidade indicaram interações desfavoráveis entre inibidor e enzima. O perfil de Verify-3D (Figura 43) indica uma região negativa em torno do resíduo 120, correspondendo a uma incompatibilidade química na interface. Nota-se que esta mesma região não apresenta problemas quando o modelo sozinho foi avaliado (Figura 41) indicando que a

incompatibilidade é consequência da formação do complexo. O escore da qualidade dada por WHATIF é -1,47, mais uma vez indicando uma distribuição de contatos atômicos inferior em qualidade quando comparado com o complexo com tripsina.

Figura 43: O perfil VERIFY-3D para o complexo trombina/EcTI indica uma região negativa em torno do resíduo 130 do inibidor, correspondendo a uma incompatibilidade química na interface. Esta mesma região não apresenta problemas na mesma avaliação efetuada com o modelo sozinho (Figura 41).

A principal razão estrutural que desfavorece a formação do complexo são impedimentos estéricos gerados em consequência de três inserções em trombina, comparado a tripsina, nas regiões dos resíduos 39, 60 e 149 (numeração de quimotripsinogênio) (destaques (3), (4), e (5) da Figura 45 e Figura 44). A inserção de nove resíduos (60A - 60I) em trombina gera uma série de impedimentos em torno da Pro70 e Pro121 do inibidor, correspondendo este último resíduo à região de mínimo no perfil Verify-3D. A mesma inserção está espacialmente próxima à região N-terminal do EcTI, como será comentado mais adiante. A inserção na região do resíduo 39 da trombina resulta também na sua aproximação com o N-terminal, onde Glu2

(EcTI) e Glu39 (trombina) estão a uma distância de 2,91 Å no modelo minimizado, gerando uma repulsão eletrostática.

Figura 44: Sobreposição das estruturas de trombina (laranja) e tripsina (verde). Em complexo com EcTI. Apenas o inibidor do complexo tripsina/EcTI é mostrado para clareza da figura. Em vermelho estão os *loops* da tripsina e em azul estão as regiões de inserção da trombina. Em rosa está o N-terminal do inibidor.

Uma segunda incompatibilidade de cargas é consequência da substituição da Glu221 (tripsina) por Arg221 (trombina), que leva à repulsão eletrostática com Arg92 do EcTI. Apesar de ser um resíduo de superfície, o rotâmero da Arg221 está fixo por consequência de duas ligações de hidrogênio feitas com Glu146 da própria trombina. A geometria desta interação, que envolve os átomos N_ϵ e N_ξ da arginina e os dois átomos O_γ do ácido glutâmico, é bastante favorável (Singh *et al.*, 1987) sugerindo que a aproximação da Arg221 à Arg92 (destaques (6) Figura 45b e (2) Figura 45a

seja uma consequência inevitável da complexação. A qualidade do resíduo Arg92 medido por WHATIF é -5,4, indicando um ambiente atômico bastante desfavorável.

Apesar disto, de uma forma geral, a complementaridade de cargas entre as superfícies da trombina e EcTI é qualitativamente razoável (Figura 45), com um poço de carga de potencial negativo no bolsão S1 (Figura 45b) cercado por uma região eletrostaticamente neutra, que complementa o inibidor. Visível também na Figura 45, é uma região no EcTI de forte potencial negativo, em função do agrupamento de cinco ácidos glutâmicos (96, 97,98 e 109,110) (destaque (1) na Figura 45a). Esta alta densidade de cargas, que é incomum, desaparece na estrutura sequenciada a partir do mapa de densidade eletrônica em função das substituições Glu96Lys, Glu97Val, Glu109Ala e Glu110Asp realizadas. Isto reduz a carga total da região de -5 para -1. Novamente, a região em escrutínio corresponde à região 91-117 na seqüência original para a qual não havia sobreposições durante o sequenciamento (Figura 45c), sugerindo que o agrupamento de cargas negativas poderia ser um resultado de dúvidas no sequenciamento.

Figura 45: Superfície de potencial eletrostático gerada pelo programa GRASP (Nicholls *et al.*, 1991) em a) e b) ilustrando a complementaridade de carga na interface do complexo EcTI/trombina. As regiões delimitadas em a): (1) trata-se de uma região bastante ácida na sequência original e, como pode ser visto em c), as substituições realizadas nesta região levaram a uma redistribuição de cargas tornando a região não carregada, em (2) a Arg92 após a formação do complexo estará próxima à Arg221 da trombina (destaque b6). As regiões 3, 4 e 5 em b) se referem a *loops* com grandes inserções na trombina. Em c) está destacado a Ser63 (P1) que está interagindo com o Asp149 (S1).

3.7.3 Complexo EcTI/Fator Xa

Foi comprovado que EcTI não inibe fator Xa (Batista, 1997). O modelo deste complexo apresentou alguns aspectos em comum com o caso da trombina. O escore Verify-3D chega a zero para a região em torno da Arg92 (Figura 46). No complexo com fator Xa, pela mesma razão detalhada acima para a trombina, esta Arg está especialmente próxima a Arg221 do fator Xa, gerando repulsão eletrostática. O fator Xa também apresenta uma inserção, relativamente à tripsina, na região do resíduo 60. Devido a uma diferença na conformação e no tamanho deste *loop*, quando comparado a trombina, potenciais impedimentos estéricos envolvem uma região diferente do inibidor. A saber, a saída do *loop* reativo, entre Ile67 e Thr69. A combinação destes fatores provavelmente explica a inatividade do EcTI contra Fator Xa.

Figura 46: Perfil VERIFY-3D do complexo fator Xa/EcTI. A região em torno da Arg92 (inibidor) chega a zero refletindo a incompatibilidade local gerada pela repulsão eletrostática entre a Arg92 e a Arg221(fator Xa).

3.7.4 Complexo EcTI/Quimotripsina

À primeira vista pode parecer contraditório o fato de EcTI também inibir quimotripsina quando a especificidade em P1 desta enzima é notadamente diferente de tripsina, com preferência para aminoácidos aromáticos (Blow *et al.*, 1976). Mesmo assim, vários inibidores são capazes de inibir as duas enzimas (Tabela 3). Uma explicação plausível foi oferecida pela estrutura do complexo entre quimotripsina e BPTI, um inibidor da família Kunitz com uma lisina em P1 (Scheidig *et al.*, 1997). Neste complexo, a cadeia lateral da lisina faz uma “curva” expondo a cadeia alifática ao bolsão S1 e com o grupo amino formando ligações de hidrogênio com carbonilas tanto do inibidor quanto da protease. Após sobrepor EcTI em BPTI utilizando os resíduos do *loop* canônico como referência, foi averiguado que a Arg63 poderia assumir uma conformação equivalente à aquela vista no complexo de BPTI. A Figura 47 mostra a conformação final e as ligações de hidrogênio gerados. Isto poderia representar um mecanismo comum de inibição de quimotripsina por inibidores de tripsina, inclusive de famílias diferentes, uma vez que as interações geradas envolvem apenas resíduos da quimotripsina e do *loop* canônico do inibidor, cuja estrutura é conservada nas diversas famílias.

Figura 47: A Arg63 do EcTI (em azul) poderia assumir uma conformação equivalente à Lys em BPTI (em amarelo).

3.7.5 tPA e Fator XIIa

A observação que EcTI inibe fator XIIa mas não tPA é bastante curiosa, pois, de uma forma geral, inibidores apresentam um comportamento equivalente frente estas duas enzimas. de Meester *et al.*, (1998) atribuíram essa seletividade principalmente ao N-terminal do inibidor. Seqüências bloqueadas, com resíduos adicionais (comparando-se com o inibidor de *E.caffra*) ou com aminoácidos ácidos no N-terminal (Teixeira *at al.* ,1994) não inibem tPA devido a impedimentos estéricos na região do resíduo 60 da protease em função de uma inserção quando comparado com tripsina ou quimotripsina. Segundo este critério, tanto a seqüência original do EcTI (que apresenta uma Lys a mais) quanto a seqüência obtida à partir do mapa de densidade eletrônica (que inicia-se com um Glu) não deveria inibir tPA, como observado experimentalmente. A Tabela 15 mostra que as seqüências de tPA e Fator XIIa são muito parecidas nesta região, que levaria a uma

explicação possível, por dedução, do padrão de inibição das duas enzimas para a maioria dos inibidores do tipo Kunitz.

Tabela 15: Inserção na região do resíduo 60 em tPA e fator Xlla podem estar levando a impedimentos estéricos com o N-terminal do inibidor. Alinhamento da seqüência de tPA, fator Xa e quimotripsina bovina na região do resíduo 60.

	57
tPA	AAHCFQERFPP
fator Xlla	AAHCLQDRPAP
quimotripsina bovina	AAHC GVTTS

A falta de uma estrutura cristalográfica do fator Xlla dificulta uma explicação estrutural de porque EcTI apresenta um comportamento anômalo neste sentido. Seria de grande interessa, do ponto de vista de uma maior compreensão do fenômeno de reconhecimento molecular entre protease e inibidor, a obtenção da estrutura cristalográfica do complexo EcTI e fator Xlla.

Conclusões

Capítulo Quatro

4 Conclusões

Foi possível determinar a estrutura cristalográfica do EcTI a partir de um cristal obtido com a técnica de acupuntura em gel, na condição 15% PEG8000 - 50mM LiSO₄. Nestas condições a proteína cristaliza no grupo espacial P2₁, com parâmetros de rede $a=37,12 \text{ \AA}$, $b=38,42 \text{ \AA}$ e $c=54,08 \text{ \AA}$ e $\beta=98,08^\circ$ possuindo uma molécula por unidade assimétrica.

A molécula de EcTI tem enovelamento tipo β -*trefoil*, típico de inibidores de protease tipo Kunitz, composto por três motivos *trefoil* formando um barril de seis fitas e três grampos- β formando a “tampa” do barril. Os aspectos estruturais canônicos deste enovelamento são conservados em EcTI. O padrão de ligações de hidrogênio encontrado no barril de EcTI e em outros inibidores Kunitz ainda não foi comentado na literatura. Essas ligações de hidrogênio ocorrem em maior número entre fitas de motivos diferentes (média de 5,67 ligações) e em menor número entre fitas de um mesmo motivo (média de 2,33 ligações), coerente com a

proposta de que o β -*trefoil* teria sua origem numa triplicação gênica a partir de um gene que codificava um único motivo β -*trefoil*.

Cada grampo- β assume uma estrutura de *coiled coil* atípica. Onde, ao invés de alternar as coordenadas (ϕ, ψ) sistematicamente num diagrama de Ramachandran para resíduos pares e ímpares ao longo da fita, o parâmetro n associado à cada resíduo tem seu valor mínimo no centro das fitas correspondendo a uma região plana e divergindo nas extremidades para uma região do espaço $\phi\psi$ que corresponde à estrutura tipo poliprolina. Não se sabe quão comum é este motivo em proteínas de modo geral, mas merece investigação.

Um aspecto estrutural incomum do motivo *trefoil* no qual a primeira fita de cada grampo de cabelo é obrigada a passar por baixo da conexão (chamada aqui de “arco”) entre a segunda fita do grampo e a segunda fita do barril em cada um dos três motivos foi examinado. É interessante entender como a proteína “resolve” este problema estereoquímico. A análise dos três arcos em EcTI mostrou que uma série de mecanismos são usados incluindo o uso de glicinas estrategicamente posicionadas, arcos mais longos para evitar impedimentos estéricos e resíduos aromáticos provenientes da fita de entrada empacotados contra o lado inferior do arco. Uma análise das demais proteínas que apresentam um enovelamento β -*trefoil* revela um padrão nos três motivos que não é observado nos inibidores Kunitz, a saber uma seqüência Gly-X-Aro (Aro = aromático) na entrada da fita onde a glicina

geralmente adota ângulos ϕ/ψ de α_L e o aromático empacota do lado inferior do arco.

O *loop* reativo em EcTI se localiza na região que faz a conexão entre o primeiro motivo e o segundo motivo e sua seqüência de aminoácidos é semelhante às seqüências encontradas nas estruturas ETI (PDB 1TIE) e SBTI (1BA7). EcTI é a primeira estrutura cristalográfica de um inibidor deste tipo com prolinas em ambas as posições, P2 e P3, mas isto não compromete a conformação do *loop* canônico. A boa definição do *loop* no mapa de densidade eletrônica e os fatores de temperatura comparáveis com o resto da estrutura são coerentes com o conceito dos inibidores possuírem um *loop* pré-formado para interação com proteases.

Uma incompatibilidade entre a seqüência de aminoácidos determinada quimicamente e o mapa de densidade eletrônica levou à proposta de uma seqüência alternativa para o EcTI. Isto recaiu no problema clássico de como determinar uma seqüência de aminoácidos a partir do mapa. É provável que a nova seqüência não seja completamente correta devido à ambigüidades geradas por cadeias laterais similares (Asp e Asn; Glu e Gln; Val e Thr) e devido à flexibilidade das mesmas em alguns casos. Mesmo assim a seqüência proposta é compatível com os demais inibidores da família Kunitz e é idêntica à seqüência original na sua maior parte, além de levar a um refinamento satisfatório da estrutura cristalina. Ainda não está claro se a amostra cristalizada representa uma isoforma diferente daquela seqüenciada, ou se de fato, a seqüência original apresenta erros. Alguns

resíduos da seqüência original (por exemplo, Arg95 e Glu117) parecem improváveis e vem de uma região que não apresentava *overlaps* no sequenciamento original. No mínimo, nossos resultados sugerem a necessidade de um reexame da seqüência do EcTI e da existência de isoformas produzidas pela planta.

A partir de modelos de diversos complexos entre EcTI e diferentes serinoproteases foi possível explicar a maioria dos dados experimentais de K_i relatados na literatura (Batista *et al.*, 1997). O parâmetro Sc de complementaridade de superfície não se mostrou um índice discriminador adequado entre complexos estáveis e não estáveis, uma vez que os resultados para o complexo com tripsina ($K_i = 0,88$ nM) e trombina (não inibe) foram parecidos (0,73 e 0,71, respectivamente). Isto pode ser em parte devido à minimização de energia a que foram submetidas as estruturas complexadas que tende a resolver problemas de impedimentos estéricos, por exemplo. A avaliação utilizando o programa Verify 3D de concordância entre estrutura e seqüência mostrou-se mais útil neste sentido, com o complexo trombina/EcTI apresentando valores negativos para alguns resíduos nesta avaliação, indicando incompatibilidade química na interface. Por outro lado, o complexo tripsina/EcTI apresenta boa compatibilidade química na interface.

A falta de atividade contra trombina parece ter uma contribuição que vem da repulsão eletrostática, porém a maior parte é devida à presença de inserções grandes na superfície de interação da trombina (em torno dos

resíduos 39,60 e 149). No caso do fator Xa apenas uma dessas inserções em torno do resíduo 60 está presente, mas leva a um impedimento estérico com a saída do *loop* reativo do inibidor. Espera-se que uma explicação parecida também seja aplicável para o caso do tPA, apesar de não ter sido modelado neste estudo. Nesse caso, é conhecido que o *loop* em torno de 60 gera impedimentos com o N-terminal do inibidor. Continua curiosa a observação da capacidade do EcTI inibir fator XIIa, uma vez que a seqüência de aminoácidos na região do resíduo 60 é muito parecida com tPA. Isto levanta a possibilidade de outros fatores estruturais estarem agindo no caso destas duas proteases, merecendo mais investigação. Particularmente a determinação da estrutura do complexo ente EcTI e fator XIIa seria de grande interesse.

No caso da inibição da quimotripsina propusemos um mecanismo parecido com o observado no complexo entre quimotripsina e BPTI. Ajustes na cadeia lateral da Arg63 demonstraram a possibilidade desta apontar a sua região alifática na direção do bolsão S1 e a formação simultânea de ligações de hidrogênio através do grupo guanidina. Estas ligações são formadas ou exclusivamente com resíduos da quimotripsina ou com resíduos do *loop* canônico do inibidor; isto representa uma explicação genérica para a inibição cruzada (tripsina e quimotripsina) em várias famílias de inibidores de protease.

Referências

Capítulo Cinco

5 Referências Bibliográficas

- Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. e Terwilliger, T. C. - PHENIX: Building New Software for Automated Crystallographic Structure Determination, *Acta Cryst.*, **D58**, 1948-1954, 2002.
- Anderson, C.M., McDonald, R.C. e Steitz, T.A. - Sequencing a Protein by X-ray Crystallography I. Interpretation of Yeast Hexokinase B, *J.Mol.Biol.*, **123**, 1-13, 1978a.
- Anderson, C.M., McDonald, R.C. e Steitz, T.A. - Sequencing a Protein by X-ray Crystallography II. Refinement of Yeast Hexokinase B, *J.Mol.Biol.*, **123**, 15-33, 1978b.
- Apostoluk, W. e Otlewski, J. - Variability of Canonical *Loop* Conformations in Serine Proteinases Inhibitors and Other Proteins. *Proteins: Struc. Function and Gen.*, **32**, 459-474, 1998.
- Barton, G. J. - Protein Multiple Sequence Alignment and Flexible Pattern Matching, *Methods .Enzymol.*, **183**, 403-428, 1990.
- Batista, I.F.C., Oliva, M.L.V., Araújo, M.S., Sampaio, M.U., Richardson, M., Fritz, H., Sampaio, M.U., Sampaio, C.A.M. e Garratt, R.C. - Primary Structure of a Kunitz-type Trypsin Inhibitor from *Enterolobium contortisiliquum* Seeds, *Phytochemistry*, **41**, 1017-1022, 1996.
- Batista, Isabel F. Correia - *Estrutura e especificidade de inibidores do tipo SBTI da família Kunitz purificados das sementes de Bauhinia unguiculata, Bauhinia variegata, Enterolobium contortisiliquum, Leucaena leucocephala e Swartzia pickellii*. São Paulo, 1997, 147p. Tese (Doutorado) - Universidade Federal de São Paulo.

- Batista, I.F.C., Bonfadini, M.R., Nonato, M.C., Beltramini, L.M., Oliva, M.L.V., Sampaio M.U., Sampaio, C.A.M. e Garratt, R.C. - Preliminary Crystallographic Studies of EcTI, a Serine Proteinase Inhibitor from *Enterolobium contortisiliquum* Seeds, *Acta Cryst.*, **D57**, 602-604, 2001.
- Berman, H.M., Westbrook, Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. e Bourne, P.E. - The Protein Data Bank, *Nucleic Acids Research*, **28**, 1, 235-242, 2000.
- Betzl, C., Lange, G., Pal, G.P., Wilson, K.S., Maelicke, A. e Saenger, W. - The Refined Crystal Structure of α -Cobratoxin from *Naja naja siamensis* at 2.4 Å Resolution, *J.Biol.Chem.*, **266**, 21530- 21536, 1991.
- Billings, P.C., Morrow, A.R., Ryan, C.A. e Kennedy, A.R. - Inhibition of Radiation-Induced Transformation of C3H/10T1/2 Cells by Carboxypeptidase Inhibitor I and Inhibitor II from Potatoes, *Carcinogenesis*, **10**, 687-691, 1989.
- Billings, P.C., Newberne, P.M. e Kennedy, A.R. - Protease Inhibitor Suppression of Colon and Anal Gland Carcinogenesis Induced by Dimethylhydrazine, *Carcinogenesis*, **11**, 1083-1086, 1990.
- Billings, P.C., St Clair, W., Ryan, C.A. e Kennedy, A.R. - Inhibition of Radiation-Induced Transformation of C3H/10T1/2 Cells by Chymotrypsin Inhibitor 1 from Potatoes, *Carcinogenesis*, **8**, 809-812, 1987.
- Blundell, T.L., Sibanda, B.L., Sternberg, M.J. e Thornton, J.M. - Knowledge-Based Prediction of Protein Structures and the Design of Novel Molecules, *Nature*, **326**, 347-352, 1987.
- Bode, W. e Huber, R. - Natural Protein Proteinase Inhibitors and Their Interaction with Proteinases, *Eur.J.Biochem.*, **204**, 433-451, 1992.
- Borges, D., Sampaio, C., de la Llosa, P. e Prado, J.L. - The Liver is the Main Organ to Clear Plasma and Tissue Kallikreins from Rat Plasma, in vivo, *Adv.Exp.Med.Biol.*, **198**, 229-233, 1986.
- Bowie, J.U., Lüthy, R. e Eissenberg, D. - A Method to Identify Protein Sequences That Fold Into a Known 3-Dimensional Structure, *Science*, **253**, 164-170, 1991.
- Brandén, C. e Tooze, J. - Introduction to Protein Structure, Garland Publishers, New York, 1991, ISBN 0-8153-0344-0.
- Brandstetter, H., Kühne, A., Bode, W., Huber, R., von der Saal, W., Wirthensohn, K. e Engh, R.A. - X-ray Structure of Active Site-Inhibited Clotting Factor Xa. Implications for Drug Design and Substrate Recognition, *J.Bio.Chem.*, **271**, 47, 29988-29992, 1996.

- Braun, W. e Go, N. – Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm, *J.Mol.Biol.*, **186**, 611-626, 1985.
- Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swamonathan, S. e Karplus, M. – CHARMM – A Program for Macromolecular Energy Minimization and Dynamics Calculations, *J.Comp.Chem.*, **4**, 187-217, 1983.
- Brünger, A.T. – Free R Value: a Novel Statistical Quantity for Assessing the Accuracy of Crystal Structures, *Nature*, **355**, 472-475, 1992.
- Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang-Sheng J., Kuszewski, J., Nigels, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. e Warren, G.L. – Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination, *Acta Cryst.*, **D54**, 905-921, 1998.
- CCP4 - Collaborative Computational Project, Number 4. “The CCP4 Suite: Programs for Protein Crystallography”, *Acta Cryst.*, **D50**, 760-763, 1994.
- Chothia, C. e Janin, J. - Orthogonal Packing of β -pleated Sheets in Proteins, *Biochemistry*, **21**, 3955-3965, 1982.
- Coombs, G.S., Rao, M.S., Olson, A.J., Dawson, P.E. e Madison, E.L. – Revisiting Catalysis by Chymotrypsin Family Serine Proteases Using Peptide Substrates and Inhibitors with Unnatural Main Chains, *J.Biol.Chem.*, **274**, 24074-24079, 1999.
- Dattagupta, J.,K., Podder, A., Chakrabarti, C., Sem, U., Dutta, S.K. e Singh, M. - Structure of a Kunitz-type Chymotrypsin Inhibitor from Winged Bean Seeds at 2.95 Å Resolution, *Acta Cryst.*, **D52**, 521-528, 1996.
- DeLano, W.L. – Manual Pymol, DeLano Scientific, San Carlos, CA, USA, copyright 1998-2001. (www.delanoscientific.com)
- de Meester, P., Brick, P., Lloyd, L.F., Blow, D.M. e Onesti, S. - Structure of the Kunitz-Type Soybean Trypsin Inhibitor (STI): Implication for the Interactions Between Members of the STI Family and Tissue-Plasminogen Activator, *Acta Cryst.*, **D54**, 589-597, 1998.
- do Socorro, M., Cavalcanti, M.S., Oliva, M.L.V., Fritz, M., Jochum, M., Mentele, R., Sampaio, M.U., Coelho, L.C., Batista, I,F,C,. Sampaio, C.A.M. – Characterization of a Kunitz Trypsin Inhibitor with One Disulfide Bridge Purified from *Swartzia pickellii*, *Biochem.Biophys.Res.Comm.*, **291**, 635-639, 2002.

- Dodson, E., Kleywegt, G.J., e Wilson, K. - Report of a Workshop on the Use of Statistical Validators in Protein X-ray Crystallography, *Acta Cryst.*, **D52**, 228-234, 1996.
- Doi, T., Tokunaga, T., Ohtsuka, E., Hiraki, Y., Suzuki, F., Ikehara, M. - Expression of the Amino Terminal Part of Synthetic Human Growth Hormone Gene and Somatomedin-like Activity of Expressed Protein, *Prot.Eng.*, **1**, 17-21, 1986.
- Drenth, J. - Principles of Protein Crystallography, Springer Advanced Texts in Chemistry, New York, 1999.
- Esmon, C.T. – The Roles of Protein C and Thrombomodulin in the Regulation of Blood Coagulation, *J.Biol.Chem.*, **264**, 4743-4746, 1989.
- Finzel, B.C., Clancy, L.L., Holland, D.R., Muchmore, S.W., Watenpaugh, K.D. e Einspahr, H.M. - Crystal Structure of Recombinant Human Interleukin-1 β at 2.0 Å Resolution, *J.Mol.Biol.*, **209**, 779-791, 1989.
- Fujinaga, M., Read, R.J., Sielecki, A., Ardelt, W., Laskowski, M., Jr., e James, M.N.G. – Refined Crystal Structure of the Molecular Complex of *Streptomyces griseus* Protease B, a Serine Protease, With the Third Domain of the Ovomucoid Inhibitor from Turkey , *Proc.Natl.Acad.Sci USA*, **79**, 4868-4872, 1982.
- García-Ruiz J.M. e Moreno A. – Investigations on Protein Crystal Growth by the Gel Acupuncture Method, *Acta Cryst.*, **D50**, 484-490, 1994.
- Graves, B.J., Hatada, M.H., Hendrickson, W.A., Miller, J.K., Madison, V.S. e Satow, Y. – Structure of Interleukin-1 α at 2.7Å Resolution, *Biochemistry*, **29**, 2679-2684, 1990.
- Greer, J. - Comparative Modeling Methods: Applications to the Family of the Mammalian Serine Proteases, *Proteins, Structure, Function & Genetics*, **7**, 317-334, 1990.
- Hazes, B., Magnu, K.A., Bonaventura, C., Bonaventura J., Dauter, Z., Kalk, K. H. e Hol, W.G.J. - Crystal Structure of Deoxygenated *Limulus polyphemus* Subunit I1 Hemocyanin at 2.18 Å Resolution: Clues for a Mechanism for Allosteric Regulation, *Prot.Science*, **2**, 597-619, 1993.
- Holton, T., Ioerger, T. R., Christopher, J. A. e Sacchettini, J. C. - Determining Protein Structure from Electron-Density Maps Using Pattern Matching, *Acta Cryst.*, **D56**, 722-734, 2000.
- Hubbard, S.J., Campbell, S.F. e Thornton J.M. - Molecular Recognition, Conformational Analysis of Limited Proteolytic Sites and Serine Proteinase Protein Inhibitors, *J.Mol.Biol.*, **220**, 507-530, 1991.

- Hubbard, S.J., Beynon, R.J. e Thornton, J.M. - Assessment of Conformational Parameters as Predictors of Limited Proteolytic Sites in Native Protein Structures, *Prot.Eng.*, **11**, 5, 349-359, 1998.
- Huber, R. e Bode, W. – Structural Basis of the Activation and Action of Trypsin, *Acc.Chem.Res.*, **11**, 114-122, 1978.
- Hutchinson, E.G. e Thornton, J.M. – PROMOTIF – A Program to Identify and Analyse Structural Motifs in Proteins, *Protein Sci.*, **5**, 212-220, 1990a.
- Hutchinson, E.G. e Thornton, J.M. – HERA - A Program to Draw Schematic Diagrams of Protein Secondary Structures, *Proteins*, **8**, 203-212, 1990b.
- Jackson, R.M. e Russel, R.B. – The Serine Protease Inhibitor Canonical Loop Conformation: Examples Found in Extracellular Hydrolases, Toxins, Cytokines and Viral Proteins, *J.Mol.Biol.*, **296**, 325-334, 2000.
- Janin, J. e Chothia, C. - The Structure of Protein-Protein Recognition Sites, *J.Mol.Chem.*, **265**, 16027-16030, 1990.
- Janson, J.C. e Rydén, L. Editors Protein Purification: Principles, High Resolution Methods and Applications, 2ª Edição, New York, Wiley-VCH, 1998.
- Jones, T.A. e Thornton, J. - Principles of Protein-Protein Interactions, *Proc.Natl.Acad.Sci. USA*, **93**, 13-20, 1996.
- Jones, T.A., Zou, J.Y., Cowan, S.W., e Kjeldgaard, M. – Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in These Models, *Acta Cryst.*, **A47**, 110-119, 1991.
- Katona, G., Wilmouth, R.C., Wright, P.A., Berglund, G.I., Hajdu, J., Neutze, R. e Schofield, J.C. - X-ray Structure of a Serine Protease Acyl-Enzyme Complex at 0.95Å Resolution, *J.Bio.Chem.*, **277**, 24, 21962-21970, 2002.
- Kendrew, J.C., Watson, H.C. - Strandberg, B.E., Dickerson, R.E., Phillips, D.C. e Shore, U.C. - A Partial Determination by X-ray Methods, and its Correlation with Chemical Data, *Nature*, **190**, 666-670, 1961.
- Kleywegt, G.J. - Validation of Protein Crystal Structures, *Acta Cryst.*, **D56**, 249-265, 2000.
- Kleywegt, G.J. e Jones, T.A. - Where Freedom is Given, Liberties Are Taken, *Structure*, **3**, 535-540, 1995.
- Kleywegt, G.J. e Jones, T.A. - Model-building and Refinement Practice, *Methods Enzymol.*, **277**, 208-230, 1997.

- Kleywegt, G.J. e Jones, T.A. – Databases in Protein Crystallography, *Acta Cryst.*, **D54**, 1119-1131, 1998.
- Kunitz, M. - Crystalline Soybean Trypsin Inhibitor, *J.Gen.Physiol.*, **30**, 291-310, 1947a.
- Kunitz, M. - Isolation of a Crystalline Protein Compound of Trypsin and Soybean Trypsin Inhibitor, *J.Gen.Physiol.*, **30**, 311-320, 1947b.
- Laskowski, M.Jr. e Kato, I. - Protein Inhibitor of Proteinases, *Annu.Rev.Biochem.*, **49**, 593-626, 1980.
- Laskowski, M.Jr. e Qasim, M.A. – What Can the Structures of Enzyme-Inhibitor Complexes Tell us About the Structures of Enzyme Substrate Complexes?, *Biochim.Biophys.Acta*, **1477**, 324-337, 2000.
- Laskowski, R.A., McArthur, M.W., Moss, D.S. e Thornton, J.M. – PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures, *J.Appl.Cryst.*, **26**, 283-291, 1993.
- Lawrence, M.C. e Colman, P.M. – Shape Complementarity at Protein/Protein Interfaces, *J.Mol.Biol.*, **234**, 946-950, 1993.
- Lehninger, A.L., Nelson, D.L. e Cox, M.M. - Principles of Biochemistry. 2^a Edição, Worth Publishers, New York, 1993.
- Lesk, A.M., Branden, C.I., Chothia, C. - Structural Principles of α/β Barrel Proteins: the Packing of the Interior of the Sheet, *Proteins*, **5**, 139-148, 1989.
- Liernier, I.E. - Possible Adverse Effects of Soybean Anticarcinogens, *J.Nutr.* **125**, 744S-759S, 1995.
- Lipscomb, W.N., Hartsuck, J.A., Quijcho, F.A. e Reeke, G.N. - The Structure of Carboxypeptidase A, IX. The X-ray Diffraction Results in the Light of the Chemical Sequence, *PNAS, USA*, **64**, 28-35, 1969.
- Longstaff, C., Campbell, A.F. e Fersht, A.R. – Recombinant Chymotrypsin Inhibitor 2: Expression, Kinetic Analysis of Inhibition with α -Chymotrypsin and Wild Type and Mutant Subtilisin BPNl and Protein Engineering to Investigate Inhibitory Specificity and Mechanism, *Biochemistry*, **29**, 7339-7347, 1990.
- Lorenzi, H., Árvores Brasileiras, Editora Plantarum, Nova Odessa, 1992.
- Lüthy, R, Bowie, J.U. e Eisenberg, D. - Assessment of Protein Models with Three-Dimensional Profiles, *Nature*, 356 (6364), 83-85, 1992.

- Maignan, S., Guilloteau, J.P., Pouzieux, S., Choi-Sledeski, Y.M., Becker, M. R., Klein, S.I., Ewing, W.R., Pauls, H.W., Spada, A.P. e Mikol, V. - Crystal Structures of Human Factor Xa Complexed with Potent Inhibitors *J.Med.Chem.*, **43**, 3226-3232, 2000.
- Marquart, M., Walter, J., Deisenhofer, J., Bode, W. e Huber, R. – The Geometry of the Reactive Site and of the Peptide Groups in Trypsin, Trypsinogen and Its Complexes With Inhibitors, *Acta Cryst.*, **B39**, 480-490, 1983.
- Martin, P.D., Makowski, M.G., DiMaio, J., Konishi, Y., Ni, F. e Edwards, B.F.P. – Bovine Thrombin Complexed with Uncleavable Analog os Residues 7-19 of Fibrinogênio A α : Geometry of the Catalytic Triad and interactions of the P1', P2 and P3' Substrate Residues, *Biochemistry*, **35**, 13030-13039, 1996.
- Marzzoco, A. e Torres, B.B., Bioquímica Básica, 2^a edição, Editora Guanabara/Koogan, Rio de Janeiro, 1999, ISBN 852770462-5.
- Matthews, B.W. - Solvent Content of Protein Crystals, *J.Mol.Biol.*, **33**, 491-497, 1968.
- Moran, L.A. e Scrimgeur, K.G. – Biochemistry Resource Book, Neil Patterson Publishers/Prentice Hall, Eaglewood Cliffs, 1994, ISBN 0-13-816679-X.
- McLachlan, A.D. - Three-fold Structural Pattern in The Soybean Trypsin Inhibitor (Kunitz), *J.Mol.Biol.*, **133**, 557-563, 1979.
- Moreno, A. e Soriano-García, M. - Crystal-growth Kinetics of Protein Single Crystals Along Capillary Tubes in the Gel-acupuncture Technique, *Acta Cryst.*, **D55**, 577-580, 1999.
- Moreno A., González-Ramírez, L.A., Hernández-Hernández, M.A., Oliver-Salvador, C., Soriano-García, M. e Rodríguez-Romero, A. – Investigations on Gravity Influence Upon Protein Crystallization by the Gel Acupuncture Technique, *J.Cryst.Growth*, **196**, 587-594, 1999.
- Morris, A.L., MacArthur, M.W., Hutchinson, E.G. e Thornton, J.M. - Stereochemical Quality of Protein Structure Coordinates, *Proteins*, **12**, 345-364, 1992.
- Murshudov, G.N., Vagin, A.A. e Dodson, E.J. - Refinement of Macromolecular Structures by the Maximum-Likelihood Method, *Acta Cryst.*, **D53**, 240-255, 1997.
- Murzin, A.G., Lesk, A.M. e Chothia, C. – β -trefoil Fold. Patterns of Structures and Sequence in the Kunitz Inhibitors Interleukins-1 β and 1 α and Fibroblast Growth Factors, *J.Mol. Biol.*, **223**, 531-543, 1992.

- Murzin, A.G., Lesk, A.M. e Chothia, C. – Principles Determining the Structure of β -Sheet Barrels in Proteins, I. A Theoretical Analysis, *J.Mol. Biol.*, **236**, 1369-1381, 1994a.
- Murzin, A.G., Lesk, A.M. e Chothia, C. – Principles Determining the Structure of β -Sheet Barrels in Proteins, II. The Observed Structures, *J.Mol. Biol.*, **236**, 1382-1400, 1994b.
- Nagai, K. e Thogersen, HC. - Generation of Beta-Globin by Sequence-Specific Proteolysis of a Hybrid Protein Produced in *Escherichia coli*, *Nature*, **309**, 810-812, 1984.
- Nagarajaram, H.A., Sowdhamini, R., Ramakrishnan C. e Balaram, P. – Termination of Right Handed Helices in Proteins by Residues in Left Handed Helical Conformations, *FEBS*, **321**, 1, 79-83, 1993.
- Navaza, J. e Saludjian, P. - AMoRE: An Automated Molecular Replacement Program Package, *Methods Enzymol.*, **276**, 581-594, 1997.
- Negreiros, A.N.M., Carvalho, M.M., Xavier-Filho, J., Blanco-Labra, A., Shewry, P.R. e Richardson, M., - The Complete Amino Acid Sequence of the Major Kunitz Trypsin Inhibitor from the Seeds of *Prosopis juliflora*, *Phytochemistry*, **30**, 2829-2833, 1991.
- Nicholls, A., Sharp, K.A., Honig, B. - Protein Folding and Association: Insights from the Interfacial and Thermodynamic Properties of Hydrocarbons, *Proteins*, **11**, 281-296, 1991.
- Oliva, M.L.V. - São Paulo, 2003, Inibidores de Proteinases isolados de Plantas, Tese (Livre Docência) - Universidade Federal de São Paulo.
- Onesti, S, Brick, P. e Blow, D.M. – Crystal Structure of a Kunitz-type Trypsin Inhibitor from *Erythrina caffra* seeds, *J.Mol.Biol.*, **217**, 153-176, 1991.
- Osawa, K. e Laskowski, M.Jr. - The Reactive Site of Trypsin Inhibitors, *J.Biol.Chem.*, **241**, 3955-3961, 1966.
- Otlewski, J. e Zbyryt, T. - Single Peptide Bond Hydrolysis/Resynthesis in Squash Inhibitors of Serine Proteinases, *Biochemistry*, **33**, 200-207, 1994.
- Otwinowski, Z. e Minor, W. - Processing of X-ray Diffraction Data Collected in Oscillation Mode, *Methods Enzymol.*, **276**, 307-326, 1997.
- Pavelcik, F., Zelinka, J. e Otwinowski, Z. - Methodology and Applications of Automatic Electron-Density Map Interpretation by Six-Dimensional Rotational and Translational Search for Molecular Fragments, *Acta Cryst.*, **D58**, 275-283, 2002.

- Peräkylä, M. e Kollman, P.A. – Why Does Trypsin Cleave BPTI So Slowly? *J.Am.Chem.Soc.*, **122**, 3436-3444, 2000.
- Perrakis, A., Morris, R. e Lamzin, V.S. – Automated Protein Model Building Combined with Interactive Structure Refinement, *Nature Struc.Bio.*, **6**, 5, 458-463, 1999.
- Polikarpov, I., Perles, L.A., de Oliveira, R.T., Oliva, G., Castellano, E. E., Garratt, R. C. e Craievich, A. - Set-up and Experimental Parameters of the Protein Crystallography Beamline at the Brazilian National Synchrotron Laboratory, *J. Synchrotron Rad.*, **5**, 72-76, 1998.
- Ponting, C.P. e Russel, R.B. – Identification of Distant Homologues of Fibroblast Growth Factor Suggests a Common Ancestor for all β -Trefoil Proteins, *J.Mol.Biol.*, **302**, 1041-1047, 2000.
- Press, W.H., Flannery, B.P., Tenkolsky, S.A. e Vetterling, W.T., Numerical Recipes, Cambridge University Press, Cambridge, 1988.
- Priestle, J.P., Schar, H.P. e Gruntter, M.G. – Crystallographic Refinement of Interleukin 1- β at 2.0 Å Resolution, *PNAS USA*, **86**, 9667-9671, 1989.
- Radisky, E.S. e Koshland, Jr., D.E. – A Clogged Gutter Mechanism for Protease Inhibitors, *Proc.Natl.Acad.Sci. USA*, **99**, 10316-10321, 2002.
- Ramakrishnan, C. e Ramachandran, G.N. – Stereochemical Criteria for Polypeptide and Protein Chain Conformations. II. Allowed Conformations for a Pair of Peptide Units, *Biophys.J.*, 5(6), 909-933, 1965.
- Richards, F.M. – Areas, Volumes, Packing and Protein Structures, *Annu.Rev.Biophysics.Bioeng.*, **6**, 151-176, 1977.
- Richardson, M. - Seed Storage Proteins: the Enzyme Inhibitors, *Methods Plant.Biochem.*, **5**, 259-305, 1991.
- Richardson, J.S. e Richardson, D.C. – Amino-acid Preferences for Specific Locations at the Ends of α -Helices, *Science*, **240**, 1648-1652, 1988.
- Robertus, J.D., Alden, R.A., Birktoft, J.J., Kraut, J., Powers, J.C. e Wilcox, P.E. - An X-ray Crystallographic Study of the Binding of Peptide Chloromethyl Ketone Inhibitors to Subtilisin BPN', *Biochemistry*, **11**, 2439-2449, 1972.
- Rossmann, M.G., Editor – The molecular Replacement Method. New York, Gordon and Breach, 1972.

- Salemme, F.R. e Weatherford, D.W. – Conformation and Geometrical Properties of β -Sheets in Proteins. I. Parallel β -Sheets, *J.Mol.Biol.*, **146**, 101-117, 1981a.
- Salemme, F.R. e Weatherford, D.W. – Conformation and Geometrical Properties of β -Sheets in Proteins. II. Antiparallel and Mixed β -Sheets, *J.Mol.Biol.*, **146**, 119-141, 1981b.
- Sali, A. e Blundell, T.L. - Comparative Protein Modelling by Satisfaction of Spatial Restraints, *J.Mol.Biol.*, **234**, 779-815, 1993.
- Sampaio, C.A.M., Oliva, M.L.V, Tanaka, A.S. e Sampaio M.U. - Proteinase inhibitors in brazilian leguminosae, *Mem.Inst.Oswaldo Cruz*, **86**, 207-209, 1991.
- Sanchez, R. e Sali, A., - Advances in Comparative Protein-Structure Modelling, *Curr.Opin.Str.Biol.*, **7**, 206-214, 1997.
- Schechter, I. e Berger, A. - On the Size of Active Site in Proteases I. Papain, *Biochem.Biophys.Res.Comm.*, **27**, 157-162, 1967.
- Scheidig, A.J., Hynes, T.R., Pelletier, L.A., Wells, J.A. e Kossiakoff, A.A. - Crystal Structures of Bovine Chymotrypsin and Trypsin Complexed to the Inhibitor Domain of Alzheimer's Amyloid β -Protein Precursor (APPI) and Basic Pancreatic Trypsin Inhibitor (BPTI): Engineering of Inhibitors with Altered Specificities, *Protein Sci*, **6**, 1806-1824, 1997.
- Singh, J., Thornton, J.M., Snarey, M. e Campbell, S.F. - The Geometries of Interacting Arginine-Carboxyls in Proteins, *FEBS Letts*, **224**, 161-171, 1987.
- Smith, D.B. e Johnson, K.S. - Single-step Purification of Polypeptides Expressed in *Escherichia coli* as Fusions with Glutathione S-Transferase, *Gene*, **67**, 31-40, 1988.
- Song, H.K. e 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**, 347-363, 1998.
- Stubbs, M.T. e Bode, W. – Structure and Specificity in Coagulation and Its Inhibition, *Trends Cardiovasc.Med.*, **5**, 157-166, 1995.
- Takeuchi, Y., Satow, Y., Nakamura, K.T. e Mitsui, Y. – Refined Crystal Structure of the Complex of Subtilisin BPN^o and Streptomyces Subtilisin Inhibitor at 1.8 Å Resolution, *J.Mol.Biol.*, **221**, 309-325, 1991.
- Terwilliger, T.C. - Automated Structure Solution, Density Modification and Model Building, *Acta Cryst.*, **D58**, 1937-1940, 2002.

- Travis, J. e Salvenssen, G. - Human Plasma Proteinase Inhibitors, *Annu.Rev.Biochem.*, **52**, 655 – 709, 1983.
- Troll, W., Wiesner, R. e Frenkel, K. - Anticarcinogenic Action of Protease Inhibitors, *Adv.Cancer Res.*, **49**, 265-283, 1987.
- Troll, W., Wiesner, R., Shellabarger, C.J., Holtzman, S. e Stone, J.P. - Soybean Diet Lowers Breast Tumor Incidence in Irradiated Rats, *Carcinogenesis*, **1**, 469-472, 1980.
- van de Locht, A., Bode, W., Huber, R., Le Bonniec, B.F., Stone, S.R., Esmon, C.T. e Stubbs, M.T. - The Thrombin E192Q-BPTI Complex Reveals Gross Structural Rearrangements: Implications for the Interaction with Antithrombin and Thrombomodulin, *EMBO J.*, **16**, 2977 - 2984, 1997.
- Voet, D. e Voet, J. – Biochemistry, 2^a edição, John Wiley & Sons, INC, N.York, 1995.
- Vriend, G. - WHAT IF: A Molecular Modeling and Drug Design Program, *J.Mol.Graph.*, **8**, 52-56, 1990.
- Vriend, G. e Sander, C. - Quality Control of Protein Models: Directional Atomic Contact Analysis, *J.Appl.Cryst.*, **26**, 47-60, 1993.
- Walenga, J.M., Jeske, W.P., Hoppensteadt, D. e Fareed, J. - Factor Xa Inhibitors: Today and Beyond, *Curr.Opin.Investig.Drugs*, 4(3), 272-281, 2003.
- Watenpaugh, K.D., Sieker, L.C., Herriott, J.R. e Jensen, L.H. - Refinement of the Model of a Protein: Rubredoxin at 1.5 Å Resolution, *Acta Cryst.*, **B29**, 943-956, 1973.
- Weiss, M.S. e Schulz, G.E. - Structure of Porin Refined at 1.8 Å Resolution, *J.Mol.Biol.*, **227**, 493-509, 1992.
- Wilmot, C.M. e Thornton, J.M. – *Prot.Eng.*, **3**, 479-493, 1990.
- Wilmouth, R.C., Edman, K., Neutze, R., Wright, P.A., Clifton, I.J., Scheider, T.R., Schofield, C.J. e Hadju, J. - X-ray Snapshots of Serine Protease Catalysis Reveal a Tetrahedral Intermediate, *Nature Struc.Bio.*, **8**, 689–694, 2001.
- Witschi, H. e Kennedy, A.R. - Modulation of Lung Tumor Development in Mice with the Soybean-Derived Bowman-Birk, *Carcinogenesis*, **10**, 2275-2277, 1989.

Yavelow, J., Finlay, T.H., Kennedy, A.R. e Troll, W. - Bowman-Birk Soybean Protease Inhibitor as an Anticarcinogen, *Cancer Res.*, **43**, 2554-2459, 1983.

Zhu, X., Komiya, H., Chirino, A., Faham, S., Fox, G.M., Arakawa, T., Hsu, B.T. e Rees, D.C. – Three-Dimensional Structures of Acidic and Basic Fibroblast Growth Factors, *Science*, **251**, 90-93, 1991.