



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS BIOLÓGICAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM  
BIOINFORMÁTICA



# ANÁLISE DE PADRÕES DE INTERAÇÃO ENTRE SERINO PROTEASES E SEUS INIBIDORES PROTÉICOS

Tese apresentada ao Curso de Pós-graduação em Bioinformática da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Doutor em Bioinformática.

**CRISTINA RIBEIRO**

**Orientador: Marcelo Matos Santoro**  
**Co-Orientador: Goran Neshich**

Belo Horizonte

08 de abril de 2009



UNIVERSIDADE FEDERAL DE MINAS GERAIS



## Folha de Aprovação

Análise de Padrões de Interação entre Serino Proteases e seus Inibidores Protéicos

CRISTINA RIBEIRO

Tese defendida e aprovada pela banca examinadora constituída por

Prof. Ph. D. MARCELO MATOS SANTORO – Orientador  
Universidade Federal de Minas Gerais

Ph. D. GORAN NESHICH – Co-Orientador  
Empresa Brasileira de Pesquisa Agropecuária

Prof. Ph. D. MONIQUE ARAÚJO DE BRITO  
Universidade Federal Fluminense

Prof. Ph. D. CAIO JÚLIO MARTINS VELOSO  
Pontifícia Universidade Católica de Minas Gerais

Prof. Ph. D. RONALDO ALVES PINTO NAGEM  
Universidade Federal de Minas Gerais

Ph. D. HUMBERTO D’MUNIZ PEREIRA  
Universidade de São Paulo

Belo Horizonte 08 de abril de 2009

*“Matar o sonho é matarmo-nos. É mutilar a nossa alma. O sonho é o que temos de realmente nosso, de impenetravelmente e inexpugnavelmente nosso”*

**Fernando Pessoa**

# Agradecimentos

A Deus, pela vida.

Aos meus pais simplesmente por tudo não há como citar ou descrever o que fizeram por mim ao longo desses 30 anos.

Agradeço ao Guilherme, pelo carinho e apoio incondicional para que eu fizesse esse doutorado, que tornou a minha vida mais leve e radiante. A você um beijo bom.

A minhas irmãs, Patrícia e Renata, pelos momentos de apoio e amor. Agradeço também ao meu cunhado Rodrigo “legalzão demais” que foi mais que um companheiro um grande amigo, um irmão.

Aos professores e mestres Marcelo Santoro e Goran Neshich, que, mais do que orientadores, foram amigos que ofereceram sua grande inteligência e capacidade a serviço dessa tese e de meu auxílio.

Aos amigos que conheci ao longo desses quatro anos; os amigos da Bioinformática, os amigos Biólogos, a todos vocês um abraço bem apertado, carrego um pouco de vocês comigo, e deixo um pouco de mim. Em especial o meu grupo de laboratório Kloro, Caio, Raquel, Waisberg vocês contribuíram muito para engrandecer o meu conhecimento.

A todos professores que estiveram presente durante esse doutorado e contribuíram muito com os seus conhecimentos.

Ao professor Miguel pela grande amizade desde a primeira vez que estive na UFMG, por me receber de braços abertos.

A minha amiga Ceci, pela vitalidade louca, pela ternura, pelo companheirismo, pela integridade de sua pessoa, pela sua alma que não tem tamanho.

A minha nova grande amiga Maíra pelo grande apoio e serenidade na reta final.

## Resumo

Para estudar a especificidade *in silico* de serino proteases, é primeiramente necessário providenciar um volume de dados suficiente para analisar a interface ao redor do sítio catalítico daquelas enzimas que não possuem informações sobre a formação de complexo de uma protease particular com alguma proteína/substrato ou inibidor específico. Uma abordagem *in silico* foi desenvolvida para construirmos os novos complexos. A característica chave deste trabalho é mapear os resíduos formadores da interface (Interface Forming Residues - IFR) em um perfil tridimensional e depois em uma matriz bidimensional, a partir do conhecimento estrutural enzima-inibidor protéico daqueles complexos com estruturas conhecidas. O mapeamento é feito depois do alinhamento estrutural de todas as serino proteases com sequência não redundantes. Propusemos uma nova metodologia para definir uma tabela bem curada de especificidade de serino proteases determinadas pelos resíduos formadores da interface. Os IFRs foram obtidos através do "hard body docking" entre o alinhamento estrutural de 70 serino proteases, com sequências não redundantes, em conjunto com três complexos serino proteases - inibidor. Os inibidores são: ecotina, ovomucoide do terceiro domínio e BPTI (Bovine Pancreatic Trypsin Inhibitor). Depois de calcularmos quais aminoácidos perdiam acessibilidade ao solvente após os complexos entre serino proteases e os três diferentes inibidores serem formados, foi criada uma tabela com todas as posições de aminoácidos que faziam parte da interface e sua respectiva área de ocupação. Pelo mapeamento desses aminoácidos, foi possível analisar suas características, o que nos permitiu marcar posições específicas no alinhamento dentre as diferentes subfamílias de serino proteases. Nas superfícies que não estavam incluídas nas interfaces das serino proteases, foi encontrada uma prevalência geral de resíduos carregados, enquanto as interfaces continham mais resíduos polares e glicinas, com exceção de algumas subfamílias. O bolsão do IFR não era formado somente por interações hidrofóbicas, como poderia ser esperado, sendo um ambiente em geral mais polar. Entretanto, a interface contém muito menos resíduos carregados, exceto em uma ou duas subfamílias, se comparado com o resto da superfície da enzima. Nosso trabalho gerou uma ferramenta original para analisar o relacionamento estrutura/função, bem como indicações concretas de como podemos encontrar ou alterar a especificidade de interação de várias serino proteases com os inibidores protéicos.

## Abstract

In order to study serine protease specificity *in silico*, it is first necessary to provide sufficient volume of data for analysis of the interfaces around catalytic sites of those enzymes even if there is no currently available information on complex formation of one particular protease with any specific substrate and/or inhibitor. An *in silico* approach was designed for building corresponding complexes. The key feature of our work is mapping the Interface Forming Residue (IFR) 3D profile into 2D matrix, from known enzyme-inhibitor structure to those complexes with no known structure. Mapping is done after structurally aligning all serine proteases with non-redundant sequences. We propose a novel method for defining an exhaustive table of serine proteases specificity by determining interface forming residues (IFR). The IFRs are obtained by “hard body docking” among 70 structurally aligned, sequence wise non redundant, serine protease structures, with 3 inhibitors: ecotine, ovomucoid third domain inhibitor and bovine pancreatic trypsin inhibitor (BPTI). After calculating which amino acids are becoming shielded from a solvent upon complex formation among serine proteases and different types of inhibitors, the table (matrix) of all amino acid positions at the interface and their respective occupancy is created. By mapping those amino acids, we are able to analyze their characteristics and by doing so we also can make position specific alignment among different sub families of serine proteases. The surfaces (not including interface area) of serine proteases are showing a general prevalence for charged residues while interfaces are containing more polar and glycine residues (with some exceptions holding). Thus, the IFR pocket is not formed only by hydrophobic interactions as one could expect. The IFR pocket is a rather polar environment in general. However, the interface contains much less (with one or two exceptions) charged residues if compared to the rest of the enzyme surface. Our work here is providing a unique tool for both structure/function relationship analysis, as well as concrete indications on how specificity of various serine proteases is achieved or could be altered.

# Sumário

Introdução .....	12
1. As Proteases .....	13
1.1. Protease, Proteinase ou Peptidase? .....	14
1.2. Serino Proteases .....	16
2. Os Inibidores de Serino Proteases.....	21
2.1. Inibidores do Plasma Sanguíneo .....	22
2.1.1. As Macroglobulinas .....	23
2.1.2. As Serpinas .....	24
2.1.3. As Hirudinas .....	26
2.2. Inibidores com Sítio Ativo Classe-Específicos .....	27
2.3. Os Inibidores Canônicos de Serino Proteases .....	27
2.4. O Início dos Estudos Sobre Inibidores .....	28
2.5. Mecanismo de Ação dos Inibidores.....	29
2.6. Proteólise Limitada .....	31
2.7. <i>Loop</i> Reativo Canônico .....	32
2.8. Famílias de Inibidores de Serino Proteases .....	33
2.8.1. Família BPTI Kunitz .....	35
2.8.2. Família Kazal.....	38
2.8.3. Família Ecotina .....	39
2.8.4. Outras Famílias de Inibidores .....	40
2.9 Funções e Aplicações dos Inibidores de Proteases.....	47
Objetivo Geral .....	48
Objetivos Específicos.....	48
Materiais e Métodos .....	49
1. Bancos de Dados de Proteínas.....	49
1.1. Uniprot.....	49
1.2. PDB.....	49
1.3. SCOP .....	50
1.4. STING.....	51

2. Seleção da Base de Dados de Serino Proteases .....	52
3. Seleção dos Complexos Proteína/Inibidor .....	58
4. Alinhamento Estrutural Usando o PrISM .....	59
5. <i>Hard Body Docking</i> .....	60
6. Determinação da Superfície em Complexos <i>in silico</i> .....	62
7. Tabela de Resíduos Formadores da Interface .....	62
Resultados e Discussão.....	64
1. Análise das Interfaces .....	65
2. Superfície versus Interface .....	75
3. Impedimentos Estéricos .....	78
4. Validação da Metodologia .....	81
Conclusão .....	83
Perspectivas.....	84
Referências Bibliográficas .....	86
Anexo A .....	93
Anexo B.....	114



## Lista de Figuras

<b>Figura 1.</b> Esquema moderno de nomenclatura usada na classificação das proteases.	15
<b>Figura 2.</b> Nomenclatura dos subsítios de uma protease e os resíduos complementares de seu substrato (Schechter & Berger, 1967).....	15
<b>Figura 3.</b> Arranjo característico das serino proteases. ....	17
<b>Figura 4.</b> Esquema do mecanismo catalítico das serino proteases (Jeremy <i>et al.</i> , 2002). .....	19
<b>Figura 5.</b> Estrutura 3D da $\alpha$ -2 macroglobulina (código PDB 1QSJ) mostrando as quatro cadeias idênticas que compõem a molécula. ....	24
<b>Figura 6.</b> (A) estrutura tridimensional da serpina $\alpha$ 1-antitripsina humana (PDB 1QLP), (B) estrutura tridimensional da hirudina extraída de <i>Hirudo medicinalis</i> (PDB 2HIR). ....	25
<b>Figura 7.</b> Reação que ilustra o mecanismo geral seguindo pelos inibidores canônicos. .....	30
<b>Figura 8.</b> Superposição dos átomos das cadeias principais dos 6 resíduos de aminoácidos do <i>loop</i> reativo.....	33
<b>Figura 9.</b> Estrutura tridimensional de BPTI extraída de <i>Bos taurus</i> (Boi, PDB 1BPI). .	36
<b>Figura 10.</b> Estrutura tridimensional do inibidor de tripsina humano do tipo Kazal (PDB 1HPT). ....	38
<b>Figura 11.</b> Estrutura tridimensional da ecotina de <i>Escherichia coli</i> (Bactéria, PDB 1ECY). ....	40
<b>Figura 12.</b> Estrutura tridimensional de STI extraído da <i>Glycine Max</i> (soja, PDB 1AVU). ....	42
<b>Figura 13.</b> (A) Estrutura tridimensional de BBI extraído da <i>Pisum sativum</i> (ervilha, PDB 1PBI). ....	43
<b>Figura 14.</b> Estrutura tridimensional do inibidor tipo batata I extraído de <i>Cucurbita maxima</i> (abóbora, PDB 1MIT). ....	43
<b>Figura 15.</b> Estrutura tridimensional do inibidor tipo batata II extraído de <i>Nicotiana glauca</i> (fumo-de-jardim, PDB 1TIH). ....	44
<b>Figura 16.</b> Estrutura tridimensional do inibidor tipo abóbora extraído de <i>Cucurbita maxima</i> (abóbora, PDB 3CTI). ....	44

<b>Figura 17.</b> Estrutura tridimensional do inibidor de subtilisina extraído de <i>Streptomyces albogriseolus</i> da família SSI (Bactéria, PDB 3SSI).....	45
<b>Figura 18.</b> Estrutura tridimensional do inibidor da família das antistasinas extraído da saliva de sanguessuga <i>Hirudo medicinalis</i> (PDB 1HIA).....	45
<b>Figura 19.</b> Estrutura tridimensional da elafina humana (PDB 2REL) que possui o arranjo característico da família chelonianina.....	46
<b>Figura 20.</b> Estrutura tridimensional do inibidor da família ascaris extraído de <i>Ascaris lumbricoides</i> (verme, PDB 1ATA).....	46
<b>Figura 21.</b> Fluxograma de busca e seleção das serino proteases. ....	53
<b>Figura 22.</b> Alinhamento estrutural de 68 serino proteases cujas sequências possuem no máximo 95 % identidade entre si, alinhadas pelo pacote PrISM.....	60
<b>Figura 23.</b> <i>Hard body docking</i> das coordenadas do inibidor com outras proteases.....	61
<b>Figura 24.</b> Diferença das classes dos resíduos de ocupação, entre o total da superfície livre da enzima versus interface.....	77
<b>Figura 25.</b> Proximidade entre resíduos pertencentes aos IFRs de 70 serino proteases e os resíduos pertencente aos IFRs dos inibidores Ecotina(A), BPTI (B) e Ovomucoide (C). ....	80

## Lista de Tabelas

<b>Tabela 1.</b> Classificação atual das serino proteases. (Fonte: <a href="http://merops.sanger.ac.uk/">http://merops.sanger.ac.uk/</a> ). .....	20
<b>Tabela 2.</b> Principais inibidores de proteases presentes no plasma sanguíneo humano. ....	23
<b>Tabela 3.</b> As famílias de inibidores de serino proteases com complexos determinados por cristalografia de raio-X .....	34
<b>Tabela 4.</b> Membros da Família Kunitz BPTI, inibidores de serino proteases. ....	37
<b>Tabela 5.</b> Membros da Família Kazal, inibidores de serino proteases. ....	39
<b>Tabela 6.</b> 67 serino proteases não redundantes com estruturas 3D resolvidas encontradas dentro do PDB, agrupadas de acordo com as suas subfamílias. ....	54
<b>Tabela 7.</b> 67 serino proteases não redundantes e o seu respectivo sítio de clivagem preferencial, as proteínas estão agrupadas por subfamílias de acordo com a Tabela 6. ....	56
<b>Tabela 8.</b> Posições dos IFRs de 70 serino proteases ligadas com os inibidores: Ecotina (A), BPTI (B), e Ovomucoide (C). ....	67
<b>Tabela 9.</b> Posições dos IFRs para quatro subfamílias de serino proteases, ligadas a ecotina. ....	71
<b>Tabela 10.</b> Posições dos IFRs para quatro subfamílias de serino proteases, ligadas a BPTI. ....	73
<b>Tabela 11.</b> Posições dos IFRs para quatro subfamílias de serino proteases, ligadas a Ovomucoide. ....	74
<b>Tabela 12.</b> Restrição ao tipo de resíduos presente nas interfaces entre quatro subfamílias de serino proteases e os inibidores protéicos, ecotina, BPTI e Ovomucoide. ....	75
<b>Tabela 13.</b> Porcentagem da área da interface ocupada por quatro classes de aminoácidos. ....	76
<b>Tabela 14.</b> Valores usados para calcular matriz de confusão (VP,VN, FP, FN) e acurácia (AC) do1FY8 (controle) e dos outros 16 complexos complementares. ....	82

## Introdução

Praticamente todos os processos biológicos requerem o reconhecimento de um ligante por uma macromolécula biológica, um processo conhecido como reconhecimento molecular. Um dos desafios da biofísico-química é poder prever a afinidade entre uma proteína e um ligante com base no conhecimento estrutural. Essa habilidade para prever a afinidade baseada na estrutura serve como base para a compreensão da relação estrutura-função e para o desenho racional de fármacos (Greighton, 1983; Lehninger *et al.*, 1993).

Uma abordagem elegante para entender esse tipo de relação envolve os estudos que investigam os determinantes estruturais da especificidade das serino proteases da subfamília da quimotripsina. São estudos estruturais e mutacionais desenvolvidos com a finalidade de entender as estratégias que a natureza usa para alterar essa especificidade. Em alguns casos essas estratégias são fáceis de entender, pois a alteração de alguns resíduos da enzima que fazem contatos com o substrato é suficiente para mudar a especificidade (Bone *et al.*, 1989a; Tsu *et al.*, 1997). Entretanto, é necessário mapear o maior número de descritores desta interação para entender completamente a especificidade das moléculas em estudo.

As serino proteases da família da quimotripsina estão envolvidas num grande número de importantes processos celulares, frequentemente relacionados com doenças fisiopatológicas severas, incluindo digestão, coagulação sangüínea, dissolução de coágulos, regulação da pressão sangüínea, maturação protéica, ativação do sistema imune e troca de tecido celular danificado (Neurath, 1989).

Com o conhecimento das estruturas tridimensionais destas macromoléculas é possível o desenvolvimento de fármacos que atuem de forma específica, minimizando efeitos colaterais. Pode-se citar como exemplo os inibidores que compõe o coquetel usado no tratamento da AIDS, que foram desenhados para uma ação específica contra a HIV-protease sem afetar colateralmente qualquer proteína

humana. Obviamente isso se tornou possível a partir do conhecimento da estrutura atômica tridimensional, particularmente do sítio ativo da protease do vírus (Hooper, 2002).

## 1. As Proteases

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

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

- 1) **Proteólise limitada**, na qual uma protease cliva apenas uma ou um número limitado de ligações peptídicas de uma determinada proteína levando à ativação ou maturação ou a formação de uma proteína inativa.
- 2) **Proteólise ilimitada**, na qual proteínas são degradadas em seus aminoácidos constituintes.

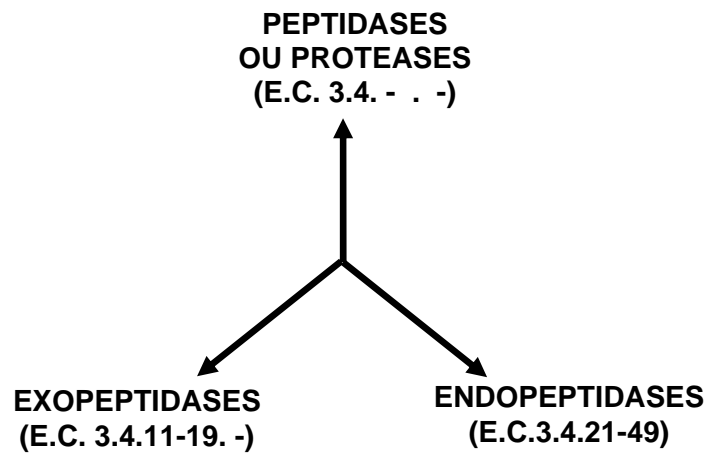
A degradação das proteínas pode se dar de duas formas: na primeira, as proteínas são inicialmente conjugadas a um marcador e então rapidamente hidrolisadas pelo proteosoma na presença de ATP; a outra via consiste na compartimentalização das proteases, por exemplo, em lisossomas, com transferência das proteínas para este compartimento, onde são rapidamente degradadas.

## 1.1. Protease, Proteinase ou Peptidase?

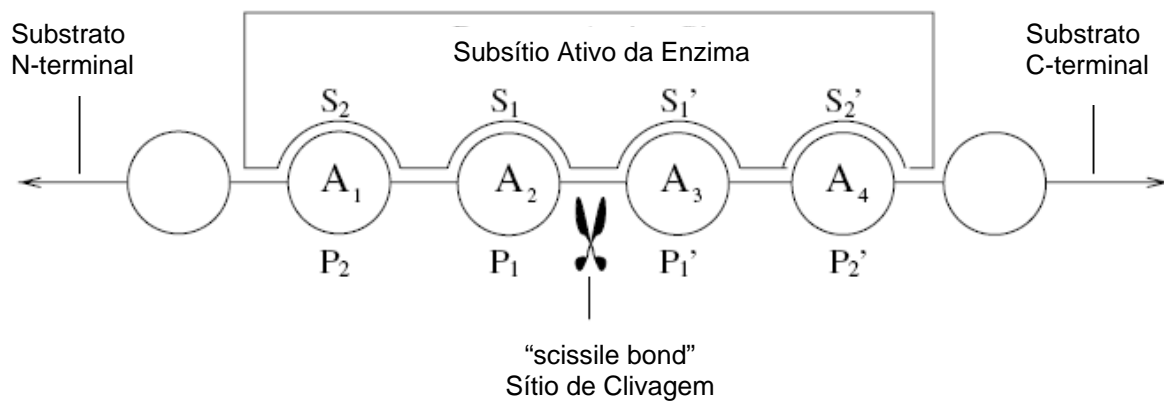
A União Internacional de Bioquímica e Biologia Molecular (UIBBM) tem recomendado desde 1984 o uso do termo *peptidase* para o subgrupo de hidrolases de ligações peptídicas (Subclasse E.C 3.4.) (Enzyme Nomenclature, (1992)). O termo *protease* amplamente usado é sinônimo de *peptidase*. Nessa tese adotaremos o termo *proteases*. *Peptidases* compreendem dois tipos de enzimas: as endopeptidases e as exopeptidases, as quais clivam ligações peptídicas em pontos na proteína e removem aminoácidos seqüencialmente do N-terminal ou C-terminal. O termo *proteinase* é também usado como sinônimo para *endopeptidase* (Barrett *et al.*, 2004). O esquema moderno de nomenclatura pode ser representado como mostrado na Figura 1.

A nomenclatura de interação de um substrato com uma protease foi introduzida em 1967 por Schechter e Berger e é agora amplamente usada na literatura.

Nesse sistema é considerado que os resíduos de aminoácidos do substrato polipeptídico se ligam em subsítios da enzima ativa. Por convenção, esses subsítios na protease são chamados S (subsítios) e os resíduos de aminoácidos do substrato são chamados P (para peptídeo). Os resíduos de aminoácidos do lado N-terminal da ligação peptídica do substrato que é hidrolisada pela protease são numerados P<sub>n</sub>, P<sub>(n-1)</sub>, ..., P<sub>3</sub>, P<sub>2</sub>, P<sub>1</sub> e os resíduos do lado C-terminal são numerados P<sub>1</sub>' , P<sub>2</sub>' , P<sub>3</sub>' , ..., P<sub>(n-1)</sub>' , P<sub>n</sub>'. Os resíduos P<sub>1</sub>-P<sub>1</sub>' formam uma ligação chamada "scissile" ou ligação peptídica do substrato que é clivada pela protease na hidrólise. Os subsítios na protease que complementam os resíduos do substrato ligado são numerados S<sub>n</sub>, S<sub>(n-1)</sub>, ..., S<sub>3</sub>, S<sub>2</sub>, S<sub>1</sub>, S<sub>1</sub>' , S<sub>2</sub>' , S<sub>3</sub>' , ..., S<sub>(n-1)</sub>' , S<sub>n</sub>' como mostrado na Figura 2 (Schechter & Berger, 1967).



**Figura 1.** Esquema moderno de nomenclatura usada na classificação das proteases.



**Figura 2.** Nomenclatura dos subsítios de uma protease e os resíduos complementares de seu substrato (Schechter & Berger, 1967).

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

Quatro classes mecanísticas tem sido reconhecidas pela UIBBM: proteases aspárticas, proteases cisteínicas, metaloproteases e serino proteases. Além destas quatro classes mecanísticas, existe ainda uma seção da nomenclatura das enzimas que é destinada para proteases cujo mecanismo catalítico ainda não foi identificado (Enzyme Nomenclature, 1992). Neste trabalho iremos focar o nosso estudo somente nas serino proteases, pois dentro dessas citadas acima é a que possui uma maior quantidade de proteínas com estrutura tridimensional resolvida, o que foi fundamental para o desenvolvimento dessa tese.

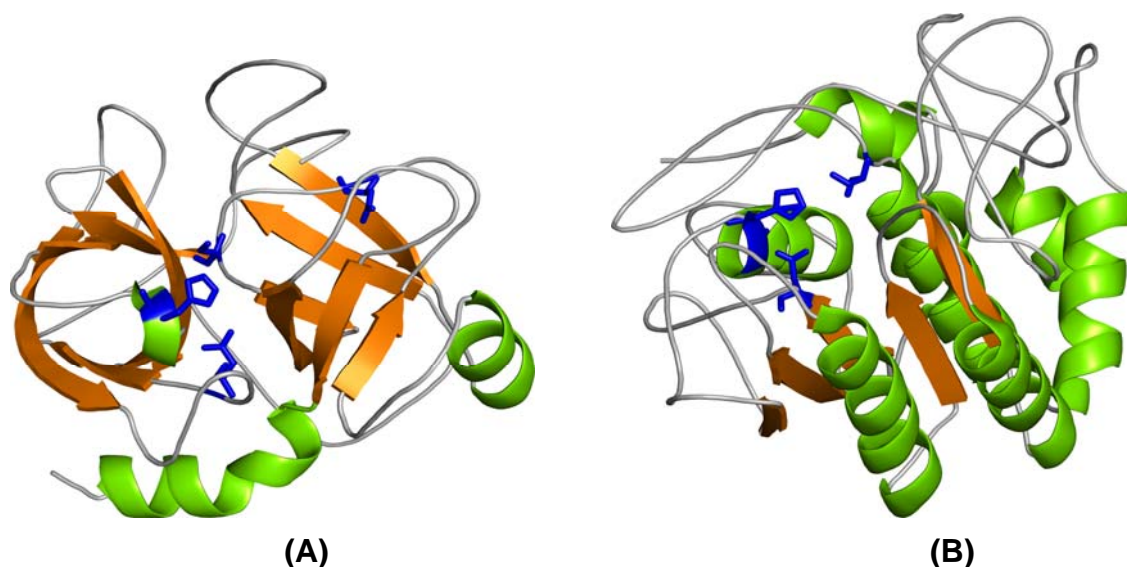
## **1.2. Serino Proteases**

Essa classe compreende duas famílias distintas:

- 1) Família da quimotripsina, a qual inclui as enzimas de mamíferos como quimotripsina, tripsina, elastase, calicreína, trombina, etc.;
- 2) Família da subtilisina, a qual inclui as enzimas de bactéria (como a subtilisina) (Hedstrom, 2002).

A estrutura 3D geral, na Figura 3, é diferente nas duas famílias, mas elas possuem a mesma geometria para o sítio ativo e por isso a catálise se dá pelo mesmo mecanismo. As serino proteases exibem diferentes especificidades por substrato o que está relacionado com as substituições de aminoácidos nos vários subsítios das enzimas, os quais são nomeados de acordo com a nomenclatura Schechter e Berger em 1967, interagindo com resíduos do substrato. Algumas enzimas têm um sítio de interação estendido com o substrato enquanto outras interagem restritamente com resíduo na posição P1 do substrato (Hedstrom, 2002).



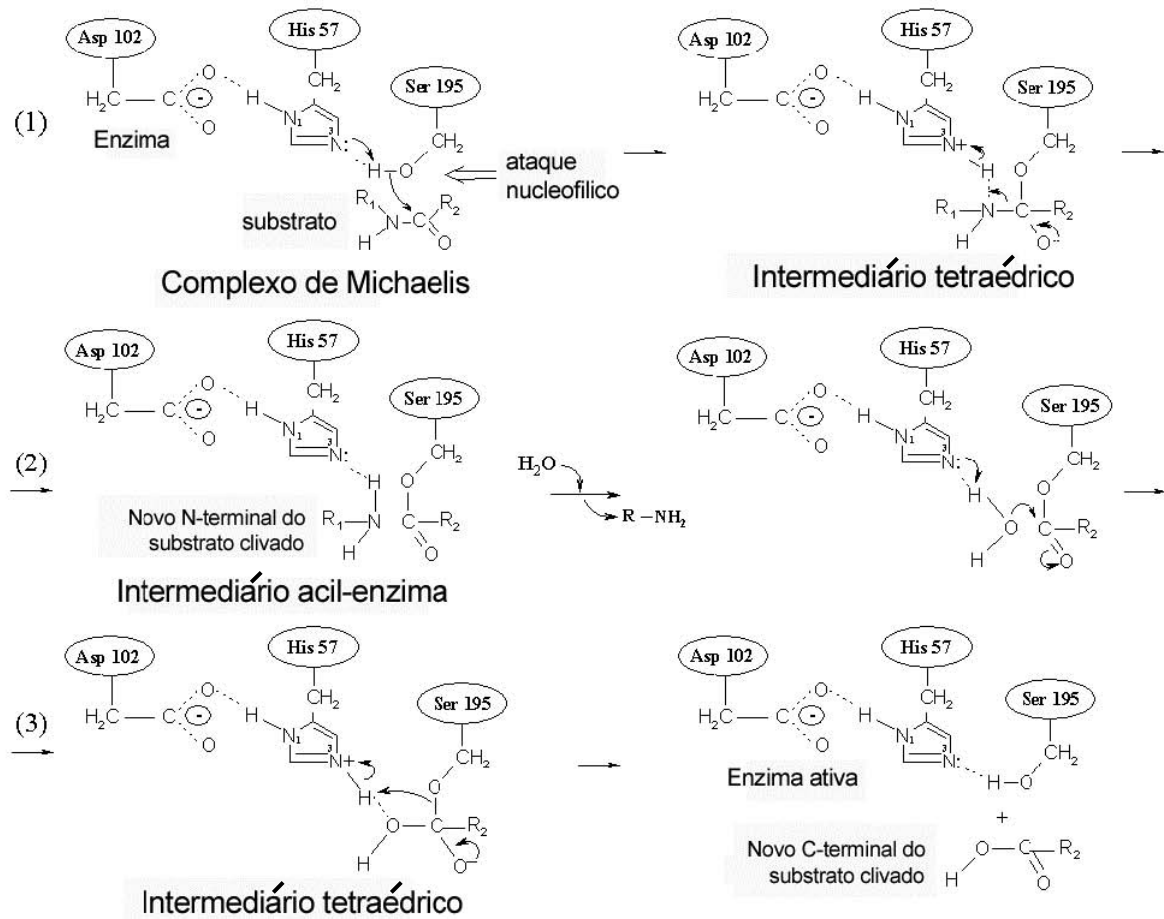


**Figura 3.** Arranjo característico das serino proteases. Os resíduos da tríade catalítica Asp102-Ser195-His57 estão representados pela cor azul. Tripsina (A) PDB 1FY8 e a subtilisina (B) PDB 1S01. Note como as estruturas são diferentes tridimensionalmente: quimotripsinas possui uma maior quantidade de folhas- $\beta$ s, enquanto as subtilisinas maior quantidades de  $\alpha$ -hélices. Na tripsina (A) mostramos também representado em azul o Asp189 (mais a direita) que confere a especificidade.

Três resíduos, os quais formam a chamada tríade catalítica, são essenciais no processo catalítico: His 57, Asp 102 e Ser 195 (seguindo a numeração do quimotripsinogênio, e da Tripsina) (Blow & Steitz, 1970; Jeremy *et al.*, 2002).

O centro ativo da quimotripsina é marcado pela serina 195. A cadeia lateral da serina 195 forma ponte de hidrogênio com o anel imidazólico da histidina 57. O grupamento -NH deste anel, por sua vez, forma ponte de hidrogênio (ou ligação de hidrogênio) com o grupo carboxilato do aspartato 102. Esta constelação de radicais é denominada de *tríade catalítica*. Este arranjo estrutural leva à alta reatividade da serina 195. A histidina serve para posicionar a cadeia lateral da serina e para polarizar a sua hidroxila. Ao fazer isto, age como catalisador básico geral como acceptor de íon de hidrogênio, porque a hidroxila polarizada da serina fica pronta para desprotonação. A retirada do próton da hidroxila gera um íon alcóxido, que é um nucleófilo muito mais poderoso do que um álcool. O aspartato ajuda a orientar a histidina e tornar a histidina um melhor acceptor de prótons, por efeitos eletrostáticos (Jeremy *et al.*, 2002).

Estas observações sugerem um mecanismo para uma hidrólise de peptídeos. Após a ligação ao substrato (etapa 1 da Figura 4), a reação começa com o ataque nucleofílico da hidroxila da serina 195 ao carbono da carbonila do substrato. Este ataque altera a geometria em torno deste carbono, de triangular plana para tetraédrica. O *intermediário tetraédrico* inerentemente instável formado é portador de uma carga negativa no átomo de oxigênio derivado da carbonila. Esta carga é estabilizada por interações com grupamentos NH da proteína em um local denominado *cavidade do oxianionte*. Estas interações também ajudam a estabilizar o estado de transição que precede a formação do intermediário tetraédrico. Este último intermediário sofre então um colapso, gerando acil-enzima (etapa 2 da Figura 4). Esta etapa é facilitada pela transferência de um próton da histidina, com carga positiva, para amina formada pela clivagem da ligação peptídica. O componente amínico está agora livre para sair da enzima e é então hidrolisado por um processo que é essencialmente as etapas anteriores. Uma molécula de água ataca a carbonila, enquanto um próton é concomitantemente removido pelo radical da histidina, que agora age como catalisador ácido geral, formando um intermediário tetraédrico (etapa 3 da Figura 4). Esta estrutura se quebra, formando o ácido carboxílico produto que deixa pronta a enzima para outra rodada de catálise (Jeremy *et al.*, 2002). A seguir é mostrada a representação esquemática do mecanismo catalítico das serino proteases.



**Figura 4.** Esquema do mecanismo catalítico das serino proteases (Jeremy *et al.*, 2002).

Na Tabela 1 temos a classificação atual das subfamílias de serino proteases retirado do banco de dados de peptidases e inibidores Merops (Rawlings *et al.*, 2008).

**Tabela 1.** Classificação atual das serino proteases. (Fonte: <http://merops.sanger.ac.uk/>).

FAMÍLIA	SUBFAMÍLIA	TIPO DE PEPTIDASE
S1	S1A	chymotrypsin A ( <i>Bos taurus</i> )
	S1B	glutamyl peptidase I ( <i>Staphylococcus aureus</i> )
	S1C	peptidase Do ( <i>Escherichia coli</i> )
	S1D	lysyl peptidase ( <i>Achromobacter lyticus</i> )
	S1E	Streptogrisin A ( <i>Streptomyces griseus</i> )
	S1F	astrovirus serine peptidase ("Human astrovirus")
S3		togavirin ("Sindbis virus")
S6		IgA1-specific serine peptidase ( <i>Neisseria gonorrhoeae</i> )
S7		flavivirin ("yellow fever virus")
S8	S8A	subtilisin Carlsberg ( <i>Bacillus licheniformis</i> )
	S8B	kexin ( <i>Saccharomyces cerevisiae</i> )
S9	S9A	prolyl oligopeptidase ( <i>Sus scrofa</i> )
	S9B	dipeptidyl-peptidase IV ( <i>Homo sapiens</i> )
	S9C	acylaminoacyl-peptidase ( <i>Homo sapiens</i> )
	S9D	glutamyl peptidase ( <i>Arabidopsis thaliana</i> )
S10		carboxypeptidase Y ( <i>Saccharomyces cerevisiae</i> )
S11		D-Ala-D-Ala carboxypeptidase A ( <i>Geobacillus stearothermophilus</i> )
S12		D-Ala-D-Ala carboxypeptidase B ( <i>Streptomyces lividans</i> )
S13		D-Ala-D-Ala peptidase C ( <i>Escherichia coli</i> )
S14		peptidase Clp ( <i>Escherichia coli</i> )
S15		Xaa-Pro dipeptidyl-peptidase ( <i>Lactococcus lactis</i> )
S16		Lon-A peptidase ( <i>Escherichia coli</i> )
S21		cytomegalovirus assemblin ( <i>human herpesvirus 5</i> )
S24		repressor LexA ( <i>Escherichia coli</i> )
S26	S26A	signal peptidase I ( <i>Escherichia coli</i> )
	S26B	signalase ( <i>Saccharomyces cerevisiae</i> )
	S26C	TraF peptidase ( <i>Escherichia coli</i> )
S28		lysosomal Pro-Xaa carboxypeptidase ( <i>Homo sapiens</i> )
S29		hepacivirin ("hepatitis C virus")
S30		potyvirus P1 peptidase ("plum pox virus")
S31		pestivirus NS3 polyprotein peptidase ("bovine viral diarrhea virus 1")
S32		equine arteritis virus serine peptidase ("equine arteritis virus")
S33		prolyl aminopeptidase ( <i>Neisseria gonorrhoeae</i> )
S37		PS-10 peptidase ( <i>Streptomyces lividans</i> )

S39	S39A	Sobemovirus peptidase ("cocksfoot mottle virus")
	S39B	luteovirus peptidase ("potato leaf roll luteovirus")
S41	S41A	C-terminal processing peptidase-1 ( <i>Escherichia coli</i> )
	S41B	tricorn core peptidase ( <i>Thermoplasma acidophilum</i> )
S45		penicillin G acylase precursor ( <i>Escherichia coli</i> )
S46		dipeptidyl-peptidase 7 ( <i>Porphyromonas gingivalis</i> )
S48		HetR peptidase ( <i>Anabaena variabilis</i> )
S49		signal peptide peptidase A ( <i>Escherichia coli</i> )
S50		infectious pancreatic necrosis birnavirus Vp4 peptidase (infectious pancreatic necrosis virus)
S51		dipeptidase E ( <i>Escherichia coli</i> )
S53		sedolisin ( <i>Pseudomonas sp. 101</i> )
S54		Rhomboid-1 ( <i>Drosophila melanogaster</i> )
S55		SpoIVB peptidase ( <i>Bacillus subtilis</i> )
S58		aminopeptidase DmpA ( <i>Ochrobactrum anthropi</i> )
S59		nucleoporin 145 ( <i>Homo sapiens</i> )
S60		lactoferrin ( <i>Homo sapiens</i> )
S62		influenza A PA peptidase ("influenza A virus")
S63		EGF-like module containing mucin-like hormone receptor-like 2 ( <i>Homo sapiens</i> )
S64		Ssy5 peptidase ( <i>Saccharomyces cerevisiae</i> )
S66		murein tetrapeptidase LD-carboxypeptidase ( <i>Pseudomonas aeruginosa</i> )
S68		PIDD auto-processing protein unit 1 ( <i>Homo sapiens</i> )
S69		Tellina virus 1 VP4 peptidase ("Tellina virus 1")
S71		MUC1 self-cleaving mucin ( <i>Homo sapiens</i> )
S72		Dystroglycan ( <i>Homo sapiens</i> )

## 2. Os Inibidores de Serino Proteases

As proteases desempenham muitas funções importantes no organismo, mas elas podem ser muito perigosas, por isso a sua atividade é rigidamente controlada para que elas sejam eficientes. Ter o sangue coagulado da cabeça aos pés por causa de um simples corte ou digerir o próprio pâncreas podem ser exemplos de uma proteólise não controlada (Laskowski & Qasim, 2000). Naturalmente para este controle foram desenvolvidos muitos mecanismos, mas dois predominam (Laskowski & Kato, 1980).

O primeiro deles é que quase todas as proteases são biosintetizadas como precursores inativos chamados pró-proteínas ou zimogênios. Esses são armazenados

e então ativados quando necessário. A ativação envolve proteólise com a remoção de uma ou várias ligações peptídicas da porção N-terminal em direção ao sítio catalítico da enzima. Pelo fato das proteínas serem sintetizadas a partir do N-terminal para o C-terminal, o produto parcialmente sintetizado nunca será ativo. Outro mecanismo é necessário para controlar a enzima uma vez ativada. Isso é providenciado pela presença de proteínas conhecidas como inibidores de proteases. Esses inibidores formam complexos completamente inativos ou parcialmente ativos com suas enzimas cognatas (Laskowski & Kato, 1980).

O termo *inibidor de proteases* poderia a princípio denotar qualquer proteína ou molécula que diminui a atividade enzimática de uma protease. Isso poderia fazer alguma confusão como, por exemplo, as proteases que “inibem” outras proteases digerindo-as, mas elas não são consideradas inibidores de proteinases. Entretanto dois grupos de proteínas completamente dissimilares são chamados inibidores de proteinases. O primeiro deles consiste das macroglobulinas que são proteínas de alto peso molecular aparentemente presente no plasma de todos os mamíferos, frequentemente em múltiplas formas, e que não seguem um mecanismo padrão de associação com proteases, não sendo, portanto, específicos para uma determinada classe de proteases. O segundo grupo é composto pela classe de inibidores específicos (Laskowski & Kato, 1980).

## **2.1. Inibidores do Plasma Sanguíneo**

Muitos processos proteolíticos como a coagulação sanguínea, dissolução de coágulos de sangue e formação e destruição de hormônios ocorrem no plasma sanguíneo e requerem controle acurado. Por isso, não é surpreendente que o plasma sanguíneo contenha muitos tipos de inibidores de proteases, alguns em concentração bastante elevadas, como mostra a Tabela 2 (Laskowski & Kato, 1980).

**Tabela 2.** Principais inibidores de proteases presentes no plasma sanguíneo humano.

Inibidor	Concentração (mg/100mL)	PM (kDa)	Cadeias polipepticas	Sítios Inibitórios
$\alpha_1$ -proteinase	290 $\pm$ 45	52	1	1
A <sub>1</sub> -antiquimotripsina	49 $\pm$ 7	69	1	1
C1-inativador	24 $\pm$ 3	70	1	1
$\alpha_2$ -antiplasmina	7 $\pm$ 1	70	1	1
Antitrombina	24 $\pm$ 2	65	1	1
Inter- $\alpha$ -tripsina	50	160	1	1
$\alpha_2$ -macroglobulina	260 $\pm$ 70	720	4	1-2 ou mais

Dos inibidores listados na Tabela 2, a cadeia do inibidor inter- $\alpha$ -tripsina termina em dois domínios pertencentes à família BPTI-Kunitz e o sítio reativo se encontra nessa porção da cadeia. Por esse aspecto ele é discutido detalhadamente em tópico separado sobre as famílias dos inibidores canônicos de serino proteases. A  $\alpha_2$ -macroglobulina e os demais inibidores que pertencem à família conhecidas como serpinas (do inglês *serine proteinases inhibitors*) serão discutidos a seguir.

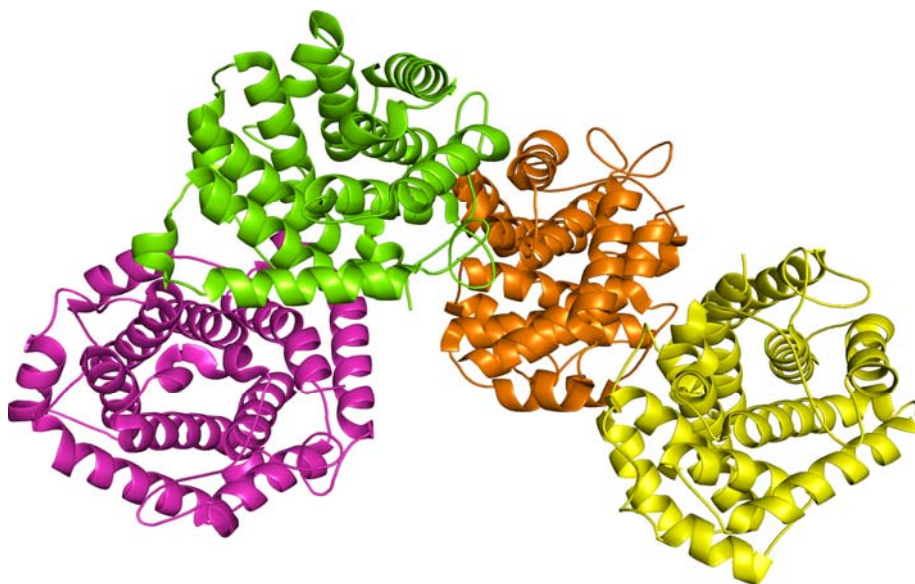
### 2.1.1. As Macroglobulinas

As macroglobulinas são exemplificadas pela  $\alpha_2$ -macroglobulina humana ( $\alpha_2M$ ), mostrada na Figura 5. Ela tem um peso molecular de 729kDa e é composta de quatro cadeias polipeptídicas idênticas. Ela dissocia em pares de cadeias após a desnaturação e em cadeias simples após redução das pontes de dissulfeto (Sottrup-Jensen *et al.*, 1984).

Quando  $\alpha_2M$  se combina com proteases, apenas a atividade proteolítica sobre substratos protéicos grandes é diminuída ou eliminada. A atividade sobre substratos específicos e pequenos substratos sintéticos não é interrompida. Os complexos  $\alpha_2M$ -proteases formados podem ainda ser inibidos por pequenos inibidores de proteases, como BPTI-Kunitz ou PSTI-Kazal, por exemplo, mas não por inibidores maiores como STI-Kunitz, por exemplo. Isso mostra claramente que nos complexos com  $\alpha_2M$

o sítio ativo das proteases permanece aberto, apenas o acesso de substratos e inibidores a esse sítio é estericamente impedido. Esses fatos enfatizam o mecanismo de aprisionamento pelo qual as macroglobulinas agem. A protease hidrolisa uma ou mais ligações peptídicas particularmente susceptíveis da  $\alpha_2M$  o que induz uma mudança conformacional da  $\alpha_2M$  a qual aprisiona a molécula da enzima (Sottrup-Jensen, 1989).

As macroglobulinas são de central importância no processamento de proteases no sangue. Entretanto a sua interação com proteases é tão diferente que pelo menos do ponto de vista químico, a sua classificação como inibidor parece inadequada (Sottrup-Jensen, 1989).



**Figura 5.** Estrutura 3D da  $\alpha$ -2 macroglobulina (código PDB 1QSJ) mostrando as quatro cadeias idênticas que compõem a molécula.

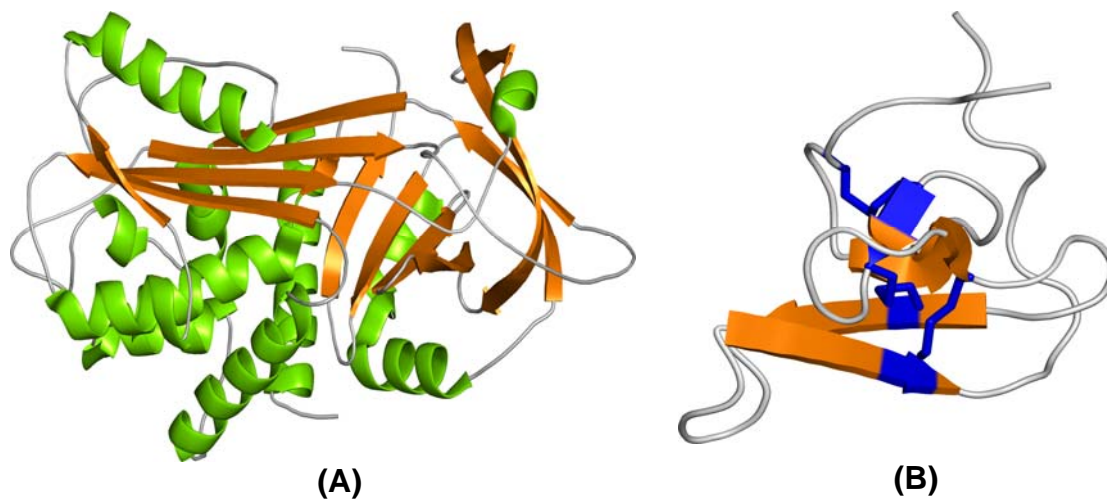
### 2.1.2. As Serpinas

Os inibidores chamados serpinas formam uma família de grandes glicoproteínas homólogas. Assim como inibidores canônicos, as serpinas parecem interagir com suas proteases cognatas via um *loop* de ligação exposto. Os complexos



resultantes são, entretanto, apenas transientes e colapsam com a liberação de uma forma clivada de diferente estrutura e estabilidade. As serpinas exibem atividade inibitória contra uma ampla variedade de serino proteases. As serpinas são constituídas de dois domínios, um predominantemente alfa e outro beta como representado na Figura 6A (Carrell & Travis, 1985).

Com a determinação da sua estrutura 3D verificou-se que os inibidores serpinas possuem dois sítios de interação, sendo um deles responsável pelo reconhecimento da proteinase e outro contendo o *loop* reativo no qual ocorre à formação do complexo enzima/inibidor propriamente dito (Loebermann *et al.*, 1984).



**Figura 6.** (A) estrutura tridimensional da serpina  $\alpha$ 1-antitripsina humana (PDB 1QLP), (B) estrutura tridimensional da hirudina extraída de *Hirudo medicinalis* (PDB 2HIR). As pontes dissulfeto estão representadas em azul.

### 2.1.3. As Hirudinas

Uma outra família de inibidores que não segue o mecanismo padrão de associação descrito para os inibidores canônicos é a das hirudinas conhecidos também como antitrombina (Rydel *et al.*, 1991).

A saliva das sanguessugas contém uma variedade de inibidores de proteases. A fim de manter o estado líquido do sangue ingerido, a maioria desses inibidores é dirigida contra proteases que estão envolvidas na coagulação sangüínea. A hirudina purificada da saliva da sanguessuga *Hirudo medicinalis* é um exemplo desses inibidores, sendo o mais potente inibidor natural de trombina conhecido. A constante de inibição de trombina por hirudina é da ordem de  $10^{-14}\text{M}$ , sendo o melhor  $K_i$  descrito até agora. Entretanto, como mencionado, a hirudina se liga à trombina seguindo um mecanismo de interação não canônico diferente do mecanismo padrão (Rydel *et al.*, 1990). A família das hirudinas é composta por inibidores homólogos de cerca de 7kDa com uma cadeia polipeptídica estabilizada por três pontes dissulfeto altamente conservadas. A estrutura desses inibidores está organizada em dois domínios, um na porção N-terminal (resíduos 3-48) formando um arranjo rígido estabilizado pelas pontes de dissulfeto, e outro na porção C-terminal (resíduos 49-65) formando uma cauda flexível (Rydel *et al.*, 1990; Rydel *et al.*, 1991).

Esse grupo de inibidores é atualmente produzido sinteticamente por várias empresas farmacêuticas e são usados na profilaxia e no tratamento de trombose e de outros distúrbios (Fenton *et al.*, 1998; Folkers *et al.*, 1989).

A estrutura das hirudinas consiste de um domínio globular N-terminal e um domínio C-terminal estendido (Folkers *et al.*, 1989), como mostrado na Figura 6B.

## 2.2. Inibidores com Sítio Ativo Classe-Específicos

Existem outros tipos de inibidores que diferem bastante das macroglobulinas. A maioria desses inibidores pode ser enquadrado nas três afirmações a seguir:

1. Nos complexos formados com as enzimas a atividade enzimática sobre substratos é totalmente abolida.
2. A inibição é estritamente competitiva.
3. Um sítio ativo inibitório particular pode inibir apenas proteases pertencentes a uma das quatro classes mecanísticas (Laskowski & Kato, 1980).

Os inibidores são divididos em quatro classes de acordo com a proteínase inibida. São eles os inibidores de proteases aspárticas, inibidores de proteínase cisteínicas, inibidores de metaloproteases e os inibidores de serino proteases (Laskowski & Kato, 1980). Apesar do enorme crescimento de informação sobre os três primeiros, a quantidade de informações sobre serino proteases ainda é muito maior. Portanto iremos concentrar os nossos estudos em serino proteases, que possui uma grande quantidade de estruturas determinadas por cristalografia, o que será de extrema importância na nossa pesquisa.

## 2.3. Os Inibidores Canônicos de Serino Proteases

O número de inibidores bem caracterizados de serino proteases excede o número de inibidores descritos para as outras três classes mecanísticas (Hedstrom, 2002) (Berman *et al.*, 2000). Entretanto, não está claro se esta abundância relativa reflete na verdade a distribuição de inibidores na natureza, ou somente a conveniência e interesse bioquímico em isolá-las, por que atualmente existem dentro do PDB mais inibidores de serino proteases do que das outras classes mecanísticas existentes.

Em contraste com outros inibidores, o mecanismo de interação dos inibidores de serino proteases com suas enzimas cognatas é consideravelmente bem detalhado.

Estudos de sequenciamento e cristalografia de raios-X têm mostrado que esses inibidores não são todos homólogos, consistindo de muitas (pelo menos 18) famílias (Laskowski & Kato, 1980). Surpreendentemente, entretanto, a grande maioria dos inibidores de serino proteases interage seguindo um “mecanismo padrão” que será descrito. Esta situação é análoga a existência de pelo menos duas famílias de serino proteases não homólogas, a família da subtilisina e a da quimotripsina, as quais não compartilham uma estrutura tridimensional comum, mas hidrolisam seus substratos e são inibidas pelos seus inibidores pelo mesmo mecanismo (Laskowski & Kato, 1980) (Hedstrom, 2002).

Existem apenas três tipos de famílias de inibidores de serino proteases que não seguem o mecanismo padrão, as macroglobulinas, as serpinas e as hirudinas, todas elas são encontradas no plasma sanguíneo (Laskowski & Kato, 1980).

## **2.4. O Início dos Estudos Sobre Inibidores**

Em 1936, o interesse em inibidores de proteases tomou um impulso quando Kunitz e Northrop começaram a estudar o inibidor de tripsina bovina extraído do pâncreas bovino, que recebeu o nome de BPTI - *Bovine Pancreatic Trypsin Inhibitor*. E em 1945, começou também a estudar outro inibidor de tripsina extraído da soja, que recebeu o nome de STI - *Soybean Trypsin Inhibitor*. Os experimentos de Kunitz abriram portas para esse tipo de pesquisa e deixaram uma série de definições e padronizações na pesquisa sobre inibidores (Kunitz & Northrop, 1936).

Surgiram então dois grandes problemas do legado deixado por Kunitz:

1. Que a maioria esmagadora de inibidores na natureza eram de proteases; foram testados inúmeros materiais biológicos quanto à inibição em tripsinas, o que resultou numa maior descoberta de inibidores de tripsinas.
2. Que tanto a tripsina quanto os seus muitos inibidores eram verdadeiras “caixas pretas”.

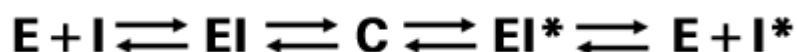
Este legado foi piorando, pois enquanto os pesquisadores ficavam intrigados em responder como a tripsina se combina com seus inibidores, os estudos de modificações químicas e dicróísmo circular davam resultados complexos (Laskowski Jr *et al.*, 2000).

Em 1945, no laboratório de Nord foi feita uma sugestão correta que os inibidores são análogos aos substratos das proteases e que os complexos enzimas/inibidor formados seriam especialmente estáveis. O desenvolvimento de tecnologia de seqüenciamento de proteínas levou Kassel e Laskowski a seqüenciar a BPTI. Logo em seguida foi feito o seqüenciamento do inibidor de calicreína bovina. Após correção de pequenos erros na sequência do inibidor de calicreína, foi percebido que os dois eram na verdade a mesma proteína (Laskowski Jr *et al.*, 2000).

A estrutura da BPTI foi determinada pouco tempo depois da cristalografia de proteínas ser possível. Uma vez conhecida a estrutura tridimensional do inibidor livre, ainda se tornava difícil desvendar o mecanismo de combinação. Isso foi amenizado mais tarde quando as estruturas tridimensionais dos complexos de tripsina-inibidor foram determinadas (Kunitz & Northrop, 1935, 1936).

## **2.5. Mecanismo de Ação dos Inibidores**

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



**Figura 7.** Reação que ilustra o mecanismo geral seguindo pelos inibidores canônicos. Onde E representa a enzima, I o inibidor, EI enzima e inibidor, C o complexo, EI\* a Enzima mais o inibidor modificado e I\* o inibidor modificado.

A presença de um resíduo de Arg na posição P1 é consistente com a especificidade da tripsina, mas podemos observar que, dentre os muitos resíduos de Lys e Arg em STI, apenas um, Arg63, é consistentemente selecionado. Em muitos inibidores de tripsina, STI por exemplo, o resíduo na posição P1 é uma Arg, enquanto em muitos outros, por exemplo BPTI, é uma Lys. Isso pode explicar os resultados aparentemente anômalos previamente obtidos por uma modificação química que tanto intrigaram os pesquisadores: a acetilação dos inibidores cujo resíduo é Lys extinguiu completamente sua atividade inibitória enquanto a acetilação dos inibidores cujo resíduo é Arg não tinha qualquer efeito sobre sua atividade. Isso acontece porque o grupo NH<sub>2</sub> da Lys é bloqueado por acetilação, o que impede a interação iônica com o Asp 102 da tripsina necessária para associação enzima-inibidor e, no caso da Arg, apenas um dos grupos NH<sub>2</sub> é bloqueado, permanecendo o outro livre para a interação. Isso foi endossado por experimentos de mutação enzimática nos quais o resíduo Lys na posição P1 do inibidor era substituído por uma Arg, fazendo com que a atividade sobre a tripsina não fosse alterada. Essa talvez tenha sido a primeira mutagênese sítio específica realizada com serino protease (Laskowski Jr *et al.*, 2000).

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

A conformação canônica do *loop* reativo é muito importante para um posicionamento adequado e uma adaptação efetiva ao sítio ativo da enzima proteolítica, o que resulta na formação do intermediário acil-enzima. O resíduo P1 desempenha um papel chave durante o processo, definindo a especificidade do inibidor. Por exemplo, a presença de uma Arg ou Lys na posição P1 permite uma interação iônica com Asp189 no sítio de ligação da tripsina. A geometria do grupo carbonila do resíduo na posição P1 define a orientação adequada para o ataque nucleofílico da serina catalítica, a qual primeiro forma o acil-enzima e então faz o ataque nucleofílico ao grupo abandonador, favorecendo a religação. Além disso, o posicionamento e a orientação do novo N-terminal do inibidor clivado são também cruciais para um ataque nucleofílico efetivo. Isso é providenciado por uma extensa rede de interações intra e intermoleculares entre inibidor e enzima (Laskowski Jr *et al.*, 2000).

## 2.6. Proteólise Limitada

O funcionamento e a compreensão da interação da protease com o inibidor tornou-se possível através da determinação das estruturas dos seus complexos. Por exemplo, porque os inibidores não são completamente clivados na presença da protease como a maioria das proteínas? Tripsina por exemplo deveria clivar cada uma das ligações Lisina-X e Arginina-X (com exceção de Prolina em X) e nesse caso, cerca de 5 a 10% das ligações peptídicas em uma proteína seria suscetível ao ataque proteolítico. Todavia, os inibidores de proteases fornecem um paradigma uma vez que apenas a ligação peptídica “scissile” P1-P1’ no *loop* reativo é proteoliticamente atacada, resultando na formação do complexo enzima-inibidor no qual o *loop* reativo do inibidor está ligado ao sítio ativo da enzima numa formação de “substrato perfeito”, enquanto, baseando-se em sua sequência, deveria ser esperado que eles fossem rapidamente proteolisados em muitos outros sítios de clivagem (Hubbard *et al.*, 1991)

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

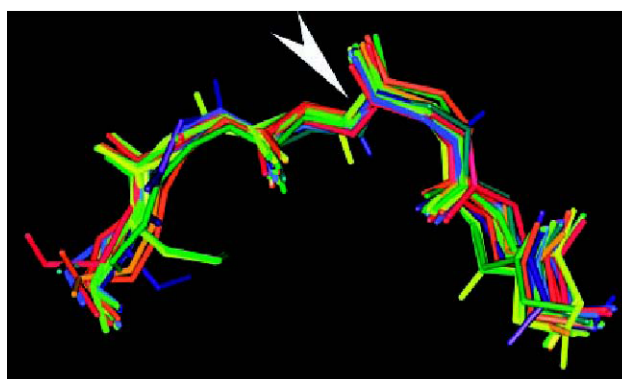
## **2.7. Loop Reativo Canônico**

Estudos estruturais de complexos enzima-inibidor confirmaram e estenderam as primeiras observações de que todos os inibidores têm um grande *loop* exposto ao redor do resíduo P1. O resíduo P1 é bem exposto nos inibidores livres. Esse *loop* atua como uma das duas fitas de uma folha- $\beta$  distorcida que é formada nos complexos com enzimas da família da quimotripsina. A segunda fita é fornecida pela enzima. Na interação com enzimas da família da subtilisina, a enzima contribui com duas fitas e a folha- $\beta$  distorcida resultante é composta de três fitas. Além dos resíduos próximos à ligação peptídica do sítio ativo que compõem o *loop* reativo, outros



resíduos de diferentes partes do inibidor também fazem contatos com a enzima tanto com a cadeia principal quanto com as cadeias laterais. Embora essas interações não sejam tão conservadas em todos os inibidores quanto aquelas feitas pelos resíduos do *loop* reativo, elas são compartilhadas por inibidores dentro das diferentes famílias.

Os resíduos de P4 a P3' de todos os inibidores exibem os mesmos ângulos de Ramachandram (Laskowski Jr *et al.*, 2000), como mostrado na Figura 8. Apesar do *loop* reativo dos inibidores da maioria das 16 famílias de inibidores serem tão próximos, é espantoso observar que cada uma dessas 16 famílias possuem estruturas tridimensionais globais bem diferentes, mas possuem similaridade onde realmente elas são importantes para a atividade inibitória.



**Figura 8.** Superposição dos átomos das cadeias principais dos 6 resíduos de aminoácidos do *loop* reativo. As estruturas são de 12 diferentes famílias de inibidores não relacionadas em sequência e em arranjo, em complexo com 9 serino proteases incluindo membros tanto da subfamília da quimotripsina quanto da subtilisina. A seta branca indica o sítio de clivagem entre P1-P1'.

## 2.8. Famílias de Inibidores de Serino Proteases

Existem pelo menos 18 diferentes famílias de inibidores de mecanismo padrão de serino proteases reconhecidas até agora, mas somente 12 famílias possuem complexos cristalizados (Laskowski Jr *et al.*, 2000; Laskowski & Kato, 1980). É muito provável que muitas novas famílias sejam descritas no futuro como também muitos inibidores novos que estão sendo descobertos. As principais famílias de inibidores de serino proteases são listadas na Tabela 3. Tentou-se dividir essas famílias em

inibidores de origem animal, de plantas e de microorganismos e apesar de se ter conseguido bastante êxito, essa divisão não é perfeita (Laskowski Jr *et al.*, 2000). Uma exceção notável é a *eglin c*, a qual é um inibidor de origem animal extraído de sanguessuga *Hirudo medicinalis* e é um membro da família Batata I. Existem muitas outras exceções. Nessa tese iremos focar os nossos estudos em três inibidores: BPTI, kazal e ecotina. Para cada uma dessas famílias é apresentado a seguir um pequeno histórico bem como uma descrição das características e peculiaridades de seus membros. Para as outras famílias será feita apenas uma visão geral (Berman *et al.*, 2000) (Laskowski Jr *et al.*, 2000).

**Tabela 3.** As famílias de inibidores de serino proteases com complexos determinados por cristalografia de raio-X

FAMÍLIA	MONÔMERO		ESTRUTURA 3D (exemplo do PDB)	
	MMr (kDa)	½ CYS	LIVRE	COMPLEXO
BPTI Kunitz <sup>a</sup>	6	6	1BPI	2PTC
STI Kunitz <sup>b</sup>	21-22	4	1AVU	1AVW
BBI <sup>b</sup>	8-9	14	1PI2	1SMF
Batata I <sup>b</sup>	8-9	0-2	1MIT	1ACB
Batata II <sup>b</sup>	6	8	1TIH	1ACB
Abóbora <sup>b</sup>	3	6	2CTI	1PPE
Kazal <sup>a</sup>	6	6	2OVO	1CHO
SSI <sup>c</sup>	12-14	2-4	2SSI	2SIC
Antistasina <sup>a</sup>	8-10	20	1SKZ	1HIA
Chelonianina <sup>a</sup>	6	6	2REL	1FLE
Ascaris <sup>a</sup>	6	10	1ATA	1EAI
Ecotina <sup>c</sup>	12-14	2	1ECY	1AZZ

Inibidores de origem <sup>a</sup>animal, de <sup>b</sup>plantas e de <sup>c</sup>microorganismos.

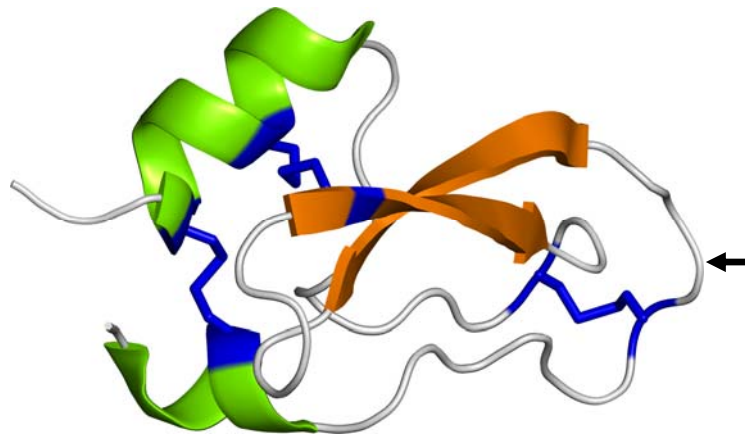
### 2.8.1. Família BPTI Kunitz

A família BPTI Kunitz é composta de pequenas proteínas com massa molecular em torno de 6kDa (cerca de 55–60 aminoácidos) com representantes em espécies de vertebrados e invertebrados, incluindo humanos, bovinos, ratos, camundongos, peixes, cavalos, sapos, vermes e bicho-da-seda. Apesar dos membros da família BPTI Kunitz compartilharem um alto grau de similaridade seqüencial e estrutural, o que faz com que todos apresentem o mesmo arranjo global característico, eles exibem uma faixa enorme de funcionalidades e especificidades (Kunitz & Northrop, 1936).

O arranjo do tipo BPTI Kunitz é altamente adaptável e tem sido encontrado em proteínas que desempenham muitas funções diferentes na natureza, desde a inibição de proteases até poderosas neurotoxinas (Marshall & Harvey, 1992; Skarzynski, 1992). Dentre as proteínas que possuem domínios BPTI Kunitz estão às chamadas bikuninas, que compreendem os inter- $\alpha$  inibidores de tripsina (I $\alpha$ I) que possuem o domínio BPTI Kunitz que é a parte inibitória do I $\alpha$ I. As bikuninas possuem dois domínios tipo BPTI Kunitz: o domínio N-terminal (domínio 1) tem um sítio de ligação neutro e específico para elastases enquanto o domínio C-terminal (domínio 2) tem um sítio de ligação básico sendo específico para as serino proteases tripsina e plasmina. Um impedimento estérico que um domínio impõe ao outro previne que ambos se liguem simultaneamente a enzimas e também restringe o tamanho do alvo que será ligado. Por essa versatilidade na inibição, as bikuninas são atualmente prescritas no Japão para o tratamento de pancreatite aguda e choque hemorrágico. Sabe-se também que as bikuninas estão envolvidas no controle do crescimento celular e os I $\alpha$ I participam em processos na matriz extracelular como morfogênese e diferenciação de tecidos (Itoh *et al.*, 1994; Lindqvist & Akerstrom, 1996).

O arranjo tridimensional da estrutura dos membros da família BPTI Kunitz, mostrado na Figura 9, é composto por uma folha- $\beta$  antiparalela central a qual é

fechada em um dos lados por uma volta- $\beta$  formando um grampo- $\beta$  e no outro lado por uma ponte dissulfeto. A seção da cadeia próxima ao N-terminal forma uma  $\alpha$ -hélice curta e algumas vezes distorcida, enquanto a seção próxima ao C-terminal forma uma  $\alpha$ -hélice mais regular. A  $\alpha$ -hélice C-terminal é covalentemente ligada tanto à folha- $\beta$  central quanto à  $\alpha$ -hélice N-terminal por pontes dissulfeto em pH fisiológico. As três pontes dissulfeto na estrutura são altamente conservadas, provavelmente para manter e reforçar a rigidez e estabilidade estrutural, o que é confirmado pela sua resistência extremamente forte à desnaturação por temperatura - temperatura de desnaturação do BPTI é de 95°C (Bode & Huber, 1992). Quando essas três pontes dissulfeto são reduzidas, a proteína perde o seu arranjo estrutural e fica inativa.



**Figura 9.** Estrutura tridimensional de BPTI extraída de *Bos taurus* (Boi, PDB 1BPI). As pontes dissulfeto são representadas em azul e o *loop* reativo é indicado pela seta.

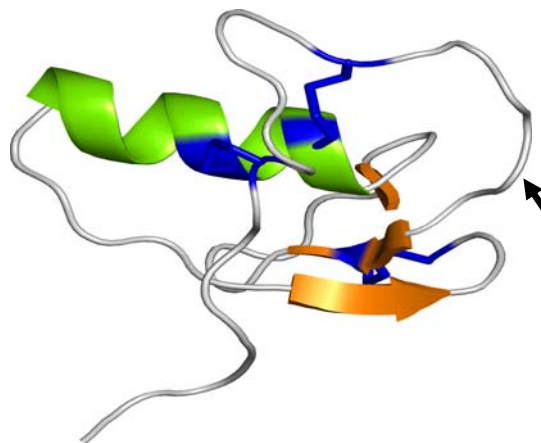
**Tabela 4.** Membros da Família Kunitz BPTI, inibidores de serino proteases.

Inibidor	Atividade Inibitória Principal	Outras Funções Conhecidas
BPTI	Forte inibidor de tripsina , calicreína quimotripsina, matriptase MTSP1 e fator de coagulação VII. Também inibe muitas outras serino proteases.	BPTI e seus fragmentos 1-15, 18-39, e 40-58 possuem atividade bactericida contra muitas bactérias gram positivas e gram negativas. Inibe cisteinoproteases de adenovírus com uma constante de ligação de $5 \times 10^5 \text{ M}^{-1}$ e num sítio de inibição diferente do que liga tripsina. Inibe o canal $\text{K}^+/\text{Ca}^{++}$ ativado. A subunidade $\alpha$ do canal $\text{K}_{\text{Ca}}$ contém um domínio inibidor de proteinase com sítio de ligação conservado.
APPI Isoformas da proteína precursora de amilóide (APP) que contém um domínio tipo BPTI-Kunitz.	APPI é um forte inibidor de tripsina e inibe muitas outras serino proteases.	Acredita-se que a função das APPs como receptores nucleares está envolvida na via de sinalização intracelular através da proteína de ligação (Go) de GTP. Na doença de Alzheimer, um fragmento de APP, a proteína $\beta$ amilóide é depositada em vários tecidos cerebrais.
Dendrotoxinas (encontradas no veneno de muitas cobras) Exemplos: toxina I, toxina K, toxina E.	Inibe tripsina e calicreína.	Bloqueiam alguns canais neuronais de $\text{K}^+$ e aumentam a liberação de neurotransmissores.
Bikunina	Inibe tripsina, quimotripsina, PPE e HLE.	Bikunina é co-transportada com $\alpha_1$ microglobulina a qual pertence à superfamília das lipocalinas. Esse precursor passa por um processamento pós-translacional para produzir $\alpha_1$ microglobulina e bikunina. Bikunina então se associa covalentemente com uma ou duas cadeias da família multioxidase formando I $\alpha$ I.
BPTI/Bikunina/Tripstatina	Inibe tripsina, quimotripsina, PPE e HLE	Inibem a ligação da HIV glicoproteína gp120 com a triptase TL.

## 2.8.2. Família Kazal

Essa família foi nomeada depois que L. Kazal descobriu o primeiro dos inibidores de tripsina da secreção pancreática (PSTI - *Pancreatic Secretory Trypsin Inhibitor*). Os inibidores da família Kazal estão presentes em todos os vertebrados, inclusive no homem, e a sua função fisiológica é evitar a ativação prematura de pró-enzimas, inibindo traços de tripsina ativa que possam estar livres no pâncreas. Os inibidores do tipo Kazal inibem uma série de serino proteases, como tripsina e elastase. Mas eles não estão limitados aos vertebrados. Extratos da sanguessuga *Hirudo medicinalis* contém numerosos inibidores de tripsina e um deles mostra pelo menos um domínio do tipo Kazal. Outras proteínas que se destacam são os ovo-inibidores, proteínas abundantes em ovos de aves (Rawlings & Barrett, 1993).

Cada domínio do tipo Kazal é composto por cerca de 56 resíduos de aminoácidos e três pontes dissulfeto. A estrutura do domínio Kazal lembra a estrutura do BPTI, tendo também a forma de pinça com uma haste sendo a  $\alpha$ -hélice e a outra haste sendo a folha- $\beta$  formada pelas três fitas- $\beta$  (Schlott *et al.*, 2002).



**Figura 10.** Estrutura tridimensional do inibidor de tripsina humano do tipo Kazal (PDB 1HPT). As pontes dissulfeto estão representadas em azul e o *loop* reativo é indicado pela seta.

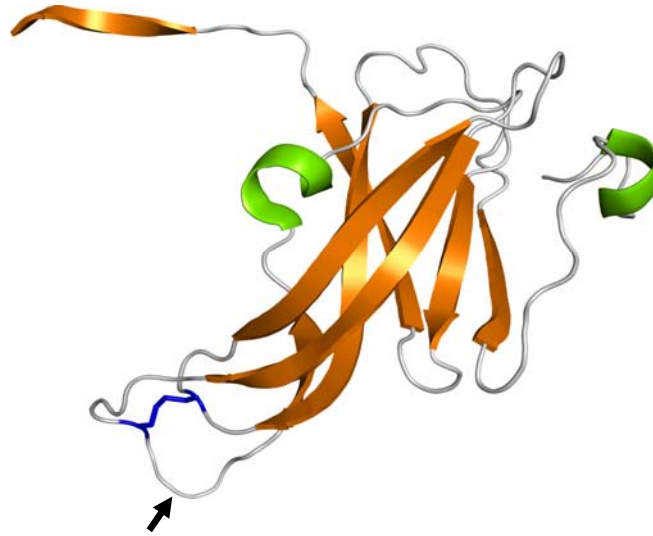
**Tabela 5.** Membros da Família Kazal, inibidores de serino proteases.

Inibidor	Atividade Inibitória Principal	Outras Funções Conhecidas
Ovoinibidor	Inibe uma grande variedade de serino proteases dentre elas podemos citar subtilisina, protease B, elastase, quimotripsina e tripsina tendo vários domínios de inibição.	Inibe o canal ativado por $Ca^{++}/K^+$ .
PEC-60	Inibe tripsina.	É estruturalmente similar ao fator de liberação da colecistokinina. Também é conhecido por suprimir a secreção de insulina induzida por glicose.
Trombospondina 1 (uma glicoproteína com multidomínios)	Contem dois domínios tipo Kazal e inibe HLE e catepsina G.	Acredita-se estar envolvida no desenvolvimento e remodelamento de tecidos.
PSTI II	Inibe tripsina.	Mostra atividade de liberação de colecistokinina.
Rodnina	O primeiro domínio da rodina inibe fortemente trombina.	O segundo domínio da rodina se liga ao sítio da trombina que reconhece e liga fibrinogênio mas não possui atividade inibitória.

### 2.8.3. Família Ecotina

A ecotina é um inibidor de serino protease encontrado no periplasma de *Escherichia coli*. Um dos seus aspectos característicos é que ele é um potente inibidor de uma variedade de serino proteases com especificidades bem diferentes (Chung *et al.*, 1983; McGrath *et al.*, 1995). Os membros da família são proteínas homodiméricas.

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



**Figura 11.** Estrutura tridimensional da ecotina de *Escherichia coli* (Bactéria, PDB 1ECY). A única ponte dissulfeto é representada em azul e o *loop* reativo é indicado pela seta.

A ecotina inibe muitas subfamílias de enzimas, que incluem elastase, quimotripsina, tripsina, fator XA e calicreína. O sítio reativo da ecotina, especificamente o resíduo P1, foi determinado para todos os complexos acima como sendo a Met 84. Bons inibidores de tripsina típicos têm uma Arg ou Lys na posição P1, mas a ecotina inibe a tripsina com uma Met. Assim, a Met emerge com o compromisso de se adequar para satisfazer impedimento estérico de diferentes bolsões. Assim como outros pequenos inibidores de serino proteases, uma ponte de dissulfeto (Cys 50-Cys 87) próxima ao sítio reativo na ecotina estabiliza o *loop* do sítio reativo por uma ligação covalente. Numerosas pontes de hidrogênio também contribuem para dar estabilidade à interação (Yang *et al.*, 1998).

#### **2.8.4. Outras Famílias de Inibidores**

Nesse trabalho nós focamos os nossos estudos em apenas 3 famílias de inibidores, mas existem outras nove, são elas STI Kunitz, Bowman Birk (BBI), Batata I, Batata II, Abobora (ou Cucúrbita), Inibidores de Subtilisinas Streptomyces (SSI), Anistasinas, Chelonianina e Ascaris. Essas famílias são descritas abaixo:



1- O inibidor STI Kunitz foi extraído da soja, possui apenas um sítio de inibição para tripsina, mas existem outros membros que inibem quimotripsina, subtilisinas, catepsinas D e  $\alpha$ -amylase. Os inibidores da família STI Kunitz são encontrados principalmente nas leguminosas (Figura 12) (Ritonja *et al.*, 1996; Song & Suh, 1998).

2- A família BBI (Bowman Birk) do inibidor de protease é amplamente distribuído em sementes de leguminosas, no abacaxi *Ananás sativus*, em germe de trigo *Triticum aestivum*, arroz *Oryza sativa*, sementes de *Coix lachryma jobi*, sementes de capim *Setaria itálica* e cevada *Hordeum vulgare*. Esses inibidores possuem geralmente dois sítios de inibição e são capazes de formar complexos com diferentes proteases – usualmente tripsina e quimotripsina, raramente elastase (Figura 13) (Odani *et al.*, 1986).

3- A família de inibidores do tipo batata I inclui todas aquelas proteínas que mostram uma grande homologia com o inibidor CI-1 de quimotripsina isolado de tubérculos de batata. Esse forte inibidor inibe a quimotripsina, subtilisina e tripsina (Figura 14) (Richardson, 1991).

4- O primeiro membro da família de inibidores do tipo batata II a ter a sua sequência de aminoácidos determinada foi o inibidor de tripsina extraído do exocarpo de berinjela *Solanum melongena* (Figura 15) (Richardson, 1991).

5- Os menores inibidores conhecidos são os inibidores de tripsina isolados das sementes da família Cucurbitae (ou abóbora). Todos contêm cerca de 29 aminoácidos e, apesar de seu pequeno tamanho, são inibidores muito potentes com suas constantes de associação ( $K_a$ ) na faixa de  $5,9 \times 10^{10}$  –  $9,5 \times 10^{11}$  M<sup>-1</sup>, estando entre as maiores já reportadas para inibidores de tripsina. A substituição química da Arg do sítio ativo por Ala converte o inibidor de tripsina num inibidor de elastase (Figura 16) (Richardson, 1991).

6- A família *Streptomyces* de bactérias produz uma série de inibidores, os quais são caracterizados por sua forte atividade contra subtilisina. Eles são coletivamente conhecidos como SSI ou “*Streptomyces Subtilisin Inhibitors*”. Alguns SSI também inibem tripsina e quimotripsina (Figura 17) (Sakai *et al.*, 1980).

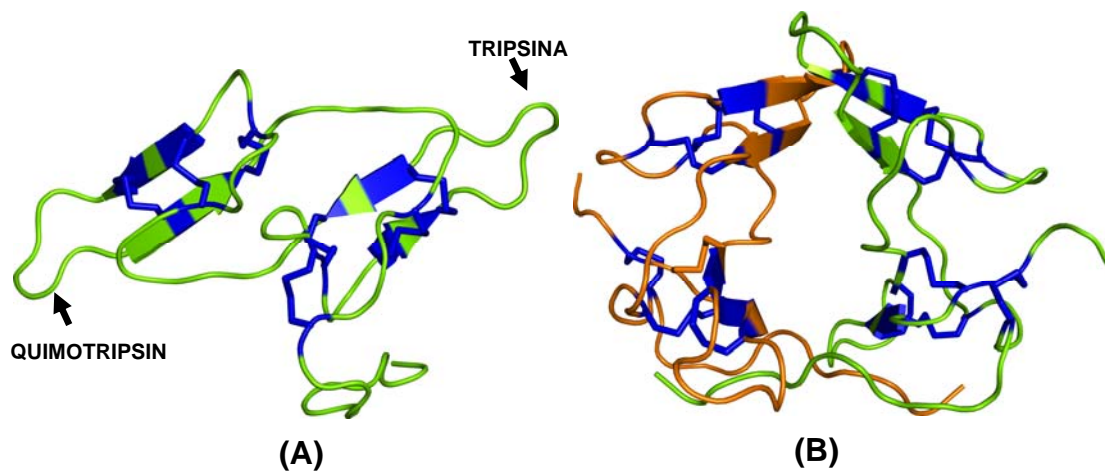
7- Apesar de extremamente pequenos, os inibidores da família das Antistasinas, de tripsina dessa família possuem um conteúdo de cisteínas bastante alto, o que lhes confere uma estrutura rígida e resistente. O padrão de resíduos de cisteínas é bastante conservado, sendo considerado como uma assinatura dessa família de proteínas (Figura 18) (Lapatto *et al.*, 1997).

8- A família Chelonianina é composta de um grupo de proteínas contendo oito resíduos de cisteínas caracteristicamente espaçados, os quais estão envolvidos em pontes dissulfeto. Embora o padrão das cisteínas seja conservado, o grau de similaridade global das sequências é baixo com poucas prolínas e glicinas, sendo razoavelmente bem conservadas (Figura 19) (Hennighausen & Sippel, 1982).

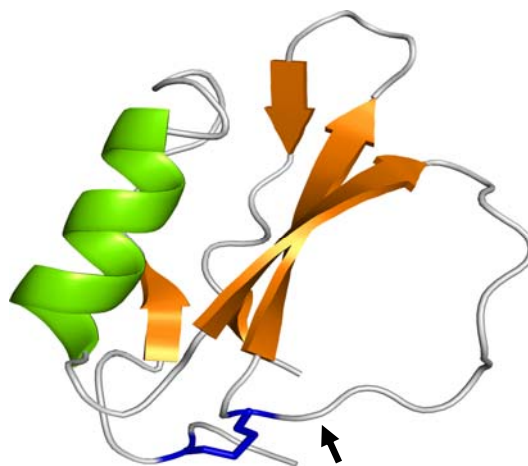
9- Os inibidores pertencentes a família ascaris possuem uma estrutura característica conhecida como domínio rico em cisteínas, que é encontrado em muitas proteínas extracelulares (Figura 20) (Grasberger *et al.*, 1994).



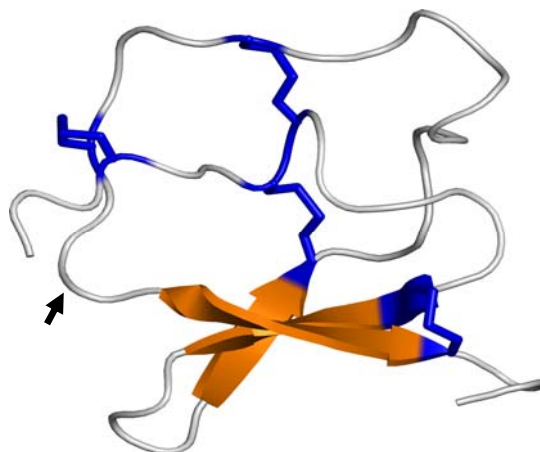
**Figura 12.** Estrutura tridimensional de STI extraído da *Glycine Max* (soja, PDB 1AVU); colorida de forma a mostrar as três unidades de repetição que formam o arranjo global conhecido como  $\beta$ -trefoil; as pontes dissulfeto estão representadas em azul; o *loop* reativo é indicado pela seta.



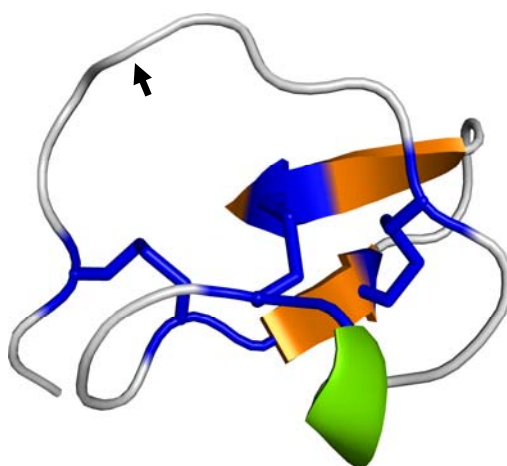
**Figura 13.** (A) Estrutura tridimensional de BBI extraído da *Pisum sativum* (ervilha, PDB 1PBI); as pontes dissulfeto estão representadas em azul; ambos os *loops* reativos que inibem tripsina e quimotripsina são indicados pelas setas; (B) estrutura do dímero formado pela junção de duas moléculas como mostradas em (A) sendo uma colorida em laranja e a outra em verde, as pontes dissulfeto estão em azul.



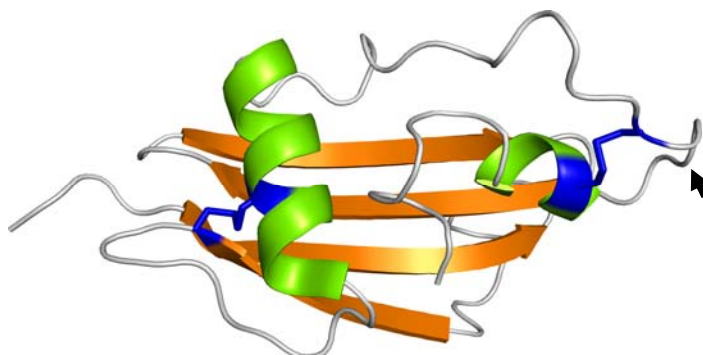
**Figura 14.** Estrutura tridimensional do inibidor tipo batata I extraído de *Cucurbita maxima* (abóbora, PDB 1MIT). A única ponte dissulfeto é mostrada em azul e o *loop* reativo é indicado pela seta.



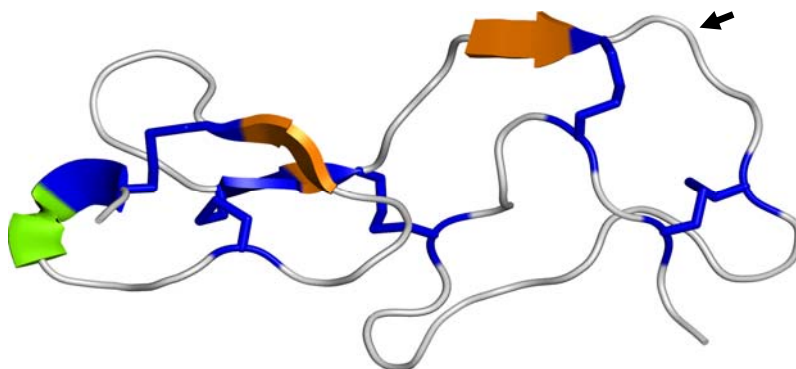
**Figura 15.** Estrutura tridimensional do inibidor tipo batata II extraído de *Nicotiana glauca* (fumo-de-jardim, PDB 1TIH). As pontes dissulfeto estão representadas em azul e o *loop* reativo é indicado pela seta.



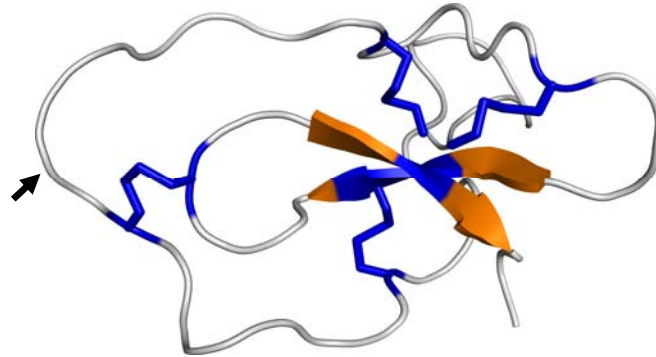
**Figura 16.** Estrutura tridimensional do inibidor tipo abóbora extraído de *Cucurbita maxima* (abóbora, PDB 3CTI). As pontes dissulfeto estão representadas em azul e o *loop* reativo é indicado pela seta.



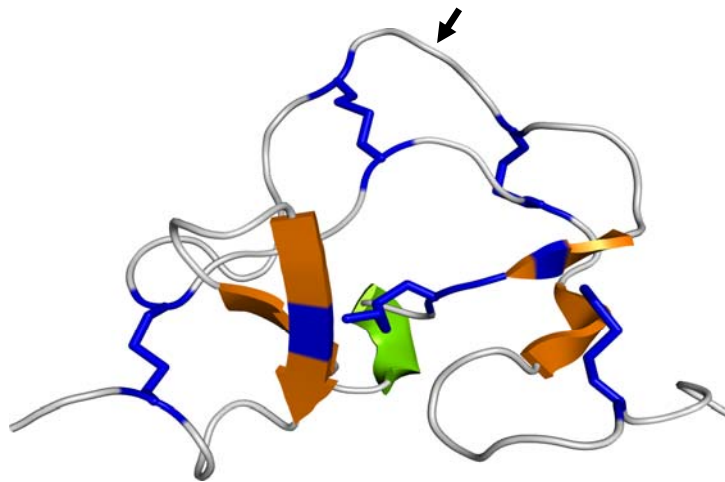
**Figura 17.** Estrutura tridimensional do inibidor de subtilisina extraído de *Streptomyces albogriseolus* da família SSI (Bactéria, PDB 3SSI). As pontes dissulfeto estão representadas em azul e o *loop* reativo é indicado pela seta.



**Figura 18.** Estrutura tridimensional do inibidor da família das antistasinas extraído da saliva de sanguessuga *Hirudo medicinalis* (PDB 1HIA). As pontes dissulfeto estão representadas em azul e o *loop* reativo é indicado pela seta.



**Figura 19.** Estrutura tridimensional da elafina humana (PDB 2REL) que possui o arranjo característico da família chelonianina. As pontes dissulfeto estão representadas em azul e o *loop* reativo é indicado pela seta.



**Figura 20.** Estrutura tridimensional do inibidor da família ascaris extraído de *Ascaris lumbricoides* (verme, PDB 1ATA). As pontes dissulfeto estão representadas em azul e o *loop* reativo é indicado pela seta.

## 2.9 Funções e Aplicações dos Inibidores de Proteases

Muitas pesquisas vêm sendo realizadas buscando determinar o papel biológico dos inibidores de proteases em plantas, e hoje se sabe que os inibidores exercem uma multiplicidade de papéis. Além do seu envolvimento nos processos de regulação das atividades proteolíticas das enzimas durante a germinação (Richardson, 1991), existem outras funções biológicas atribuídas aos inibidores, como a sua atuação como proteínas de reserva, participação nos mecanismos de defesa das plantas ao ataque de insetos, e o seu papel antifúngico e antimicrobiano.

A ação dos inibidores nesses casos está intimamente ligada às suas propriedades anti-nutricionais. Em um dos mecanismos propostos, eles atuam reduzindo a digestão das proteínas da dieta dos insetos pela inibição direta das enzimas digestivas desses predadores, que morreriam por inanição. Estudos mostram que se adicionando inibidores puros na ração artificial, reduz-se significativamente o crescimento e desenvolvimento de vários insetos e pragas. Mas outros estudos posteriores mostraram que a ingestão de inibidores causa uma significativa elevação dos níveis de tripsina no estômago dos insetos e, por isso, foi proposto que o verdadeiro modo de ação dos inibidores seria causar a hiperprodução perniciosa de tripsina (Cheng & Xue, 2003), o que levaria o inseto à morte por danos causados ao aparelho digestivo numa espécie de autodigestão.

Há um interesse muito grande no possível uso de inibidores de origem vegetal que atuam sobre proteases no tratamento de uma grande variedade de distúrbios metabólicos associados às enzimas proteolíticas, como, por exemplo, pancreatites, enfisemas, alergias, inflamações e hipertensão (Hsu *et al.*, 2006). Experimentos *in vivo* mostraram que inibidores de proteases de alguns tipos, especialmente os da família Bowman-Birk e Kunitz, possui uma boa resposta na prevenção e tratamento do câncer (McGuire, 2003). Essas propriedades incluem a habilidade para reduzir a formação de radicais de oxigênio, suprimir o crescimento de tumores anais e de

cólon induzidos quimicamente, tumor de mama em ratos e em humanos, e tumor de pulmão em camundongos (McGuire, 2003).

## **Objetivo Geral**

Analisar as interações entre serino proteases e seus inibidores, em termos das propriedades dos resíduos que estão na interface das duas moléculas.

## **Objetivos Específicos**

1. Criar um banco de dados não redundante de serino proteases.
2. Desenvolver uma nova metodologia para melhor entender os mecanismos que regem a interação entre serino proteases e inibidores.
3. Baseado nos tipos de resíduos de aminoácidos que compõe o IFR (*Interface Forming Residues*), determinar as posições que são mais ou menos restritivas em uma subfamília de serino protease com relação à especificidade.



## **Materiais e Métodos**

Neste capítulo apresentaremos todos os procedimentos realizados na seleção das bases de dados e os experimentos que foram utilizados.

### **1. Bancos de Dados de Proteínas**

#### **1.1. Uniprot**

O UniPROT (*Universal Protein Resource*) (2008) é um dos maiores bancos de dados que engloba dados de sequências de proteínas e dados de anotações, ele possui 40.5506 entradas (dados de 16-Dez-2008). Ele funciona como um hub central de informações sobre proteínas e suas funcionalidades com uma acurácia consistente e rica. Para cada entrada do UniPROT você pode encontrar basicamente a sequência de aminoácidos, o nome da proteína ou descrição, os dados taxonômicos e informações de citações. As sequências de proteínas encontradas no UniPROT são derivadas da tradução das sequências de ácidos nucleicos que são submetidas aos bancos de dados públicos, como EMBL/GenBank/DDBJ. Caso você queira também pode cruzar os dados encontrados no UniProt com dados de outros bancos de dados como por exemplo o PDB (Leinonen *et al.*, 2004).

#### **1.2. PDB**

O PDB (*Protein Data Bank*) (Berman *et al.*, 2000) é um repositório global de estruturas tridimensionais de moléculas biológicas o que incluem proteínas e ácidos nucleicos. Observando o PDB podemos notar um crescimento do número de estruturas ao longo dos anos, ele contém atualmente 55.126 estruturas de proteínas<sup>1</sup>, ácidos nucleicos e complexos protéicos. As sequências das estruturas existentes no

---

<sup>1</sup> Dados do PDB na data de 12/01/2009.

PDB podem ser muito redundantes; de acordo com o site do banco existem atualmente 20.229 cadeias<sup>1</sup> com valor menor ou igual a 95% de identidade. Para cada cadeia, podem existir dados de diversos mutantes simples ou múltiplos além da existência de múltiplos cenários experimentais nos quais a estrutura foi resolvida.

A maioria das estruturas depositadas no PDB foram resolvidas utilizando o método de cristalografia de raios-X, mas existem outras técnicas que também são usadas para resolução de estruturas, que são a ressonância nuclear magnética (NMR -Nuclear Magnetic Resonance) e a microscopia eletrônica. Em média, a resolução é de 2,18Å com desvio padrão de 1,31Å (Berman *et al.*, 2000).

### 1.3. SCOP

O SCOP (Structural Classification of Proteins) (Murzin *et al.*, 1995) é um banco de dados criado pela inspeção manual e auxiliado por um conjunto de métodos automatizados, e tem como objetivo fornecer uma detalhada descrição das relações evolutivas entre todas as estruturas protéicas conhecidas. O SCOP classifica as estruturas do PDB de acordo com os diferentes tipos de família, superfamília, enovelamento e classes. Na versão atual (1.73) do SCOP a partir de 55.126 entradas do PDB foram anotadas 34.494, o que significa 97178 cadeias de 1.086 diferentes enovelamentos. As proteínas de uma mesma família têm alta similaridade seqüencial e estrutural. Proteínas da mesma superfamília são provavelmente relacionadas evolutivamente compartilhando o mesmo enovelamento e desempenho de funções bastante similares. Proteínas compartilham o mesmo enovelamento se possuem o mesmo arranjo na sua arquitetura, ou seja, são estruturalmente muito próximas. As classes do SCOP são definidas com base na composição das cadeias em termos de estruturas secundárias: se a maioria é  $\alpha$  (formadas, na maioria, por  $\alpha$ -hélices) ou  $\beta$  (formadas, na maioria, por folhas- $\beta$ ) ou uma junção delas (Murzin *et al.*, 1995).

O SCOP é um poderoso banco de dados uma vez que ele possui uma excelente anotação do PDB e bem organizada, o que pode ajudar muito na validação

dos dados em uma pesquisa com proteínas. Dentro do SCOP também podemos encontrar arquivos texto facilmente legíveis por *scripts*, nos quais podemos ter não somente a classificação em termos de classe, superfamílias, famílias, e enovelamento, mas temos também o nome da cadeia e do organismo do qual foi extraído a proteína. Neste trabalho usamos a classificação pelo tipo de família e enovelamento.

#### 1.4. STING

O STING (Neshich *et al.*, 2005b) é um banco de dados que possui muitas ferramentas utilizadas para analisar a estrutura protéica. Dentro do STING podemos encontrar vários módulos, por exemplo, podemos visualizar a forma estrutural da proteína e também sua sequência através do Jmol. O módulo JPD (*Java Protein Dossier*) possui vários parâmetros dentro dele, mas vamos destacar um que foi usado em nosso trabalho que é o IFR (Interface Forming Residues), que são os resíduos de aminoácidos localizados na interface de duas subunidades ou cadeias de algum complexo; esses resíduos são aqueles que perderam acessibilidade ao solvente uma vez que o complexo foi formado. Através de outro módulo, o Contatos do IFR, também é possível calcular todos os resíduos que estão na interface de duas moléculas que fazem algum tipo de contato como:

- Contatos hidrofóbicos
- Contatos carregados atrativos
- Contatos carregados repulsivos
- Pontes de hidrogênio
- Empilhamento aromático
- Pontes dissulfeto

Assim sendo, podemos dizer que os resíduos formadores da interface são estritamente definidos e baseados na acessibilidade ao solvente. Entretanto, todos aqueles resíduos da superfície das duas cadeias que não perderam acessibilidade ao solvente, ou não estão localizados nas imediações e na fronteira entre as subunidades, não são considerados como IFR.

## 2. Seleção da Base de Dados de Serino Proteases

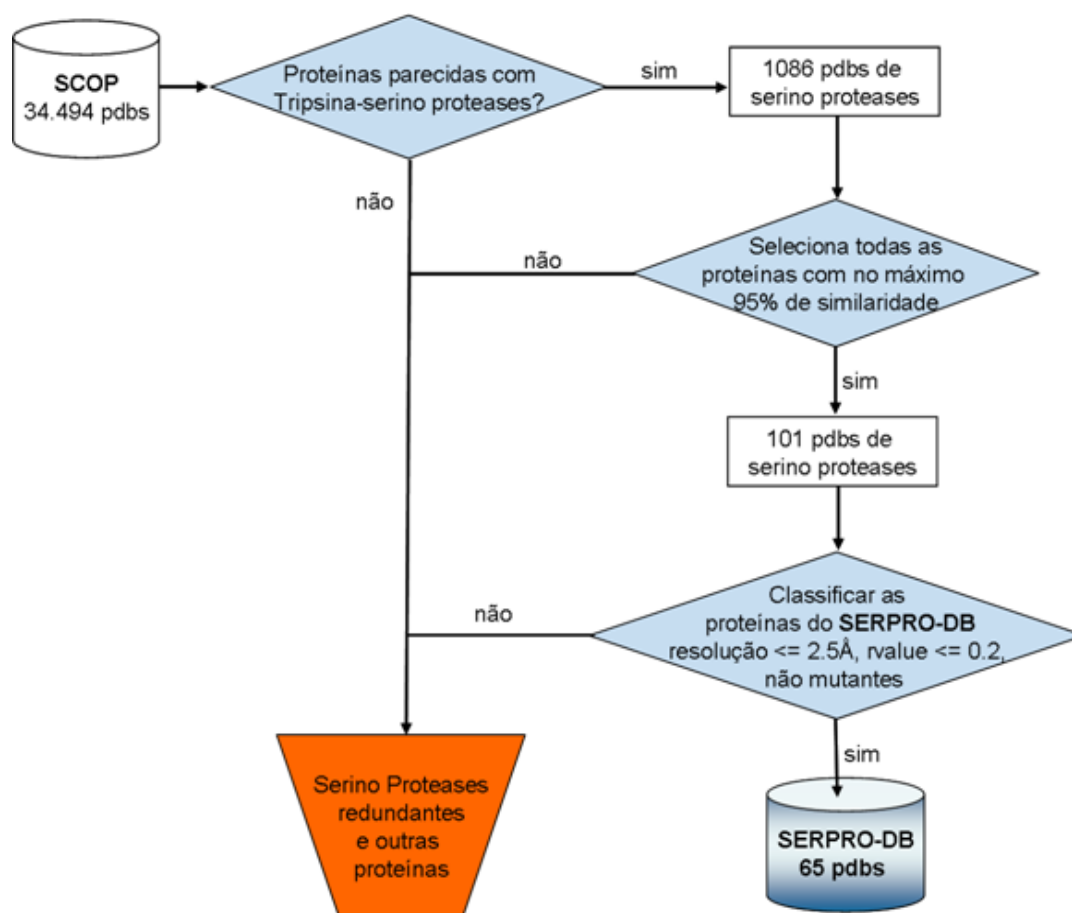
Para termos uma base de dados consistente precisávamos selecionar todas as serino proteases existentes com estruturas resolvidas. Primeiramente foi feita uma busca sobre serino proteases no UniProt; essa busca retornou 7.008 saídas com a palavra-chave serino proteases, mas como dentro do UniProt temos sequências sem estruturas resolvidas, precisávamos melhorar mais ainda a nossa busca.

Para buscarmos somente estruturas de serino proteases utilizamos então o banco de dados SCOP que classifica as estruturas por seu tipo de enovelamento; o escolhido foi proteínas parecidas com a tripsinas (*trypsin-like serine proteases*), e assim foram selecionadas 1.086 estruturas do PDB. A partir dessas 1.086 estruturas foi feito um refinamento baseado na sequência dos aminoácidos de cada estrutura, ou seja, foram selecionadas somente as estruturas que tinha valor menor ou igual a 95% de identidade entre si, assim ficamos com 101 estruturas do PDB.

Uma vez que obtivemos 101 serino proteases essas proteínas passaram pelos seguintes filtros: resolução menor ou igual a 2,5Å, *R-Value Working*<sup>2</sup> menor que 0,2, identidade de sequência menor que 95% e não mutantes, o que resultou em 67 proteínas. A Figura 21 mostra um fluxograma da seleção da base de dados.

---

<sup>2</sup> Nós usamos *R-Value Working* ou *R-Value Working Test* porque verificamos empiricamente que esses campos estão mais frequentemente anotados nos arquivos PDB que o *Free R-Value*.



**Figura 21.** Fluxograma de busca e seleção das serino proteases. A busca foi feita baseada no Banco de Dados SCOP.

Na Tabela 6, as quatro últimas famílias, foram agrupadas por alguma característica em comum, e foram usadas algumas siglas para agrupar essas características; são elas, respectivamente: VaP – Vários Procariotos, VaV – Vários Virus e Va – Vários, que eram um agrupamento de várias proteínas que eram representantes únicos. As três últimas proteínas são ligadas respectivamente com os inibidores BPTI, ovomucoide e ecotina e foram as proteínas usadas como base para nossos experimentos mais adiante, como veremos mais detalhadamente.

**Tabela 6.** 67 serino proteases não redundantes com estruturas 3D resolvidas encontradas dentro do PDB, agrupadas de acordo com as suas subfamílias.

	<b>ID-PDB</b>	<b>SUBFAMÍLIA</b>	<b>ORGANISMO</b>
1	1LTO:A	Triptase	<i>Homo sapiens</i>
2	2FPZ:A	Triptase	<i>Homo sapiens</i>
3	1NN6:A	Quimase	<i>Homo sapiens</i>
4	3RP2:A	Quimase	<i>Rattus rattus</i>
5	1K2I:1	Quimotripsina	<i>Bos taurus</i>
6	1EQ9:A	Quimotripsina	<i>Solenopsis invicta</i>
7	1GCT	Quimotripsina	<i>Bos taurus</i>
8	1PYT	Quimotripsina	<i>Bos taurus</i>
9	1ELT:A	Elastase	<i>Salmo salar</i>
10	1M9U:A	Elastase	<i>Eisenia foetida</i>
11	1HNE:E	Elastase	n/a
12	1GVK:B	Elastase	<i>Sus scrofa</i>
13	1BRU:P	Elastase	<i>Sus scrofa</i>
14	1AGJ:A	Exfoliativa toxina A	<i>Staphylococcus aureus</i>
15	1QTF:A	Exfoliativa toxina B	<i>Staphylococcus aureus</i>
16	1HPG:A	Glutamil endopeptidase	n/a
17	2O8L:A	Glutamil endopeptidase	<i>Staphylococcus aureus</i>
18	1P3C:A	Glutamil-endopeptidase	<i>Bacillus intermedius</i>
19	2PKA:A,B	Calicreína	<i>Sus scrofa</i>
20	1GVZ:A	Calicreína	<i>Equus caballus</i>
21	1LO6:A	Calicreína	<i>Homo sapiens</i>
22	1TON:A	Calicreína	<i>Rattus rattus</i>
23	2FOM:B	NS3 protease	<i>Dengue virus type 2</i>
24	2FP7:B	NS3 protease	<i>West Nile virus</i>
25	1DY9:A,C	NS3 protease	<i>Hepatitis C virus</i>
26	2SGA:A	Estreptogrisina A	<i>Streptomyces griseus</i>
27	1SGP:E	Estreptogrisina B	<i>Streptomyces griseus</i>
28	1PQ7:A	Tripsina	<i>Fusarium oxysporum</i>
29	1TRN:A	Tripsina	<i>Homo sapiens</i>
30	1HJ9:A	Tripsina	<i>Bos taurus</i>
31	1HJ8:A	Tripsina	<i>Salmo salar</i>
32	2F91:A	Tripsina	<i>Pontastacus leptodactylus</i>
33	1H4W:A	Tripsina	<i>Homo sapiens</i>
34	2A31:A	Tripsina	<i>Sus scrofa</i>
35	1A0J:A	Tripsina	<i>Salmo salar</i>
36	1OS8:A	Tripsina	<i>Streptomyces griseus</i>
37	1OP0:A	Veneno	<i>Agkistrodon acutus</i>
38	1BQY:A	Veneno	<i>Trimeresurus stejnejeri</i>
39	1C5L:H	Trombina	<i>Homo sapiens</i>
40	1VR1:H	Trombina	<i>Homo sapiens</i>
41	2PUX:B	Trombina	<i>Mus musculus</i>
42	1H8D:H	Trombina	<i>Homo sapiens</i>
43	1ETR:H	Trombina	<i>Synthetic construct</i>
44	1ARB:A	lisil endopeptidase- VaP	<i>Achromobacter lyticus</i>
45	1Y8T:A	Protease PepD - VaP	<i>Mycobacterium tuberculosis</i>

46	2SFA:A	Hipotética - VaP	<i>Streptomyces fradiae</i>
47	1SOT:A	Protease DegS - VaP	<i>Escherichia coli</i>
48	2H5C:A	$\alpha$ -lytic endopeptidase - VaP	<i>Lysobacter enzymogenes</i>
49	1NPM:A	Neuropsina - VaV	<i>Mus musculus</i>
50	1EP5:A	Proteína de Capsídio Viral- VaV	<i>Venezuelan equine encephalitis</i>
51	1FIW:A	Acrosina - Va	<i>Ovis aries</i>
52	1DLE:A	C3/C5 convertase - Va	<i>Homo sapiens</i>
53	1T32:A	Catepsina G - Va	<i>Homo sapiens</i>
54	2BZ6:H	Fator Coagulação VIIa - Va	<i>Homo sapiens</i>
55	1BIO:A	Fator Coagulação Factor D - Va	<i>Homo sapiens</i>
56	1EUF:A	Duodenase - Va	<i>Bos taurus</i>
57	1EKB:B	Enteropeptidase - Va	<i>Bos taurus</i>
58	1IAU:A	Granzime B - Va	<i>Homo sapiens</i>
59	2HLC:A	HL collagenase - Va	<i>Hypoderma lineatum</i>
60	1Q3X:A	MASP-2 - Va	<i>Homo sapiens</i>
61	1EAX:A	Matriptase - Va	<i>Homo sapiens</i>
62	1FUJ:A	Mieloblastina - Va	<i>Homo sapiens</i>
63	1MBM:A	NSP4 proteinase - Va	<i>Equine arteritis virus</i>
64	1FON:A	Procarboxipeptidase A-S6 - Va	<i>Bos taurus</i>
65	1RTF:B	Ativador T-plasminogen - Va	<i>Homo sapiens</i>
66	1C5Y:B	Ativador U-Plasminogen - Va	<i>Homo sapiens</i>
67	1A7S:A	Heparina - Va	<i>Homo sapiens</i>
68	1FY8	Tripsina - Complexo	<i>Rattus rattus</i>
69	1CHO	Quimotripsina - Complexo	<i>Bos Taurus</i>
70	1AZZ	Colagenase - Complexo	<i>Celuca pugilator</i>

**Tabela 7.** 67 serino proteases não redundantes e o seu respectivo sítio de clivagem preferencial, as proteínas estão agrupadas por subfamílias de acordo com a Tabela 6.

ID-PDB	REAÇÃO-CLIVAGEM PREFERENCIAL
1 1LTO:A	R- - , K- - , possui uma restrição maior que a tripsina
2 2FPZ:A	R- - , K- - , possui uma restrição maior que a tripsina
3 1NN6:A	F- -Xaa > Y- -Xaa > W- -Xaa > L- -Xaa.
4 3RP2:A	
5 1K2I:1	Y- -Xaa, W- -Xaa, F- -Xaa, L- -Xaa
6 1EQ9:A	Y- -Xaa, W- -Xaa, F- -Xaa, L- -Xaa
7 1GCT	Y- -Xaa, W- -Xaa, F- -Xaa, L- -Xaa
8 1PYT	Y- -Xaa, W- -Xaa, F- -Xaa, L- -Xaa
9 1ELT:A	
10 1M9U:A	
11 1HNE:E	V- -Xaa > A- -Xaa
12 1GVK:B	A- -Xaa
13 1BRU:P	L- -Xaa, M- -Xaa and F- -Xaa
14 1AGJ:A	
15 1QTF:A	
16 1HPG:A	E- -Xaa >> D- -Xaa. Preferência por P or L no P2 e F no P3. Clivagem da ligação E- -P e D- -P lenta
17 2O8L:A	D- -Xaa, E- -Xaa.
18 1P3C:A	
19 2PKA:A,B	R- -Xaa para pequenas moléculas. Alta ação de seletividade para kalidina (lisill-bradikinina) para kininogen envolve a hidrólise da M- -Xaa or L- -Xaa.
20 1GVZ:A	R- -Xaa para pequenas moléculas. Alta ação de seletividade para kalidina (lisill-bradikinina) para kininogen envolve a hidrólise da M- -Xaa or L- -Xaa
21 1LO6:A	
22 1TON:A	R- -Xaa para pequenas moléculas. Alta ação de seletividade para kalidina (lisill-bradikinina) para kininogen envolve a hidrólise da M- -Xaa or L- -Xaa
23 2FOM:B	
24 2FP7:B	Hidrólise seletiva de ligações Xaa-Xaa- -Xbbem que cada Xaa pode ser R or K e cada Xbb pode ser S or A
25 1DY9:A,C	Hidrólise de 4 peptídeos liga o precursor viral poliproteína, com A or E na posição P6, C ou T em P1 e S ou A em P1'
26 2SGA:A	Hidrólise de proteínas com especificidade similar as quimotripsinas.
27 1SGP:E	Hidrólise de proteínas com especificidade similar as quimotripsinas
28 1PQ7:A	R- -Xaa, K- -Xaa
29 1TRN:A	R- -Xaa, K- -Xaa
30 1HJ9:A	R- -Xaa, K- -Xaa
31 1HJ8:A	R- -Xaa, K- -Xaa
32 2F91:A	R- -Xaa, K- -Xaa
33 1H4W:A	R- -Xaa, K- -Xaa
34 2A31:A	R- -Xaa, K- -Xaa
35 1A0J:A	R- -Xaa, K- -Xaa
36 1OS8:A	R- -Xaa, K- -Xaa
37 1OP0:A	



38	1BQY:A	
39	1C5L:H	R- -G
40	1VR1:H	R- -G
41	2PUX:B	
42	1H8D:H	R- -G
43	1ETR:H	R- -G
44	1ARB:A	K- -Xaa, incluindo K- -P
45	1Y8T:A	
46	2SFA:A	
47	1SOT:A	
48	2H5C:A	Hidrólise de proteínas com ligações adjacentes L-A and L-V, elastina
49	1NPM:A	
50	1EP5:A	
51	1FIW:A	Hidrólise da R- -Xaa e K- -Xaa; clivagem preferencial R- -Xaa >> K- -K >> K- -Xaa
52	1DLE:A	Cliva C3 na cadeia alfa de C3a e C3b. Cliva C5 na cadeia alfa de C5a e C5b. Ambos clivam em um local perto do C-terminal de em resíduo de arginina
53	1T32:A	Especificidade similar a quimotripsina.
54	2BZ6:H	Hidrolisa uma R- -I ligada no fator X para formar fator Xá
55	1BIO:A	Cliva componentes do fator B (R- -K) quando em complexo com C3b ou com veneno de cobra fator (CVF)
56	1EUF:A	
57	1EKB:B	Clivagem seletiva de 6-K- -I-7 ligada em tripsinogênio.
58	1IAU:A	D- -Xaa >> N- -Xaa > M- -Xaa, S- -Xaa
59	2HLC:A	
60	1Q3X:A	Clivagem seletiva depois da R223 em complemento com o componente C2 (-S-L-G-R- K-I-Q-I) e depois da R76 em complemento com o componente C4 (-G-L-Q-R- A-L-Q-I)
61	1EAX:A	
62	1FUJ:A	A- -Xaa > V- -Xaa
63	1MBM:A	
64	1FON:A	
65	1RTF:B	R- V
66	1C5Y:B	R- V
67	1A7S:A	
68	1FY8	R- -Xaa, K- -Xaa
69	1CHO	Y- -Xaa, W- -Xaa, F- -Xaa, L- -Xaa
70	1AZZ	Degrada a colagenase até 75% da molécula a partir N-terminal. Baixa atividade em pequenos substratos

### 3. Seleção dos Complexos Proteína/Inibidor

Os complexos serino proteases/inibidores foram selecionados também utilizando o SCOP. Primeiramente foi feita uma revisão bibliográfica pra saber quais são os inibidores de serino proteases existentes, depois foi feita uma busca usando como palavra chave a família dos inibidores. Abaixo estão listadas todas as famílias de inibidores e os seus respectivos complexos com serino proteases:

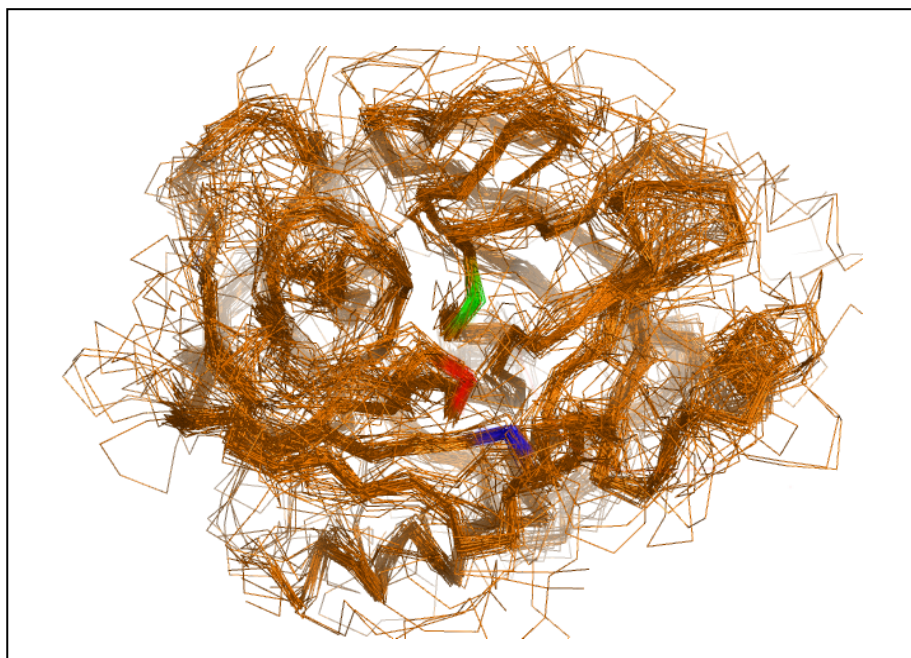
- Serpinas - 1EZX, 1OPH, 1TB6, 1SR5
- Kunitz (*Soybean Trypsin inhibitor*) - 1AVW, 1AVX
- Ecotina - 1N8O, 1AZZ
- CI-2-Subtilisinas - 1CSE, 2SEC, 2TEC, 3TEC, 1ACB, 1TEC, 1MEE, 2SNI
- Parecidas com ATI (*Ascaris trypsin inhibitor*) - 1EAI
- Ovomucoide III - 1R0R, 3SGB, 1PPF, 1CHO, 1Z7K, 1CGI, 1CGJ, 1TGS, 1TBR, 1TBQ, 1TDT
- Inibidores de plantas (*Abobora*, *Curcubita sp*) - 1MCT, 1F2C, 2STA, 1PPE, 2BTC, 2STB, 1MCV, 4CPA
- PMP- 1GL1, 1GL0
- Elafina - 1FLE
- Inibidores extraídos de plantas (*Batata*, *Solanum tuberosum*) - 4SGB
- BPTI (*Bovine Pancreatic Trypsin Inhibitors*) - 3TGI, 1TPA, 3BTK, 3TPI, 2PTC, 2TGP, 1FY8, 4TPI, 2TPI, 1BZX, 1BRB, 1FAK, 1CBW, 2KAI, 1EAW, 1MTN

Como existem 11 famílias de inibidores de serino proteases, decidimos escolher somente as famílias mais estudadas pra trabalhar, que são: ecotina, ovomucoide e BPTI, correspondendo respectivamente a g.8, b.16 e g.68 subfamílias do SCOP. Para selecionarmos um representante de cada um das três famílias de inibidores, foram utilizados os seguintes critérios: complexos enzima/inibidor com melhor resolução e não mutante; sendo assim, no final ficamos com três complexos representantes das famílias da ecotina 1AZZ (inibindo a colagenase de caranguejo) (Perona *et al.*, 1997), ovomucoide 1CHO (inibindo a alfa-quimotripsina bovina) (Fujinaga *et al.*, 1987) e BPTI 1FY8 (inibindo a tripsina de rato) (Pasternak *et al.*, 2001).

O complexo 1CHO tem a sua estrutura formada por dois barris representada por quatro cadeias diferentes E F G I.

#### **4. Alinhamento Estrutural Usando o PrISM**

O PrISM (Protein Informatics System for Modeling) (Yang & Honig, 1999) é um *software* para análise e modelagem de proteínas, que tem como vantagem suportar um grande número de cadeias; ele utiliza somente os dados das cadeias protéicas como parâmetro de entrada. Este pacote foi escolhido pelo seu bom desempenho no alinhamento estrutural de muitas moléculas (mais de 50, que é o nosso caso). Como saída o PrISM possui um alinhamento estrutural contendo todas as estruturas do PDB com coordenadas similares, alinhamento de sequência baseado na estrutura e uma matriz com todos os RMSD (Root Mean Square Deviation) de todas as cadeias entre si, que foram alinhadas estruturalmente. Do conjunto curado de serino proteases foram selecionadas 67 serino proteases e três complexos nativos totalizando 70 estruturas. O PDB do complexo 1CHO foi editado para trocar três cadeias da enzima (E, F, G) por cadeia E, para distinguir ele facilmente da cadeia do inibidor I. Para o complexo 1AZZ (que é um dímero) foi selecionado somente a cadeia A (enzima) e C (inibidor). O complexo 1FY8 não foi editado. As 70 estruturas alinhadas pelo software PrISM podem ser vistas na Figura 22 do alinhamento estrutural; somente as coordenadas da enzima foram carregadas pra fazer o alinhamento.

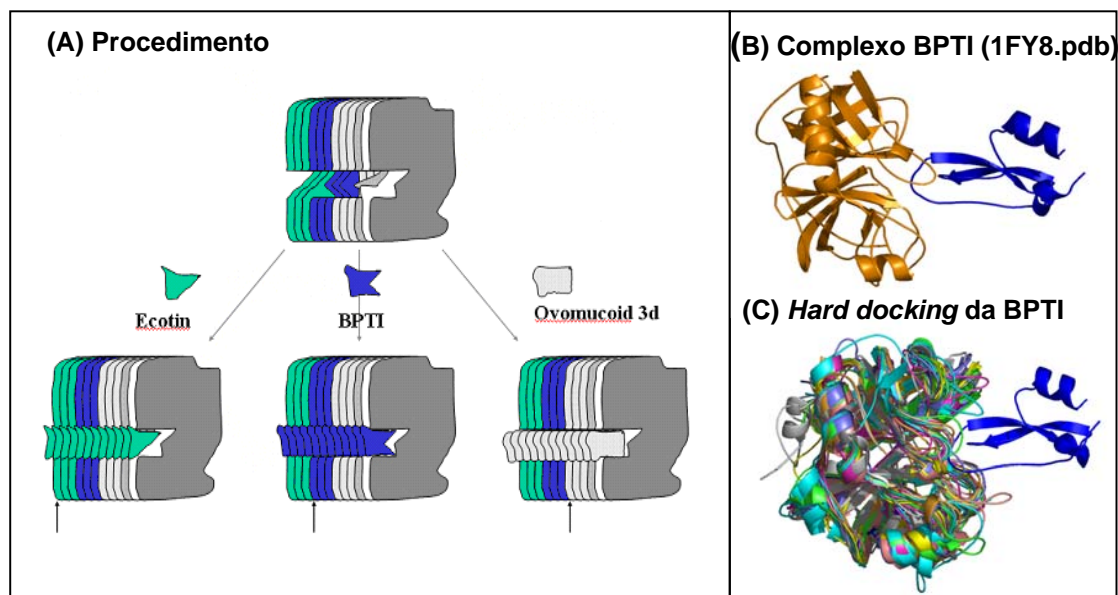


**Figura 22.** Alinhamento estrutural de 68 serino proteases cujas sequências possuem no máximo 95 % identidade entre si, alinhadas pelo pacote PrISM. A imagem foi produzida pelo pacote PyMOL. Somente a cadeia principal é representada. As posições destacadas são Ser 195 em verde , His 57 em vermelho e o Asp 102 em azul.

## **5. *Hard Body Docking***

O termo “*hard body docking*” é usado neste trabalho para definir o que foi feito: adicionando as coordenadas a cada arquivo PDB contendo uma estrutura de serino protease resultante do alinhamento do PrISM. Estes arquivos PDBs foram gerados separadamente a partir da saída do PrISM anteriormente definida como super PDB. Um par composto pelo complexo original (ex. 1FY8) e seu correspondente arquivo PDB que foi gerado pela saída do PrISM (ex. 1FY8\_PrISM) foi alinhado pelo software PyMol com a linha de comando “align(1FY8, 1FY8\_PrISM)” que mantinha as coordenadas do arquivo PDB do PrISM fixo sendo sobreposta pelo complexo nativo, assim sendo as coordenadas do inibidor foram relocadas junto com a molécula nativa. As coordenadas do inibidor foram então extraídas a partir do arquivo PDB de saída do PyMol e foi incluída em todos os arquivos das outras 69 serino proteases geradas pelo PrISM. Os detalhes do procedimento descrito acima são mostrados na Figura 23. Desde que completamos esse procedimento, todos os novos complexos

estruturais podem ser descritos como serino proteases consistentemente orientadas de uma única maneira (obedecendo ao alinhamento estrutural). Note que muitos complexos na Figura 23 (A) são de fato “forçados” por serem formados *in silico* e terem sido criados por um procedimento de “*hard body docking*”, mas isso pode acontecer devido a vários fatores, incluindo a forma não complementar do inibidor e o seu bolsão correspondente do lado da enzima, provavelmente alguns não poderiam ser formados naturalmente, como será discutido adiante.



**Figura 23.** *Hard body docking* das coordenadas do inibidor com outras proteases. (A) Diagrama esquemático de um experimento para obter a complexo estrutural de 3 diferentes conjuntos de 70 serino proteases com cada um dos três inibidores selecionados: Ecotina (verde), BPTI (azul), Ovomucoide (Branco). As setas embaixo de cada um dos três conjuntos de 70 estruturas de serino proteases alinhadas estruturalmente estão indicando que somente uma estrutura tinha coordenada de ambos da serino protease e seu inibidor correspondente (indicado pela mesma cor). Todas as outras proteases “receberam” o inibidor na posição idêntica em que ele estava no complexo enzima/inibidor (indicado pela seta abaixo dos alinhamentos estrutural). Para cada conjunto obtido, um conjunto de IFR foi extraído e analisado; (B) estrutura do complexo da Trypsina/BPTI como foi encontrado no 1FY8.pdb. (C) BPTI inserido no *hard body docking* no alinhamento de 70 serino proteases. Uma dessas 70 estruturas de enzima é exatamente como apresentada na letra (B).

## 6. Determinação da Superfície em Complexos *in silico*

Nos três conjuntos de 70 complexos enzima/inibidor foram determinados a acessibilidade do solvente à superfície de aminoácidos antes e depois da ligação do inibidor. O cálculo da acessibilidade da superfície dos aminoácidos ao solvente para ambos inibidor e enzima (isolados e depois de se ligarem) foi feita usando o algoritmo do SurfV (Sridharan et al., 1992). O algoritmo SurfV calcula a área acessível ao solvente que é definida como uma sonda que roda na superfície da molécula. O uso de uma sonda de raio zero retornará a área da superfície de Van der Waals. O tamanho do raio da sonda que em geral é usado para desenhar a superfície é de 1.4 Å (raio de uma molécula de água), e foi este o valor usado no nosso trabalho. As moléculas de hidrogênio são ignoradas no cálculo da superfície. Os resíduos que perderam acessibilidade ao solvente depois de feita a formação do complexo *in silico* são chamados de IFR (Interface Forming Residues), ou seja, eles agora estão na interface de duas moléculas enzima/inibidor.

## 7. Tabela de Resíduos Formadores da Interface

Para cada um dos conjuntos de 70 complexos, foram mapeadas todas as posições das sequências que pertenciam ao conjunto da IFR, esse mapeamento foi feito em uma das saídas do PrISM que é o alinhamento de sequências baseado na estrutura. Para numerar esse alinhamento foi usado como referência o complexo nativo tripsina-BPTI (1FY8), por que ele segue a numeração padrão das serino proteases (Greer, 1990).

Em todos os resíduos que foram detectados a perda de alguma área que era acessível ao solvente foram marcados no alinhamento de sequência baseado na estrutura obtido pelo PrISM, seguindo a numeração da sequência da cadeia E do 1FY8. No caso de *gap* na sequência usada como referência, uma letra foi adicionada ao número anterior, assim mantinha-se a numeração seguindo a referência. Colunas que não tinham nenhum resíduo marcado como IFR ou apenas um foram eliminados

para facilitar a inspeção. As posições que não foram identificadas como IFR de um par particular serino protease/inibidor são apresentadas com o sinal “-”;

Uma vez que construímos três conjuntos de 70 complexos, cada um deles com o seu respectivo inibidor fizemos também uma investigação nos relacionamentos entre características estruturais e especificidade das enzimas. Foi calculada a diferença no tipo de resíduo que ocupa a superfície da enzima (e do inibidor) como um todo e sua respectiva interface. A idéia era verificar se a interface era diferente e, se a resposta fosse sim, como eles se diferenciam do resto da superfície da proteína em termos do tipo de resíduo que forma essa interface.

## Resultados e Discussão

Nesse trabalho iremos expandir o escopo de interesse de especificidade das enzimas. Esse aumento pode contribuir para não focarmos somente na tríade catalítica e no bolsão de ligação, mas na grande categoria de aminoácidos que nós chamamos de resíduos formadores da interface (*Interface Forming Residues* - IFR) (Neshich *et al.*, 2005a; Neshich *et al.*, 2005b; Neshich *et al.*, 2006). Para ser mais exato é uma parte da superfície da molécula que era coberta pelo solvente e depois do complexo ser formado perdeu a acessibilidade ao solvente. Assim, aqueles resíduos que perderam acessibilidade ao solvente têm um importante papel no processo de *docking* e também na definição da especificidade (Da Silva *et al.*, 2000; Neshich *et al.*, 2003). A partir dos complexos formados *in silico* entre as 70 serino proteases e os três inibidores protéicos, BPTI, ecotina e ovomucoide, calculamos os aminoácidos que fazem parte do IFR. Então criamos uma tabela com todas as posições de aminoácidos da interface e sua respectiva ocupação.

Através do mapeamento desses aminoácidos, pode-se então analisar suas características e determinar posições específicas no alinhamento entre diferentes subfamílias de serino proteases. Um passo chave em nosso procedimento era encontrar um número suficiente de estruturas PDB que continham complexos de serino proteases e inibidores. Claramente precisávamos fabricar algum novo método ou abandonar o nosso trabalho, porque no PDB não poderíamos encontrar exemplos de complexos em quantidades estatisticamente significantes. A solução para esse desafio foi apresentada em detalhes no material e métodos.

Nesse trabalho desenvolvemos um método para analisar o relacionamento estrutura e função, bem como concretizar indicações de como a especificidade de várias serino proteases pode ser encontrada ou alterada.

Já foi estabelecido que o aumento da plasticidade estrutural no bolsão de ligação aumenta a variação do tamanho do substrato que pode caber no espaço



crítico em torno da tríade catalítica, ampliando então a especificidade (Bone *et al.*, 1989a; Bone *et al.*, 1989b). Conseqüentemente, é natural esperar que uma alta rigidez em torno do bolsão promove uma alta pressão seletiva para o tipo de substrato que a enzima pode operar (Perona & Craik, 1995). Especificidade então é diretamente proporcional às limitações estruturais impostas primeiramente pelo tamanho do espaço de *docking* e também pelas características físico-químicas desse espaço. Neste trabalho iremos focar nossa atenção nos tipos de resíduos que ocupam a interface da enzima quando o inibidor se liga a ela, e iremos fazer uma avaliação de como a plasticidade do bolsão de ligação influencia na especificidade.

## 1. Análise das Interfaces

Os tipos de resíduos que ocupam as posições da interface implicitamente contêm um tipo de informação. Esses resíduos que formam a interface em cada complexo serino protease/inibidor foram marcados no alinhamento baseado na estrutura de saída do PrISM. Nesse alinhamento foram selecionados todos os resíduos que faziam parte da IFR. As Tabelas 8A, 8B e 8C mostram respectivamente os resíduos formadores da IFR da enzima com o Ecotina, BPTI, e Ovomucoide. As linhas representam as 70 enzimas da família das serino proteases selecionadas, as colunas as posições dos resíduos de aminoácidos numeradas de acordo com a enzima 1FY8 do PDB. As posições em que os resíduos de IFR apareciam somente em até 15 das 70 serino proteases foram retiradas da figura; isso foi feito como uma forma de classificar os dados, ou seja, posições que tinha uma freqüência maior de dados foram mantidas na análise e as restantes foram retiradas. O alinhamento total de sequências baseado no alinhamento estrutural pode ser visto em anexo.

Na Tabela 8A podemos ver como a His 57 da tríade catalítica se conserva bem em quase todas as serino proteases com exceção da 1GCT (gama quimotripsina) e 1MBM (NSP4 protease de vírus). Analisando os dados apresentados na Tabela 8A podemos notar claramente algumas diferenças nas subfamílias com respeito à interface dos complexos formados com os inibidores protéicos. Como exemplo,

podemos citar a família das trombinas 39 1C5L - 43 1ETR, que possui algumas posições bem conservadas para todos os cinco representantes da subfamília. Os dados compilados nessa tabela podem ser considerados como uma poderosa ferramenta para decifrar função, especialmente para planejarmos mudanças necessárias na estruturas/sequência para introduzir uma alteração no tipo de atividade por exemplo, transformar uma tripsina em uma quimotripsina (ou mais generalizado: transformar uma subfamília de serino protease em outra). No passado, muitas tentativas foram feitas com a mesma finalidade de alterar ou transformar a atividade focando em um único resíduo ou até mesmo em muitos (Page & Cera, 2007; Perona & Craik, 1995).

Os dados da Tabela 8 indicam que a alteração da função pode ser realizada eliminando diferenças chaves na ocupação do IFR. Analisar 16 diferentes famílias de serino proteases seria um trabalho muito extenso, decidimos então focar em quatro subfamílias mais estudadas e conhecidas atualmente; são elas as tripsinas, quimotripsinas, elastases e trombinas.

O código de cores dos aminoácidos da Tabela 8 segue o padrão apresentado por Taylor (1997), como a seguir:

- Resíduos pequenos e hidrofóbicos: AVLIMFP são coloridos em cinza;
- Resíduos polares: STYNQWG são coloridos em verde;
- Resíduos negativamente carregados: D E são coloridos em vermelho;
- Resíduos positivamente carregados: R K são coloridos em azul;
- Resíduos formadores de ponte dissulfeto: C coloridos em amarelo;
- Resíduos de Histidina: H coloridos em turquesa.

**Tabela 8.** Posições dos IFRs de 70 serino proteases ligadas com os inibidores: Ecotina (**A**), BPTI (**B**), e Ovomucoide (**C**). As linhas representam as sequências primárias (resíduos de aminoácidos) dos IFRs para cada serino protease específica (o PDB ID de cada complexo é mostrado na segunda coluna). As colunas depois do PDB ID mostram as numerações do alinhamento de sequência baseado no alinhamento estrutural de serino proteases (somente os IFRs são mostrados), para todas as 70 serino proteases em que foi feito o *hard body docking* com o inibidor BPTI. As linhas com as subfamílias de serino proteases são listadas respectivamente a seguir: Triptase (1-2), Chimase (5-8), Quimotripsina (5-8) {marcado em vermelho}, Elastase (9-13){marcado em azul}, Exfoliative Toxina (14-15), Glutamil endopeptidase (16-18), Calicreína (19-22), NS3 protease (23-25), Streptogrisina (26-27), Tripsina (28-36){marcado em verde}, Veneno (7-38), Trombina (39-43){marcado em amarelo}, Varios-Procarotos (44-48), Varios-Virus(49-50), Varios(51-67), Complexos Nativos (68-70).







A partir das subfamílias da tripsina, quimotripsina, elastase e trombina complexadas com a ecotina podemos ver as diferenças apresentadas em 40 posições de IFR, como mostra a Tabela 9. Essas posições foram selecionadas a partir da Tabela 8A, lembrando que selecionamos somente as posições colunas em que ao menos uma das quatro subfamílias apresentava um padrão de tipo de aminoácidos de 50% para os membros da mesma.

**Tabela 9.** Posições dos IFRs para quatro subfamílias de serino proteases, ligadas a ecotina. Em cada posição (linha da tabela) podemos ver as diferenças dos tipos de resíduos entre cada uma das quatro subfamílias: tripsina, quimotripsina, elastase e trombinas, respectivamente. Os resíduos em uma determinada posição que estão presente em mais de 50% em uma dada subfamília são indicados pelo '\*'. As cores dos resíduos seguem o mesmo padrão da tabela anterior onde: verdes são polares, cinzas são pequenos hidrofóbicos, vermelhos são carregados negativamente e azuis são carregados positivamente (Taylor, 1997).

#	AA#	Tripsina	Quimotripsina	Elastase	Trombina
1	37)	R,E,M	D,Y	Y*,R	R*
2	37C)	-	-	-	P*
3	39)	Y,F	F,R	Y,R	E*,R
4	40)	P,H*	H*	H*	L*,F
5	41)	W,F*,G	F*,R,T	T*,S,F	L*
6	59)	Y	G*	-	-
7	59B)	-	-	N	Y*
8	59D)	-	-	-	P*
9	59E)	-	-	-	W*
10	59F)	-	-	-	D*
11	59G)	-	-	-	K*
12	60)	Y,K*	S	V,D	-
13	94)	Y*,F	Y*,W	W*,Y	-
14	96)	R,S,Y,G	S*,D	S,P,T	W*
15	97)	K,N,Y,D,R	L*,F*	D,V,N	R*
16	97A)	-	-	-	E*
17	98)	T*,G,N,L	T*,L*	D,S,N,Q	N*
18	99)	L*,I,N	I*,L,V	V,Y,L	L*
19	143)	N,T	L,S,R	K,R,L	N*
20	143D)	E	N	T	E*
21	143H)	G	G	G,N	T*
22	143Q)	T,Y,Q	T*	L*,I,S	Q*
23	171)	Q,S,D,A	Y,D	-	S*
24	172)	Y*	W*,Q	W*,M	T*
25	173)	P*,G	G*,W	G*	R*
26	174)	G*,S,D,N	T*	S*,A	I*
27	175)	A,K,Q,M,E	K*,R,T	T*,N	R*
28	177)	-	K*	K*	-
29	189)	D*	-	-	D*
30	190)	S*,T	S*,A	V,S	A*
31	192)	Q*	M*,H	N*,F,Q	E*
32	213)	V*	V*	T*,A,V	V*
33	215)	W*	W*,F*	F*,W	W*
34	216)	G*	G*	V*	G*
35	217)	Y*,N,D,S,H	S*	S*,V,R	E*
36	217A)	-	S*,G	S,R,S	-
37	219)	G	T,P	S,G	G*
38	221A)	R	S	-	R*
39	224)	Y,K,N,R	T,E,K	K*,Y	K*
40	226)	G*	D,T	-	-

De acordo com a Tabela 9 (inibidor ecotina) marcamos as posições que são mais restritivas em termos do tipo de aminoácido que podem ocupá-las (hidrofóbico, carregado, polar e glicina). A posição 40 é ocupada exclusivamente pela His na subfamília quimotripsina e elastase, e na tripsina ela é o aminoácido predominante em mais de 50% dos membros dessa subfamília (marcada com o '\*'), já na subfamília da trombina temos somente resíduos hidrofóbicos nesta posição. As posições 94, 172, 215 e 216 são ocupadas na maioria das vezes por resíduos polares e alguns hidrofóbicos. Algumas posições são ocupadas por um ou até mesmo dois tipos de resíduos em cada uma das quatro subfamílias (59, 59B, 60, 143D, 189, 216, 221A etc.).

Entretanto, existem posições menos restritivas que podem ser ocupadas por três diferentes classes de aminoácidos em uma ou em todas as subfamílias (posições 37, 41, 143, 171, 175 e 224). Por exemplo, a posição 97 é menos restritiva para as tripsinas (carregado positivamente: K, R; polar: N, Y; carregado negativamente: D) e elastases (polar: N; carregado negativamente: D; hidrofóbicos: V) enquanto que para as quimotripsinas (hidrofóbico: L, F) e trombinas (positivamente carregada: R) é bem restritiva.

Marcamos com cinza escuro aquelas posições de IFR que são populadas exclusivamente por um único tipo de resíduos ou por ausência no conjunto de IFR das subfamílias. Aquelas posições que foram ocupadas por um tipo predominante de aminoácidos, (posições que apresentavam um padrão 50% entre os membros de uma mesma subfamília) foram marcadas em cinza claro. Uma análise similar foi feita para as quatro subfamílias e os outros dois inibidores - BPTI com 36 posições do IFR disponíveis (Tabela 10) e ovomucoide do terceiro domínio com 34 posições do IFR disponíveis (Tabela 11).



**Tabela 10.** Posições dos IFRs para quatro subfamílias de serino proteases, ligadas a BPTI. As posições mostradas aqui seguem a mesma metodologia que a Tabela 9.

	AA#	Trypsin	Chymotrypsin	Elastase	Thrombin
1	37)	R, E, M	Y	Y*,R, L	R*
2	37C)	-	-	-	P*
3	39)	Y,G,F,S	F,S,R	Y,S,G,A,R	E*,R
4	40)	P,H*	H*	H*	L*,F
5	41)	F*,W,G	F*,R,T	T*,S,F	L*
6	59B)	G	G,N	Y*	-
7	59D)	-	-	-	P*
8	59E)	-	-	-	W*
9	59F)	-	-	-	D*
10	59G)	-	-	-	K*
11	59H)	-	-	-	N*
12	60)	K*,Y	S	D,V,S	F*
13	73)	I*	L*,Q	L,Q	R*
14	94)	Y*	Y,W	W*,Y	-
15	96)	R,S,Y,G	S*,D	S,P,T	W*
16	97)	K,N,Y,D,R	L*,F*	D,V,N	R*
17	97A)	-	-	-	E
18	99)	N,I,L*	I,L,V	V,Y,L	L*
19	141)	W*	-	-	W
20	142)	G	-	-	G*
21	143)	N,T	L,S,R	K,R,L	N*
22	143Q)	T,Y,Q	T*	L*,I,S	Q*
23	172)	-	W*	W*	-
24	174)	-	-	-	I*
25	189)	D*	S	-	D*
26	190)	S*,T	S*,A	A,V,S	A*
27	192)	Q*	M*,H	N*,F,Q	E*
28	213)	V*	V*	T*,A,V	V*
29	215)	W*	W*,F*	F*,W	W*
30	216)	G*	G*	V*,G	G*
31	217)	G*	S*	S*,V,R	E*
32	217A)	-	S*,G	S,R	-
33	219)	G*	T,P	S	G*
34	226)	G*	G,D,T	S	G*
35	227)	V*	V	-	F
36	228)	Y	Y*	F	Y*

**Tabela 11.** Posições dos IFRs para quatro subfamílias de serino proteases, ligadas a Ovomucoide. As posições mostradas seguem a mesma metodologia que a Tabela 8 e 9.

#	AA#	Trypsin	Chymotrypsin	Elastase	Thrombin
1	37)	R,E,S,M	D,L,Y	Y*,R,L	R*,A
2	37C)	-	-	-	P*
3	38)	N	-	R	Q*
4	39)	G,Y,F,S	F,S,R	Y,S,G,A,R	E*,R
5	40)	P,H*	H*	H*	L*
6	41)	F*,W,G	F*,R,T	T*,S,F	L*
7	59)	Y*,V	G	-	-
8	59B)	-	-	N	Y*
9	59D)	-	-	-	P*
10	59E)	-	-	-	W*
11	59F)	-	-	-	D*
12	59G)	-	-	-	K*
13	60)	K*,D,Y	V,L,S	V,S	F*
14	73)	R	-	-	R*
15	74)	E,N,K	-	-	T*
16	94)	Y	Y,W	W*,Y	-
17	97)	K,N,Y,D,R	F*	D,N	-
18	99)	L*,N,I	I,L,V	V,Y,L	L*
19	143)	N*,A,T	L,S,R	K,R,L	N*
20	143D)	E	N	T	E*
21	143H)	G	G	G,N	T*
22	143Q)	T,Y,Q	T*	L*,I,S	Q*
23	172)	Y*	W*,Q	W*	T
24	173)	-	G,W	G	R*
25	174)	-	-	A	I*
26	175)	K,Q,M,E	K,T	T*	-
27	189)	D*	-	-	D*
28	190)	S*,T	S*,A*	V,S	A*
29	192)	Q*	M*,H	N*,F,Q	E*
30	213)	V*	V*	T*,A,V	V*
31	215)	W*	W*,F*	F*,W*	W*
32	216)	G*	G*	V*,G	G*
33	217)	Y*,N,D,S,H	S*	S*,V,R	E*
34	217A)	-	S*,G	S,R	-

Para os dados apresentados nas Tabelas 10 e 11 (BPTI e ovomucoide do terceiro domínio, respectivamente) está claro que as posições 94 e 172 tem presente somente resíduos polares (nenhum hidrofóbico, como na Tabela 9), assim essas posições se tornaram mais restritivas. Comparando os dados das três tabelas pode-se observar que os complexos de serino proteases com inibidor ecotina totalizam 13 posições de IFR restritivas (cinza escuro), 10 meio restritivas (cinza claro) e 17 não restritivas, em um total de 40. Nos complexos com os inibidores BPTI e ovomucoide, os dados são, respectivamente: 14, 7 e 15 (36); 11, 7 e 16 (34). O ponto de vista percentual com relação a restrição dos três inibidores ecotina, BPTI e ovomucoide

ligado com as quatro subfamílias de serino proteases é mostrado na Tabela 12 . Os complexos formados entre serino proteases e o inibidor ovomucoide são em geral menos específicos do que os outros formados com a ecotina e a BPTI.

**Tabela 12.** Restrição ao tipo de resíduo presente nas interfaces entre quatro subfamílias de serino proteases e os inibidores protéicos, ecotina, BPTI e Ovomucoide.

%	Restritiva	Meio Restritiva	Não Restritiva
<b>Ecotina</b>	32,5%	25%	42,5%
<b>BPTI</b>	39%	19,5%	41,5%
<b>Ovomucoide</b>	32%	20.5%	47%

## 2. Superfície versus Interface

Continuando a analisar as interfaces, vamos explorar agora a diferença entre a área total da superfície e a interface em termos de polaridade e hidrofobicidade. A área obliterada (interface) e a área da superfície livre, que não está incluída na interface, são comparadas com respeito ao tipo de resíduo que as ocupam. Novamente usaremos as classes de aminoácidos a seguir: carregados, glicinas, hidrofóbico e polar.

A análise de dados da Figura 24 mostra que a maioria das interfaces, em todas as subfamílias de serino proteases ligadas com BPTI, ecotina ou ovomucoide, tem em geral preferência por resíduos polares e glicinas. Ao contrário, a maioria das superfícies livre das enzimas (não incluídas a área da interface) mostram preferência por resíduos carregados. Algumas subfamílias não seguem esse padrão. Surpreendentemente, na interface da subfamília da Estreptogrisina observamos uma alta presença de resíduos de glicina para complexos com os três tipos de inibidor (Barras cinza negativas na Figura 24). Elastases mostram uma diferente constituição de IFR para BPTI, onde uma pequena predominância de resíduos hidrofóbicos e carregados é registrada na interface (com ressalva para ecotina e ovomucoide, onde uma pequena predominância de polares e glicinas são observados). Além disso, a

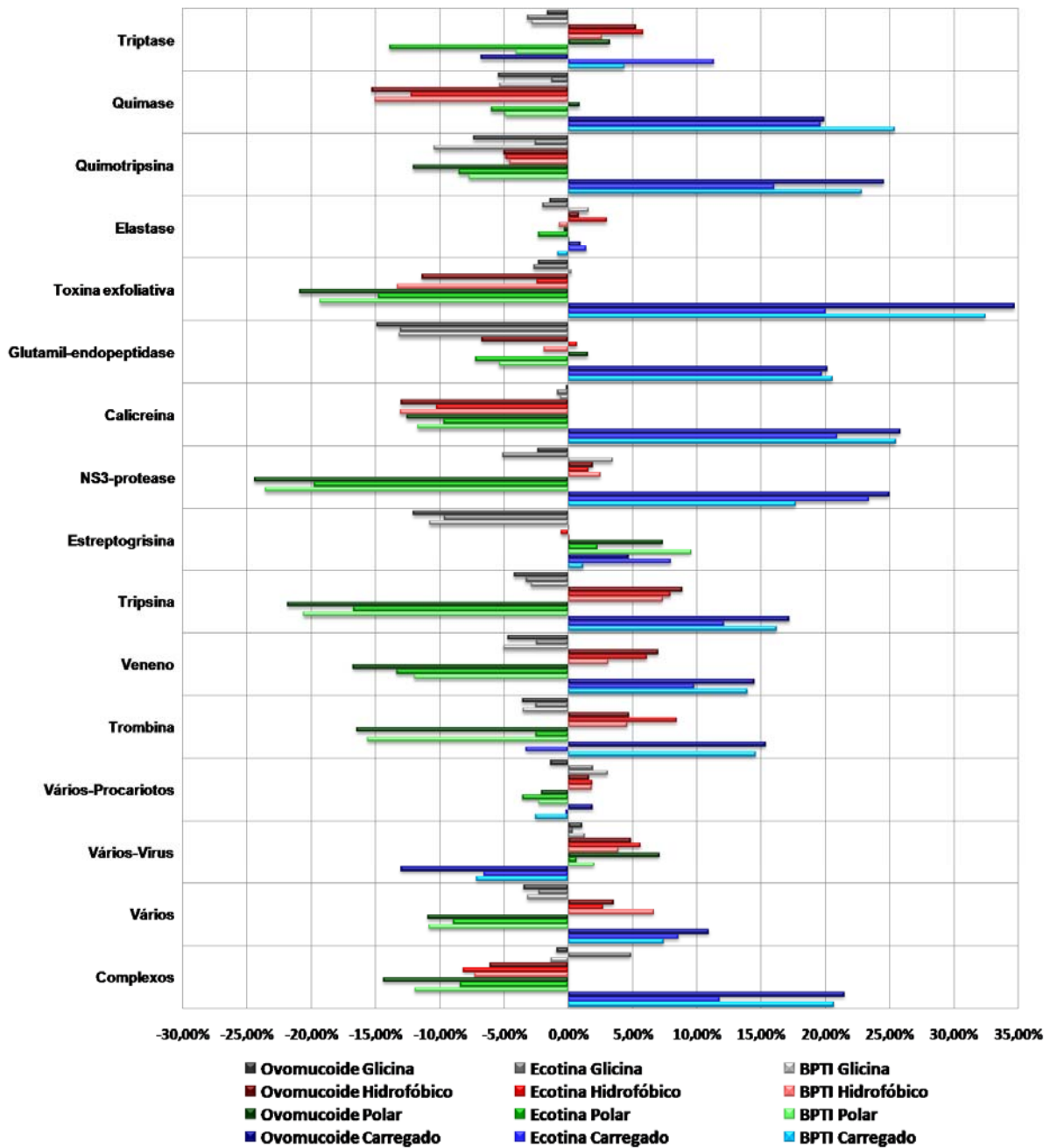
distribuição predominante de resíduos carregados na interface é observada para a subfamília “vários-vírus” para os três tipos de inibidores. Em geral, a parte superior da Figura 24 para a interface mostra uma forte presença de resíduos hidrofóbicos (junto com algumas glicinas e uns polares), enquanto a parte de inferior Figura 24 mostra uma ausência de resíduos hidrofóbicos na interface, parte esquerda do gráfico. Uma rápida olhada na parte inferior da Figura 24 para a área livre da superfície da enzima (parte da direita) mostra um aumento da presença de resíduos hidrofóbicos.

Em geral, a área livre da superfície da enzima mostra uma prevalência de resíduos carregados, enquanto a interface contém mais resíduos polares e glicinas (com algumas exceções descritas acima). Assim, o bolsão do IFR não é formado somente por interações hidrofóbicas como poderia se esperar. O bolsão do IFR em geral possui um ambiente preferencialmente polar. Entretanto a interface contém poucos resíduos carregados, salvo uma ou duas exceções. Algumas interessantes diferenças, com respeito a área ocupada por 4 classes de aminoácidos na interface e na superfície, podem ser observadas para a tripsina, quimotripsina, elastase e trombinas, como apresentadas na Tabela 13.

**Tabela 13.** Porcentagem da área da interface ocupada por quatro classes de aminoácidos. As serino proteases estão ligadas a ecotina, BPTI e ovomucoide. Os números entre parêntesis mostram a quantidade de complexos enzima /inibidor para cada uma das subfamílias.

Complexos	Área Ocupada				
	Carregado (%)	Glicina (%)	Hidrofóbico (%)	Polar (%)	(Carregado +Polar (%))
Tripsina/ Ecotina (9)	15	10	13	62	77
Quimotripsina/ Ecotina (4)	18	6	30	46	64
Elastase/ Ecotina (5)	19	9	22	50	69
Trombina/ Ecotina (5)	53	6	12	29	82
Tripsina/ BPTI (9)	10	10	14	66	76
Quimotripsina/ BPTI (4)	11	8	35	46	57
Elastase/ BPTI (5)	21	6	25	48	69
Trombina/ BPTI (5)	37	6	16	41	78
Tripsina/ Ovomucoide(9)	10	11	12	67	77
Quimotripsina /Ovomc.(4)	10	9	32	49	59
Elastase/ Ovomucoide (5)	20	9	24	47	67
Trombina/ Ovomucoide(5)	37	7	15	41	78

### Diferença entre as porcentagens de Superfície Livre - Interface



**Figura 24.** Diferença das classes dos resíduos de ocupação, entre o total da superfície livre da enzima versus interface. Nesta figura apresentamos a diferença na porcentagem da área de ocupação da superfície livre da enzima e a porcentagem do IFR, para todas 70 serino proteases ligadas ao inibidor Ecotina (A), BPTI (B) e Ovomucoide do terceiro domínio (C). As enzimas foram classificadas pelas subfamílias a seguir: Triptase (2), Quimase (2), Quimotripsina (4), Elastase (5), Exfoliative toxina (2), Glutamil endopeptidase (3), Calicreína (4), NS3 protease (3), Estreptogrisina (2), Tripsina (9), Veneno (2), Trombina (5), Vários-Procaríotos (5), Vários-Vírus (2), algumas serino proteases que não se encaixavam em nenhuma subfamília definida nos chamamos de Vários (17) e os Complexos (3). A média dos valores da porcentagem de ocupação foi feita a partir dos membros de cada subfamília. O número entre os parêntesis mostram a quantidade de enzimas em cada subfamília. As barras do lado esquerdo do gráfico indicam as classes de resíduo que são mais frequentemente encontradas na interface do que na superfície. As barras do lado direito do gráfico indicam as classes de resíduos que são mais frequentes na superfície do que na interface.

Ainda que todas as quatro subfamílias de serino proteases mostrem na prática uma porcentagem quase idêntica para área total da superfície que foi obliterada na formação do complexo *in silico* com os inibidores BPTI, ecotina, e ovomucoide, as porcentagens das classes dos resíduos que foram obliterados variam significativamente. A composição de área obliterada carregada e polar na tripsina/BPTI é de 76%, na quimotripsina/BPTI é 57%, na elastase/BPTI é 69%, enquanto que para a trombina/BPTI a soma é de 78%. Para os três inibidores examinados, as subfamílias da trombina e tripsina tem a mais alta porcentagem de área polar obliterada, enquanto a elastase e quimotripsina são menores. Os detalhes para a diferença de ocupação do total da superfície da enzima e interface com a BPTI, ecotina e Ovomucoide são apresentados na Tabela 13.

Essas diferenças claramente corroboram com a função que cada protease desempenha; tripsina (e trombina) precisam acessar cargas positivas em um polipeptídeo onde a enzima irá clivar; para contrabalançar o polipeptídeo a superfície do IFR possui um perfil carregado/polar. Quimotripsinas contrabalançam o espaço obliterado com resíduos hidrofóbicos (e polares) do polipeptídeo a ser clivado, precisa estrategicamente acomodar glicinas dentro da superfície do seu IFR. Elastase tem uma presença significativa de resíduos hidrofóbicos entre os IFR, o que estrategicamente complementaria o seu interior pequeno seria também aminoácidos hidrofóbicos no sítio de clivagem de um polipeptídeo.

A tendência geral das características das interfaces descreve sobre o que será complementado em relação à ocupação específica das posições dos IFR, como já discutimos anteriormente, como se obtêm um desenho completo da especificidade das subfamílias de serino proteases.

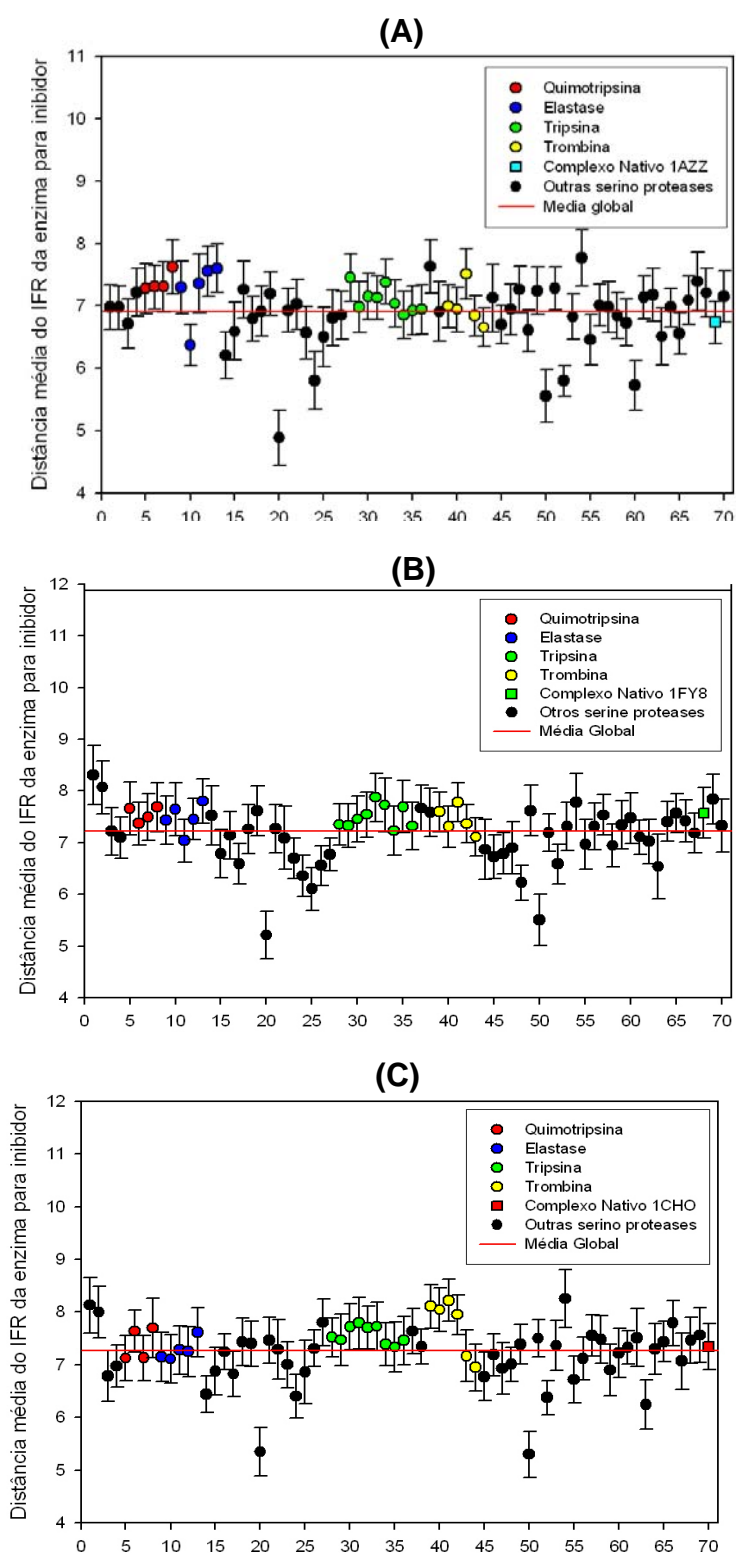
### **3. Impedimentos Estéricos**

Analisando o nosso procedimento de *hard body docking* em termos de formar modelos de complexos plausíveis, quanto a presença de impedimentos estéricos

(steric clash), decidiu-se examinar a extensão de algumas interações não favoráveis. De fato, essas interações podem representar um severo impedimento que pode ser a principal razão para fazer com que esses complexos não sejam criados *in vivo*, por exemplo, uma ausência de plasticidade na enzima, o que poderia em princípio contrabalançar para o impedimento estérico.

Calculamos as distâncias entre os resíduos da interface da enzima e do inibidor BPTI (Figura 25A), ecotina (Figura 25B) e ovomucoide do terceiro domínio (Figura 25C) complexados com as serino proteases. Todas as 70 serino proteases estão presentes, usamos a mesma numeração da Tabela 6 para o eixo X, os membros de uma mesma subfamília estão agrupados seqüencialmente. As cores das bolinhas que representam cada serino protease, seguiram as cores das quatro subfamílias das Tabelas 9, 10, 11. O eixo Y mostra as distâncias, medidas em angstroms, com o desvio padrão. As distâncias são a média dos valores obtidos pelo cálculo da soma das menores distâncias CA-CA (carbono alfa) para cada resíduo pertencente ao IFR da enzima aos resíduos do IFR do inibidor, e então divididos pelo número de resíduos do IFR da enzima.

A linha horizontal representa a média global que está sempre perto dos 7.0 Å para todos os três tipos de inibidores complexados. Em outro artigo (da Silveira *et al.*, 2009) do nosso grupo de pesquisa foi demonstrado que 7.0 Å é um *cutoff* de distância que otimiza o isolamento, em termos de resíduo, a primeira ordem de contatos (a primeira vizinhança) em proteína com um bom empacotamento globular. Isso foi uma coincidência em que a média da distância global entre os CA dos complexos enzima-inibidor tende também aos 7.0 Å. Isso confirma que enzimas e inibidores formam um bom empacotamento como aqueles encontrados no núcleo de proteínas globulares. Assim é plausível pensar que complexos em que distâncias muito menores estão presentes (abaixo da média global) contêm um excesso de impedimentos estéricos. Esses complexos poderão ser provavelmente submetidos a uma minimização e dinâmica molecular para ordenar e rearranjar esses resíduos, diminuindo assim os impedimentos.



**Figura 25.** Proximidade entre resíduos pertencentes aos IFRs de 70 serino proteases e os resíduos pertencente aos IFRs dos inibidores Ecotina(A), BPTI (B) e Ovomucoide (C). Cada ponto no gráfico representa a média e erro padrão para as menores distâncias encontradas entre cada resíduo do IFR da enzima aos resíduos do IFR do inibidor. A linha horizontal representa a medial global para todas menores distâncias.



## 4. Validação da Metodologia

Para validar essa metodologia decidimos testar se ela tinha bons resultados com alguns dos complexos que já existem naturalmente, ou seja, que existem dentro do PDB. Foram selecionados 16 complexos com o inibidor BPT; testamos com essa família de inibidores, pois ela era a que tinha um maior número de complexos nativos (veja a lista desse código PDBs na Tabela 14). O que queríamos era checar o quanto realista ou quão boa poderia ser a predição dos resíduos do IFR aplicando o mesmo procedimento usado nesse trabalho.

A partir desses 16 complexos reais nós fizemos os mesmos procedimentos já descritos anteriormente; primeiramente alinhamos todas as 16 cadeias das enzimas de cada um desses complexos no PrISM, depois alinhamos o complexo 1FY8\_original com o complexo 1FY8\_prism no software PyMol. Esse alinhamento mantinha as coordenadas do arquivo PDB do PrISM fixas sendo sobreposta pelo complexo nativo; assim, as coordenadas do inibidor foram relocadas junto com a molécula nativa. As coordenadas do inibidor foram então extraídas a partir do arquivo PDB de saída do PyMol e propagadas para todos as outras 16 cadeias da enzimas.

Depois de termos criado 16 complexos *in silico* baseado no complexo 1FY8, também rodamos o algoritmo *surfv* para determinar os resíduos que faziam parte da interface (IFR). O algoritmo *surfv* também foi usado para determinar os resíduos que eram da interface nos complexos nativos. Assim a partir desses dados foi feita uma comparação entre o IFR dos complexos gerados pelo “hard body docking” e o IFR dos complexos naturais. Para fazer essa comparação decidimos usar matriz de confusão (Kohavi & Provost, 1998). A matriz de confusão de uma hipótese *h* oferece uma medida efetiva do modelo de classificação, ao mostrar o número de classificações corretas *versus* as classificações preditas para cada classe, sobre um conjunto de exemplos *T*. No nosso caso haviam apenas duas classes, complexos naturais e complexos *in silico*. Assim sendo precisávamos determinar os verdadeiros positivos (VP), os verdadeiros negativos (VN), os falsos negativos (FN) e os falsos

negativos. Os VP são os resíduos de aminoácidos do IFR que estão presente no complexo nativo e no complexo *in silico*, os VN são os resíduos de aminoácidos que não fazem parte do IFR e estão presentes no complexo nativo e no complexo *in silico*, os FP são resíduos do IFR que estão presente no complexo *in silico* e não no complexo nativo e os FN são resíduos do IFR que não estão presente no complexo *in silico* e estão no complexo natural. A Tabela 14 mostra esses valores para o controle 1FY8 e cada um dos 16 complexos complementares. Em todos eles a acurácia (AC) ficou entre 0,97 e 0,98.

**Tabela 14.** Valores usados para calcular matriz de confusão (VP,VN, FP, FN) e acurácia (AC) do 1FY8 (controle) e dos outros 16 complexos complementares.

<b>PDB ID</b>	<b>Subfamília</b>	<b>Organismo</b>	<b>VP</b>	<b>VN</b>	<b>FP</b>	<b>FN</b>	<b>AC</b>
1FY8	Tripsina	<i>Bos taurus</i>	44	244	0	0	1
3TGI	Tripsina	<i>Bos taurus</i>	42	239	5	0	0.98
1TPA	Tripsina	<i>Bos taurus</i>	44	240	1	3	0.98
3BTK	Tripsina	<i>Bos taurus</i>	43	241	1	4	0.98
3TPI	Tripsina	<i>Bos taurus</i>	43	242	1	2	0.98
2PTC	Tripsina	<i>Bos taurus</i>	44	241	0	3	0.98
2TGP	Tripsina	<i>Bos taurus</i>	45	240	0	3	0.98
4TPI	Tripsina	<i>Bos taurus</i>	44	239	1	4	0.98
2TPI	Tripsina	<i>Bos taurus</i>	44	241	1	2	0.98
1BZX	Tripsina	<i>Bos taurus</i>	43	242	2	1	0.98
1BRB	Tripsina	<i>Rattus novergicos</i>	41	236	2	2	0.98
1FAK	Fator de Coagulação VIII	<i>Homo sapiens</i>	44	247	6	0	0.97
1CBW	Quimotripsina	<i>Bos taurus</i>	44	238	5	1	0.97
1MTN	Quimotripsina	<i>Bos taurus</i>	45	239	1	3	0.98
2KAI	Calicreína	<i>Bos taurus</i>	57	224	0	4	0.98
1EAW	Matriptase MTSP1	<i>Homo sapiens</i>	42	241	4	1	0.98

## Conclusão

A análise de dados produzida durante esse trabalho é apresentada como uma ferramenta para entender a especificidade baseada nos resíduos formadores da interface (*Interface Forming Residues* - IFR) desenhada através do perfil do alinhamento estrutural.

Baseado na predição de cada IFR dos complexos formados, pôde-se indicar as diferenças entre várias proteases com respeito à especificidade, bem como abrir a possibilidade de descrever as diferenças e suas características físico-químicas.

Nesse trabalho foi possível identificar quais eram os sítios restritivos, meio restritivos e não restritivos na interface de uma subfamília de serino protease, qual era a natureza da interface da enzima quanto às classes de aminoácidos que a ocupava e como essa natureza pode contribuir para determinar a especificidade.

Através desse trabalho foi evidenciado o que a taxonomia da natureza tem oferecido uma importante visão no relacionamento estrutura/função para a família de proteínas serino proteases. A escolha e o foco da nossa atenção na interface é uma simples consequência de que todos os parâmetros de funcionalidade crítica devem estar concentrada nela (Perona & Craik, 1995). A superposição do *backbone* no alinhamento estrutural de serino proteases claramente corrobora para essa idéia, a natureza usa essencialmente o mesmo enovelamento e já encontraram variações de muitas atividades específicas por variar simplesmente a superfície (Page & Cera, 2007). O que provavelmente descreve uma aplicação de um particular “toque final” sobre um esqueleto em comum. O leitor pode encontrar na literatura uma grande quantidade de dados onde os autores tentam modificar a funcionalidade das serino proteases pelo tratamento de uma funcionalidade específica em local e escopo limitado, o que poderia mudar caso ele entendesse mais do local geral onde essas atividades acontecem. Nossas pesquisas e abordagens corrobora com resultados apresentados por Ma *et al.*, 2005 e Novozymes Biotech, Inc. (Davis, CA, US) que

possui uma patente de tecnologia (Microbial trypsin mutants having chymotrypsin activity - United States Patent 20050037368).

Baseado nos resultados obtidos, nós pode-se equilibrar e agrupar em um futuro próximo, os dados das interfaces da super família da serino proteases como uma expansão de sequência, estrutura e informação funcional sobre os resíduos formadores da interface em volta do sitio catalítico. Uma combinação de imagens gráficas e dados numéricos foram usados para complementar a análise da relação estrutura e função.

## **Perspectivas**

Aqui serão comentadas as perspectivas e algumas limitações deste trabalho, uma vez que estas limitações podem servir para desenvolver uma segunda etapa desse projeto.

As pessoas que trabalham com proteínas devem saber que selecionar dados não viciados ou não redundantes é um dos grandes desafios da Bioinformática. Todas as hipóteses, evidências e conclusões geradas nesse trabalho foram em cima de uma base dados que nós selecionamos. Para gerarmos essa base de dados, criamos um critério de seleção e filtragem sobre o conjunto universo das serino proteases que existem dentro do PDB e SCOP (o que em geral é apenas uma pequena amostragem dos tipos de proteínas presente nos seres vivos). Aproveitando a metodologia empregada nesse trabalho, poderíamos trabalhar com outras famílias.

Nesse trabalho nós decidimos utilizar somente as serino proteases que tinham um tipo de enovelamento parecido com as tripsinas; em trabalhos futuros poderíamos trabalhar com as subtilisinas, que é uma subfamília de serino protease que possui um enovelamento diferente encontradas em bactérias (Figura 3B), ou trabalhar com outras famílias, por exemplo, as cisteíno proteases.

Uma meta muito importante seria ampliar o escopo do nosso trabalho para as outras 9 famílias de inibidores que possui estruturas conhecidas e inibem as serino proteases, uma vez que nesse trabalho analisamos somente 3 inibidores.

Como os complexos utilizados em nosso trabalho foram gerados através do *hard body docking*, esse método possui algumas limitações, em algumas análises que nós fizemos da distância dos resíduos enzima/inibidor, notamos que alguns complexos tinham impedimento estérico entre resíduos. Para resolvermos o problema do impedimento estérico nós poderíamos fazer uma minimização e dinâmica para os resíduos se acomodarem ou rearranjarem.

Dando continuidade ao tema dessa pesquisa, poderíamos montar um banco de dados na *web* com todas as informações geradas nesse trabalho desde a base de dados de serino proteases não redundante, os alinhamentos estruturais, as interfaces encontradas e as análises gráficas.

Outro item que gostaríamos de ter implementado neste trabalho é uma análise comparativa e criteriosa entre a nossa metodologia e outra proposta na literatura. O principal problema que enfrentamos foi conseguir programas de uso aberto para que pudéssemos fazer os testes com a mesma base de dados que apresentamos. A maioria dos autores não disponibiliza o software e apresenta os resultados em bases específicas e pré-computadas em interfaces web. Acreditamos que para esta análise seria necessário eleger algumas das metodologias mais interessantes e tentar conseguir os softwares dos autores ou, alternativamente, reimplementá-los.

## Referências Bibliográficas

- 1992, *Enzyme Nomenclature*: San Diego: Academic Press.
- 2008, The universal protein resource (UniProt): *Nucleic Acids Res*, v. 36, p. D190-5.
- Barrett, A. J., N. D. Rawlings, & J. F. Woessner, 2004, *Handbook of Proteolytic Enzymes*: 2nd ed, vol. 1. Amsterdam: Elsevier/ Academic Press.
- Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, & P. E. Bourne, 2000, The Protein Data Bank: *Nucleic Acids Res*, v. 28, p. 235-42.
- Blow, D. M., & T. A. Steitz, 1970, X-ray diffraction studies of enzymes: *Annu Rev Biochem*, v. 39, p. 63-100.
- Bode, W., & R. Huber, 1992, Natural protein proteinase inhibitors and their interaction with proteinases: *Eur J Biochem*, v. 204, p. 433-51.
- Bone, R., D. Frank, C. A. Kettner, & D. A. Agard, 1989a, Structural analysis of specificity: alpha-lytic protease complexes with analogues of reaction intermediates: *Biochemistry*, v. 28, p. 7600-9.
- Bone, R., J. L. Silen, & D. A. Agard, 1989b, Structural plasticity broadens the specificity of an engineered protease: *Nature*, v. 339, p. 191-5.
- Carrell, R., & J. Travis, 1985, alpha1-Antitrypsin and the serpins: Variation and countervariation: *Trends Biochem Sci*, v. 10, p. 20-24.
- Cheng, Z. Y., & Q. Z. Xue, 2003, The structure and regulation of plant proteinase inhibitor genes and their strategy in pest control: *Yi Chuan Xue Bao*, v. 30, p. 790-6.
- Chung, C. H., H. E. Ives, S. Almeda, & A. L. Goldberg, 1983, Purification from *Escherichia coli* of a periplasmic protein that is a potent inhibitor of pancreatic proteases: *J Biol Chem*, v. 258, p. 11032-8.
- Da Silva, M. C., M. F. de Sa, M. J. Chrispeels, R. C. Togawa, & G. Neshich, 2000, Analysis of structural and physico-chemical parameters involved in the specificity of binding between alpha-amylases and their inhibitors: *Protein Eng*, v. 13, p. 167-77.
- da Silveira, C. H., D. E. Pires, R. C. Minardi, C. Ribeiro, C. J. Veloso, J. C. Lopes, W. Meira, Jr., G. Neshich, C. H. Ramos, R. Habesch, & M. M. Santoro, 2009,

- Protein cutoff scanning: A comparative analysis of cutoff dependent and cutoff free methods for prospecting contacts in proteins: *Proteins*, v. 74, p. 727-43.
- Fenton, J. W., 2nd, F. A. Ofose, D. V. Brezniak, & H. I. Hassouna, 1998, Thrombin and antithrombotics: *Semin Thromb Hemost*, v. 24, p. 87-91.
- Finkenstadt, W. R., & M. Laskowski, Jr., 1965, Peptide Bond Cleavage on Trypsin-trypsin Inhibitor Complex Formation: *J Biol Chem*, v. 240, p. 962-3.
- Folkers, P. J., G. M. Clore, P. C. Driscoll, J. Dodt, S. Kohler, & A. M. Gronenborn, 1989, Solution structure of recombinant hirudin and the Lys-47----Glu mutant: a nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing study: *Biochemistry*, v. 28, p. 2601-17.
- Fujinaga, M., A. R. Sielecki, R. J. Read, W. Ardel, M. Laskowski, Jr., & M. N. James, 1987, Crystal and molecular structures of the complex of alpha-chymotrypsin with its inhibitor turkey ovomucoid third domain at 1.8 Å resolution: *J Mol Biol*, v. 195, p. 397-418.
- Grasberger, B. L., G. M. Clore, & A. M. Gronenborn, 1994, High-resolution structure of *Ascaris* trypsin inhibitor in solution: direct evidence for a pH-induced conformational transition in the reactive site: *Structure*, v. 2, p. 669-78.
- Greer, J., 1990, Comparative modeling methods: application to the family of the mammalian serine proteases: *Proteins*, v. 7, p. 317-34.
- Greighton, T. E., 1983, *Proteins: structures and molecular principles*: New York : W. H. Freeman and Co.
- Hedstrom, L., 2002, Serine protease mechanism and specificity: *Chem Rev*, v. 102, p. 4501-24.
- Hennighausen, L. G., & A. E. Sippel, 1982, Mouse whey acidic protein is a novel member of the family of 'four-disulfide core' proteins: *Nucleic Acids Res*, v. 10, p. 2677-84.
- Hooper, N. M., 2002, *Proteases in Biology and Medicine*: London: Portland Press.
- Hsu, J. T., H. C. Wang, G. W. Chen, & S. R. Shih, 2006, Antiviral drug discovery targeting to viral proteases: *Curr Pharm Des*, v. 12, p. 1301-14.
- Hubbard, S. J., 1998, The structural aspects of limited proteolysis of native proteins: *Biochim Biophys Acta*, v. 1382, 191-206 p.

- Hubbard, S. J., S. F. Campbell, & J. M. Thornton, 1991, Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors: *J Mol Biol*, v. 220, p. 507-30.
- Hubbard, S. J., F. Eisenmenger, & J. M. Thornton, 1994, Modeling studies of the change in conformation required for cleavage of limited proteolytic sites: *Protein Sci*, v. 3, p. 757-68.
- Itoh, H., H. Ide, N. Ishikawa, & Y. Nawa, 1994, Mast cell protease inhibitor, trypstatin, is a fragment of inter-alpha-trypsin inhibitor light chain: *J Biol Chem*, v. 269, p. 3818-22.
- Jeremy, M. B., John. L T., & S. Lubert, 2002, *Biochemistry: 5 ed.* New York: W H Freeman.
- Kohavi, R., & F. Provost, 1998, Glossary of Terms: *Machine Learning*, v. 30, p. 271-274.
- Kunitz, M., & J. H. Northrop, 1935, Crystalline Chymo-Trypsin and Chymo-Trypsinogen : I. Isolation, Crystallization, and General Properties of a New Proteolytic Enzyme and Its Precursor: *J Gen Physiol*, v. 18, p. 433-58.
- Kunitz, M., & J. H. Northrop, 1936, Isolation from Beef Pancreas of Crystalline Trypsinogen, Trypsin, a Trypsin Inhibitor, and an Inhibitor-Trypsin Compound: *J Gen Physiol*, v. 19, p. 991-1007.
- Lapatto, R., U. Kregel, H. A. Schreuder, A. Arkema, B. de Boer, K. H. Kalk, W. G. Hol, P. D. Grootenhuis, J. W. Mulders, R. Dijkema, H. J. Theunissen, & B. W. Dijkstra, 1997, X-ray structure of antistasin at 1.9 Å resolution and its modelled complex with blood coagulation factor Xa: *Embo J*, v. 16, p. 5151-61.
- Laskowski Jr, M., M. A. Qasim, & M. L. Stephen, 2000, Interaction of standard mechanism, canonical protein inhibitors with serine proteinases, *in C. Kleanthous, B. D. Hames, and D. M. Glover, eds., Protein-Protein recognition:* New York, Oxford University Press, p. 228-79.
- Laskowski, M., Jr., & I. Kato, 1980, Protein inhibitors of proteinases: *Annu Rev Biochem*, v. 49, p. 593-626.
- Laskowski, M., & M. A. Qasim, 2000, What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes?: *Biochim Biophys Acta*, v. 1477, p. 324-37.
- Lehninger, A. L., D. L. Nelson, & M. M. Cox, 1993, *Principles of Biochemistry: 2.ed.* New York : Worth Publisher.



- Leinonen, R., F. G. Diez, D. Binns, W. Fleischmann, R. Lopez, & R. Apweiler, 2004, UniProt archive: Bioinformatics, v. 20, p. 3236-7.
- Lindqvist, A., & B. Akerstrom, 1996, Bovine alpha 1-microglobulin/bikunin. Isolation and characterization of liver cDNA and urinary alpha 1-microglobulin: *Biochim Biophys Acta*, v. 1306, p. 98-106.
- Loebermann, H., R. Tokuoka, J. Deisenhofer, & R. Huber, 1984, Human alpha 1-proteinase inhibitor. Crystal structure analysis of two crystal modifications, molecular model and preliminary analysis of the implications for function: *J Mol Biol*, v. 177, p. 531-57.
- Ma, W., C. Tang, & L. Lai, 2005, Specificity of trypsin and chymotrypsin: loop-motion-controlled dynamic correlation as a determinant: *Biophys J*, v. 89, p. 1183-93.
- Marshall, D. L., & A. L. Harvey, 1992, Protease inhibitor homologues of dendrotoxin do not bind to dendrotoxin acceptors on synaptosomal membranes or facilitate neuromuscular transmission: *Biol Chem Hoppe Seyler*, v. 373, p. 707-14.
- McGrath, M. E., S. A. Gillmor, & R. J. Fletterick, 1995, Ecotin: lessons on survival in a protease-filled world: *Protein Sci*, v. 4, p. 141-8.
- McGuire, J. J., 2003, Anticancer antifolates: current status and future directions: *Curr Pharm Des*, v. 9, p. 2593-613.
- Murzin, A. G., S. E. Brenner, T. Hubbard, & C. Chothia, 1995, SCOP: a structural classification of proteins database for the investigation of sequences and structures: *J Mol Biol*, v. 247, p. 536-40.
- Neshich, G., L. C. Borro, R. H. Higa, P. R. Kuser, M. E. Yamagishi, E. H. Franco, J. N. Krauchenco, R. Fileto, A. A. Ribeiro, G. B. Bezerra, T. M. Velludo, T. S. Jimenez, N. Furukawa, H. Teshima, K. Kitajima, A. Bava, A. Sarai, R. C. Togawa, & A. L. Mancini, 2005a, The Diamond STING server: *Nucleic Acids Res*, v. 33, p. W29-35.
- Neshich, G., A. L. Mancini, M. E. Yamagishi, P. R. Kuser, R. Fileto, I. P. Pinto, J. F. Palandrani, J. N. Krauchenco, C. Baudet, A. J. Montagner, & R. H. Higa, 2005b, STING Report: convenient web-based application for graphic and tabular presentations of protein sequence, structure and function descriptors from the STING database: *Nucleic Acids Res*, v. 33, p. D269-74.
- Neshich, G., I. Mazoni, S. R. Oliveira, M. E. Yamagishi, P. R. Kuser-Falcao, L. C. Borro, D. U. Morita, K. R. Souza, G. V. Almeida, D. N. Rodrigues, J. G. Jardine, R. C. Togawa, A. L. Mancini, R. H. Higa, S. A. Cruz, F. D. Vieira, E. H. Santos,

- R. C. Melo, & M. M. Santoro, 2006, The Star STING server: a multiplatform environment for protein structure analysis: *Genet Mol Res*, v. 5, p. 717-22.
- Neshich, G., R. C. Togawa, A. L. Mancini, P. R. Kuser, M. E. Yamagishi, G. Pappas, Jr., W. V. Torres, T. Fonseca e Campos, L. L. Ferreira, F. M. Luna, A. G. Oliveira, R. T. Miura, M. K. Inoue, L. G. Horita, D. F. de Souza, F. Dominiquini, A. Alvaro, C. S. Lima, F. O. Ogawa, G. B. Gomes, J. F. Palandrani, G. F. dos Santos, E. M. de Freitas, A. R. Mattiuz, I. C. Costa, C. L. de Almeida, S. Souza, C. Baudet, & R. H. Higa, 2003, STING Millennium: A web-based suite of programs for comprehensive and simultaneous analysis of protein structure and sequence: *Nucleic Acids Res*, v. 31, p. 3386-92.
- Neurath, H., 1989, Proteolytic processing and physiological regulation: *Trends Biochem Sci*, v. 14, p. 268-71.
- Novotny, J., & R. E. Brucoleri, 1987, Correlation among sites of limited proteolysis, enzyme accessibility and segmental mobility: *FEBS Lett*, v. 211, p. 185-9.
- Odani, S., T. Koide, & T. Ono, 1986, Wheat germ trypsin inhibitors. Isolation and structural characterization of single-headed and double-headed inhibitors of the Bowman-Birk type: *J Biochem*, v. 100, p. 975-83.
- Page, M. J., & E. D. Cera, 2007, Is it possible to transform trypsin to thrombin?: *IUBMB Life*, v. 59, p. 413-4.
- Pasternak, A., A. White, C. J. Jeffery, N. Medina, M. Cahoon, D. Ringe, & L. Hedstrom, 2001, The energetic cost of induced fit catalysis: Crystal structures of trypsinogen mutants with enhanced activity and inhibitor affinity: *Protein Sci*, v. 10, p. 1331-42.
- Perona, J. J., & C. S. Craik, 1995, Structural basis of substrate specificity in the serine proteases: *Protein Sci*, v. 4, p. 337-60.
- Perona, J. J., C. A. Tsu, C. S. Craik, & R. J. Fletterick, 1997, Crystal structure of an ecotin-collagenase complex suggests a model for recognition and cleavage of the collagen triple helix: *Biochemistry*, v. 36, p. 5381-92.
- Radisky, E. S., & D. E. Koshland, Jr., 2002, A clogged gutter mechanism for protease inhibitors: *Proc Natl Acad Sci U S A*, v. 99, p. 10316-21.
- Rawlings, N. D., & A. J. Barrett, 1993, Evolutionary families of peptidases: *Biochem J*, v. 290 ( Pt 1), p. 205-18.
- Rawlings, N. D., F. R. Morton, C. Y. Kok, J. Kong, & A. J. Barrett, 2008, MEROPS: the peptidase database: *Nucleic Acids Res*, v. 36, p. D320-5.

- Richardson, M., 1991, Seed storage proteins: the enzyme inhibitors. *Methods in Plant Biochemistry*, p. 259-305.
- Ritonja, A., T. H. Coetzer, R. N. Pike, & C. Dennison, 1996, The amino acid sequences, structure comparisons and inhibition kinetics of sheep cathepsin L and sheep stefin B: *Comp Biochem Physiol B Biochem Mol Biol*, v. 114, p. 193-8.
- Rydel, T. J., K. G. Ravichandran, A. Tulinsky, W. Bode, R. Huber, C. Roitsch, & J. W. Fenton, 2nd, 1990, The structure of a complex of recombinant hirudin and human alpha-thrombin: *Science*, v. 249, p. 277-80.
- Rydel, T. J., A. Tulinsky, W. Bode, & R. Huber, 1991, Refined structure of the hirudin-thrombin complex: *J Mol Biol*, v. 221, p. 583-601.
- Sakai, M., S. Odani, & T. Ikenaka, 1980, Importance of the carboxyl-terminal four amino acid residues in the inhibitory activity of *Streptomyces subtilisin* inhibitor (with a revision of its carboxyl-terminal sequence): *J Biochem*, v. 87, p. 891-8.
- Schechter, I., & A. Berger, 1967, On the size of the active site in proteases. I. Papain: *Biochem Biophys Res Commun*, v. 27, p. 157-62.
- Schlott, B., J. Wohnert, C. Icke, M. Hartmann, R. Ramachandran, K. H. Guhrs, E. Glusa, J. Flemming, M. Gorchach, F. Grosse, & O. Ohlenschlager, 2002, Interaction of Kazal-type inhibitor domains with serine proteinases: biochemical and structural studies: *J Mol Biol*, v. 318, p. 533-46.
- Skarzynski, T., 1992, Crystal structure of alpha-dendrotoxin from the green mamba venom and its comparison with the structure of bovine pancreatic trypsin inhibitor: *J Mol Biol*, v. 224, p. 671-83.
- Song, H. K., & S. W. Suh, 1998, 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*, v. 275, p. 347-63.
- Sottrup-Jensen, L., 1989, Alpha-macroglobulins: structure, shape, and mechanism of proteinase complex formation: *J Biol Chem*, v. 264, p. 11539-42.
- Sottrup-Jensen, L., T. M. Stepanik, T. Kristensen, D. M. Wierzbicki, C. M. Jones, P. B. Lonblad, S. Magnusson, & T. E. Petersen, 1984, Primary structure of human alpha 2-macroglobulin. V. The complete structure: *J Biol Chem*, v. 259, p. 8318-27.
- Sridharan, S., A. Nicholls, & B. Honig, 1992, A new vertex algorithm to calculate solvent accessible surface areas.: *Biophys J*, v. 61, p. A174.

Taylor, W. R., 1997, Residual colours: a proposal for aminochromography: *Protein Eng*, v. 10, p. 743-6.

Tsu, C. A., J. J. Perona, R. J. Fletterick, & C. S. Craik, 1997, Structural basis for the broad substrate specificity of fiddler crab collagenolytic serine protease 1: *Biochemistry*, v. 36, p. 5393-401.

Yang, A. S., & B. Honig, 1999, Sequence to structure alignment in comparative modeling using PrISM: *Proteins*, v. Suppl 3, p. 66-72.

Yang, S. Q., C. I. Wang, S. A. Gillmor, R. J. Fletterick, & C. S. Craik, 1998, Ecotin: a serine protease inhibitor with two distinct and interacting binding sites: *J Mol Biol*, v. 279, p. 945-57.

## Anexo A

Alinhamento de seqüência baseado no alinhamento estrutural, saída do PrISM.  
Os IFRs estão marcados com a cor cinza.

1GCT	A	1	245	-----	-----	-----	-----	-----
1K2I	1	1	245	-----	-----	-----	-----	-----
1CHO	E	1	245	-----	-----	-----	-----	-----
1H4W	A	16	246	-----	-----	-----	-----	-----
1HJ9	A	16	245	-----	-----	-----	-----	-----
2A31	A	16	245	-----	-----	-----	-----	-----
1TRN	A	16	246	-----	-----	-----	-----	-----
1FY8	E	14	245	-----	-----	-----	-----	-----
1A0J	A	16	245	-----	-----	-----	-----	-----
1HJ8	A	16	245	-----	-----	-----	-----	-----
2F91	A	16	243	-----	-----	-----	-----	-----
1NPM	A	16	243	-----	-----	-----	-----	-----
1EUF	A	16	243	-----	-----	-----	-----	-----
3RP2	A	16	243	-----	-----	-----	-----	-----
1IAU	A	16	244	-----	-----	-----	-----	-----
1GVK	B	16	245	-----	-----	-----	-----	-----
1ELT	A	16	242	-----	-----	-----	-----	-----
1BRU	P	16	245	-----	-----	-----	-----	-----
1EKB	B	16	243	-----	-----	-----	-----	-----
1EAX	A	16	244	-----	-----	-----	-----	-----
1H8D	H	16	246	-----	-----	-----	-----	-----
1VR1	H	16	243	-----	-----	-----	-----	-----
1C5L	H	16	247	-----	-----	-----	-----	-----
2PUX	B	16	246	-----	-----	-----	-----	-----
1ETR	H	16	247	-----	-----	-----	-----	-----
1HNE	E	16	243	-----	-----	-----	-----	-----
1FUJ	A	16	243	-----	-----	-----	-----	-----
1A7S	A	1	225	-----	-----	-----	-----	-----
1AZZ	A	16	246	-----	-----	-----	-----	-----
1LO6	A	16	243	-----	-----	-----	-----	-----
2PKA	A	16	246	-----	-----	-----	-----	-----
1T32	A	16	244	-----	-----	-----	-----	-----
1Q3X	A362	686		TIVDCGPPDD	LPSGRVEYIT	GPGVTTYKAV	IQYSCEETFY	TMKVNDGKYV
1EQ9	A	16	244	-----	-----	-----	-----	-----
1FIW	A	16	254	-----	-----	-----	-----	-----
1BIO	A	16	243	-----	-----	-----	-----	-----
1OP0	A	16	245	-----	-----	-----	-----	-----
1BQY	A	16	245	-----	-----	-----	-----	-----
1PQ7	A	16	239	-----	-----	-----	-----	-----
1OS8	A	16	245	-----	-----	-----	-----	-----
2FPZ	A	16	244	-----	-----	-----	-----	-----
1LTO	A	16	244	-----	-----	-----	-----	-----
1C5Y	B	16	245	-----	-----	-----	-----	-----
1TON	A	16	246	-----	-----	-----	-----	-----
1M9U	A	16	242	-----	-----	-----	-----	-----
2BZ6	H	16	257	-----	-----	-----	-----	-----
1RTF	B	16	244	-----	-----	-----	-----	-----
2HLC	A	16	246	-----	-----	-----	-----	-----
1NN6	A	18	241	-----	-----	-----	-----	-----
1GVZ	A	16	246	-----	-----	-----	-----	-----
1DLE	A	1K250		-----	-----	-----	-----	-----
1PYT	D701	944		-----	-----	-----	-----	-----

208L	A	1	216	-----	-----	-----	-----	-----
1P3C	A	1	215	-----	-----	-----	-----	-----
1AGJ	A	1	242	-----	-----	-----	-----	-----
1QTF	A	1	246	-----	-----	-----	-----	-----
1FON	A	9	240	-----	-----	-----	-----	-----
2SGA	A	16	242	-----	-----	-----	-----	-----
2SFA	A	1	191	-----	-----	-----	-----	-----
1SGP	E	16	242	-----	-----	-----	-----	-----
1HPG	A	16	242	-----	-----	-----	-----	-----
2H5C	A	15A	245	-----	-----	-----	-----	-----
1Y8T	A	6	314	-----	-----	-----	-----	-----
1SOT	A	43	353	-----	-----	-----	-----	-----
1DY9	A	1	175	-----	-----	-----	-----	-----
1ARB	A	1	263	-----	-----	-----	-----	-----
1MBM	A	7	204	-----	-----	-----	-----	-----
2FOM	B	18	167	-----	-----	-----	-----	-----
2FP7	B	19	170	-----	-----	-----	-----	-----
1EP5	A	120	275	-----	-----	-----	-----	-----
1GCT	A	1	245	-----	-----	C GVPAIQPVLI	V--N	-----GEE
1K2I	I	1	245	-----	-----	C GVPAIQPVLI	V--N	-----GEE
1CHO	E	1	245	-----	-----	C GVPAIQPVLI	V--N	-----GEE
1H4W	A	16	246	-----	-----	-----	I V--G	-----GYT
1HJ9	A	16	245	-----	-----	-----	I V--G	-----GYT
2A31	A	16	245	-----	-----	-----	I V--G	-----GYT
1TRN	A	16	246	-----	-----	-----	I V--G	-----GYN
1FY8	E	14	245	-----	-----	-----	D K--G	-----GYT
1A0J	A	16	245	-----	-----	-----	I V--G	-----GYE
1HJ8	A	16	245	-----	-----	-----	I V--G	-----GYE
2F91	A	16	243	-----	-----	-----	I V--G	-----GTD
1NPM	A	16	243	-----	-----	-----	I L--E	-----GRE
1EUF	A	16	243	-----	-----	-----	I I--G	-----GHE
3RP2	A	16	243	-----	-----	-----	I I--G	-----GVE
1IAU	A	16	244	-----	-----	-----	I I--G	-----GHE
1GVK	B	16	245	-----	-----	-----	V V--G	-----GTE
1ELT	A	16	242	-----	-----	-----	V V--G	-----GRV
1BRU	P	16	245	-----	-----	-----	V V--G	-----GED
1EKB	B	16	243	-----	-----	-----	I V--G	-----GSD
1EAX	A	16	244	-----	-----	-----	V V--G	-----GTD
1H8D	H	16	246	-----	-----	-----	I V--E	-----GSD
1VR1	H	16	243	-----	-----	-----	I V--E	-----GSD
1C5L	H	16	247	-----	-----	-----	I V--E	-----GSD
2PUX	B	16	246	-----	-----	-----	I V--E	-----GWD
1ETR	H	16	247	-----	-----	-----	I V--E	-----GQD
1HNE	E	16	243	-----	-----	-----	I V--G	-----GRR
1FUJ	A	16	243	-----	-----	-----	I V--G	-----GHE
1A7S	A	1	225	-----	-----	-----	I V--G	-----GRK
1AZZ	A	16	246	-----	-----	-----	I V--G	-----GVE
1LO6	A	16	243	-----	-----	-----	L V--H	-----GGP
2PKA	A	16	246	-----	-----	-----	I I--G	-----GRE
1T32	A	16	244	-----	-----	-----	I I--G	-----GRE
1Q3X	A	362	686	CEADGFWTSS	KGEKSLPVC	PVCGLSARTI	Y--G	-----GQK
1EQ9	A	16	244	-----	-----	-----	I V--G	-----GKD
1FIW	A	16	254	-----	-----	-----	I I--G	-----GQD
1BIO	A	16	243	-----	-----	-----	I L--G	-----GRE
1OP0	A	16	245	-----	-----	-----	V I--G	-----GNE
1BQY	A	16	245	-----	-----	-----	V F--G	-----GDE
1PQ7	A	16	239	-----	-----	-----	I V--G	-----GTS
1OS8	A	16	245	-----	-----	-----	V V--G	-----GTR
2FPZ	A	16	244	-----	-----	-----	I V--G	-----GQE
1LTO	A	16	244	-----	-----	-----	I V--G	-----GQE

1C5Y	B	16	245	-----	-----	-----	I	I--G	-----	-----	GEF
1TON	A	16	246	-----	-----	-----	I	V--G	-----	-----	GYP
1M9U	A	16	242	-----	-----	-----	V	I--G	-----	-----	GTN
2BZ6	H	16	257	-----	-----	-----	I	V--G	-----	-----	GKV
1RTF	B	16	244	-----	-----	-----	I	K--G	-----	-----	GLF
2HLC	A	16	246	-----	-----	-----	I	I--N	-----	-----	GYE
1NN6	A	18	241	-----	-----	-----	-----	-----	-----	-----	GTE
1GVZ	A	16	246	-----	-----	-----	I	I--G	-----	-----	GWE
1DLE	A	1K250	-----	-----	AD	PDESQSLSLC	G	M	-----	-----	VKG
1PYT	D701	944	-----	-----	CGAP	IFQPNLSARV	V	G	-----	-----	GED
2O8L	A	1	216	-----	-----	-----	V	ILPNN	-----	-----	DRHQ
1P3C	A	1	215	-----	-----	-----	VV	IGDDG	-----	-----	RTK
1AGJ	A	1	242	-----	-----	EV	SAEEIKKHEE	KWNKYYGVNA	FNLPKELFSK	-----	-----
1QTF	A	1	246	-----	-----	-----	KEYSAEEIRK	LKQKFE--VP	-PTDKELYTH	-----	-----
1FON	A	9	240	-----	-----	-----	-----	-----	-----	-----	-----
2SGA	A	16	242	-----	-----	-----	-----	-----	-----	-----	-----
2SFA	A	1	191	-----	-----	-----	-----	-----	-----	-----	-----
1SGP	E	16	242	-----	-----	-----	-----	-----	-----	-----	-----
1HPG	A	16	242	-----	-----	-----	-----	-----	-----	-----	-----
2H5C	A	15A245	-----	-----	-----	-----	-----	-----	-----	-----	-----
1Y8T	A	6	314	-----	-----	-----	-----	G--S	-----	-----	VEQ
1SOT	A	43	353	-----	-----	-----	TP	A--S	-----	-----	YNL
1DY9	A	1	175	-----	-----	AP	ITAYSQQTRG	LLGCI	-----	-----	ITSL
1ARB	A	1	263	-----	-----	-----	GVS	GS--C	-----	-----	NID
1MBM	A	7	204	-----	-----	-----	KA	-----	-----	-----	-----
2FOM	B	18	167	-----	-----	-----	-----	-----	-----	-----	-----
2FP7	B	19	170	-----	-----	-----	-----	-----	-----	-----	-----
1EP5	A120	275	-----	-----	-----	-----	-----	-----	-----	-----	-----

1GCT	A	1	245	AVP	----	G	S--WPWQ-VS	LQ-DK	----	T	G	----	F
1K2I	1	1	245	AVP	----	G	S--WPWQ-VS	LQ-DK	----	T	G	----	F
1CHO	E	1	245	AVP	----	G	S--WPWQ-VS	LQ-DK	----	T	G	----	F
1H4W	A	16	246	CEE	----	N	S--LPYQ-VS	LN-S	----	G	----	----	S
1HJ9	A	16	245	CGA	----	N	T--VPYQ-VS	LN-S	----	G	----	----	Y
2A31	A	16	245	CAA	----	N	S--IPYQ-VS	LN-S	----	G	----	----	S
1TRN	A	16	246	CEE	----	N	S--VPYQ-VS	LN-S	----	G	----	----	Y
1FY8	E	14	245	CQE	----	N	S--VPYQ-VS	LN-S	----	G	----	----	Y
1A0J	A	16	245	CRK	----	N	S--ASYQ-AS	LQ-S	----	G	----	----	Y
1HJ8	A	16	245	CKA	----	Y	S--QPHQ-VS	LN-S	----	G	----	----	Y
2F91	A	16	243	ATL	----	G	E--FPYQ-LS	FQ-E	----	T	F	----	IGFS-F
1NPM	A	16	243	CIP	----	H	S--QPWQ-AA	LF-Q	----	G	E	----	R
1EUF	A	16	243	AKP	----	H	S--RPYM-AF	LL-F	----	K	T	----	SGK-S
3RP2	A	16	243	SIP	----	H	S--RPYM-AH	LD-I	----	V	T	----	EKGL-R
1IAU	A	16	244	AKP	----	H	S--RPYM-AY	LM-I	----	W	D	----	QKS-L
1GVK	B	16	245	AQR	----	N	S--WPSQ-IS	LQ-YRSGSS	----	W	----	----	A
1ELT	A	16	242	AQP	----	N	S--WPWQ-IS	LQ-YKSGSS	----	Y	----	----	Y
1BRU	P	16	245	ARP	----	N	S--WPWQ-VS	LQ-YDSSGQ	----	W	----	----	R
1EKB	B	16	243	SRE	----	G	A--WPWV-VA	LY-F	----	D	D	----	Q
1EAX	A	16	244	ADE	----	G	E--WPWQ-VS	LH-A	----	L	G	----	Q-G
1H8D	H	16	246	AEI	----	G	M--SPWQ-VM	LF-RKSP	----	Q	----	----	E
1VR1	H	16	243	AEI	----	G	M--SPWQ-VM	LF-AKH	----	E	----	----	R
1C5L	H	16	247	AEI	----	G	M--SPWQ-VM	LF-RKSP	----	Q	----	----	E
2PUX	B	16	246	AEK	----	G	I--APWQ-VM	LF-RKSP	----	Q	----	----	E
1ETR	H	16	247	AEV	----	G	L--SPWQ-VM	LF-RKSP	----	Q	----	----	E
1HNE	E	16	243	ARP	----	H	A--WPFM-VS	LQ-L	----	R	G	----	G
1FUJ	A	16	243	AQP	----	H	S--RPYM-AS	LQ-MRG	----	N	P	----	G-S
1A7S	A	1	225	ARP	----	R	Q--FPFL-AS	IQ-N	----	Q	G	----	R
1AZZ	A	16	246	AVP	----	N	S--WPHQ-AA	LF-I	----	D	----	----	D-M
1LO6	A	16	243	CDK	----	T	S--HPYQ-AA	LY-T	----	S	G	----	H
2PKA	A	16	246	CEK	----	N	S--HPWQ-VA	IY-YY	----	S	----	----	S
1T32	A	16	244	SRP	----	H	S--RPYM-AY	LQ-IQSP	----	A	----	----	GQ

1Q3X	A362	686	AKP----	G-	D--	FPWQ-VL	IL-	-----	---	G	-----	G-
1EQ9	A	16	244	APV----	G-	K--	YPYQ-VS	LR-L	-----	S-G	-----	S
1FIW	A	16	254	AAH----	G-	A--	WPWM-VS	LQ-I	-----	F-TYH	NNRR	Y
1BIO	A	16	243	AEA----	H-	A--	RPYM-AS	VQ-L	-----	N-G	-----	A
1OP0	A	16	245	CDI----	N-	E--	HRFL-VA	FF-N	-----	T-T	-----	G
1BQY	A	16	245	CNI----	N-	E--	HRSL-VV	LF-N	-----	S-N	-----	G
1PQ7	A	16	239	ASA----	G-	D--	FPFI-VS	IS-R	-----	N	-----	G-G
1OS8	A	16	245	AAQ----	G-	E--	FPFM-VR	LS-M	-----	-----	-----	-----
2FPZ	A	16	244	APR----	S	K--	WPWQ-VS	LR-VHG-PY	-----	W	-----	M
1LTO	A	16	244	APR----	S	K--	WPWQ-VS	LR-VRD-RY	-----	W	-----	M
1C5Y	B	16	245	TTI----	E-	N--	QPWF-AA	IY-R	-----	R-HR	GGSV	T
1TON	A	16	246	CEK----	N-	S--	QPWQ-VA	VI-N	-----	-----	-----	E
1M9U	A	16	242	ASP----	G-	E--	FPWQ-LS	QQ-RQSGS	-----	W	-----	S
2BZ6	H	16	257	CPK----	G-	E--	CPWQ-VL	LL-VN	-----	G	-----	A
1RTF	B	16	244	ADI----	A-	S--	HPWQ-AA	IF-A	-----	KNGE	-----	R
2HLC	A	16	246	AYT----	G-	L--	FPYQ-AG	LD-I	-----	TLQDQ	-----	R-R
1NN6	A	18	241	CKP----	H-	S--	RPYM-AY	LE-IVTSNG	-----	P	-----	S
1GVZ	A	16	246	CEK----	H-	S--	KPWQ-VA	VY-H	-----	QG	-----	H
1DLE	A	1K	250	TDY----	H-	K--	QPWQ-AK	IS-V	-----	I-R	PSKGH	H
1PYT	D701	944	AIP----	H-	S--	WPWQ-IS	LQ-YLRDNT	-----	W	-----	R	
2O8L	A	1	216	ITD----	TT	NGHYAPV-TY	IQ-V	-----	EA-PT	G	T	
1P3C	A	1	215	VAN----	TR	VAPYNSI-AY	IT-FG	-----	-----	-----	G	
1AGJ	A	1	242	VDEKD--	RQ	KYPYNTI-GN	VF-V	-----	K	G	Q	
1QTF	A	1	246	ITD----	NA	RSPYNSV-GT	VF-V	-----	K	G	S	
1FON	A	9	240	-----	---	S--	WSWQ-VS	LQ-YEKDGA	-----	F	H	
2SGA	A	16	242	-----	---	IAGGEA	IT-T	-----	G	-----	G	
2SFA	A	1	191	-----	---	IAGGEA	IY-AA	-----	G	-----	G	
1SGP	E	16	242	-----	---	ISGGDA	IY-SS	-----	-----	-----	T	
1HPG	A	16	242	-----	---	VLGGGA	IY-GG	-----	-----	-----	G	
2H5C	A	15A	245	-----	---	ANIVGGIE	YS-IN	-----	N	-----	A	
1Y8T	A	6	314	VAA----	K-	V--	VPSV-VL	ET-D	-----	L	-----	-----
1SOT	A	43	353	AVR----	R-	A--	APAV-VN	VYNRGLNTNS	HNQLE	I	R	
1DY9	A	1	175	TGR----	DK	NQVDGEV-QV	LS-T	-----	-----	A	T	
1ARB	A	1	263	VVCEGDGR-	---	RDIIRAV-GA	YS-K	-----	S	-----	G-T	
1MBM	A	7	204	-----	---	RGNV-GF	VA-G	-----	-----	-----	S	
2FOM	B	18	167	-----	---	LEDGA-YR	IK-QKGILG	-----	-----	Y	S	
2FP7	B	19	170	-----	---	TTGV-YR	IM-T	-----	-----	S	Y	
1EP5	A120	275	-MK----	L-	E--	SDKT-FP	IM-L	-----	E-G	-----	K	
1GCT	A	1	245	HFCGGSLI-N	---	E--	NWVVTAAH	C-G	-----	V	TT	
1K2I	1	1	245	HFCGGSLI-N	---	E--	NWVVTAAH	C-G	-----	V	TT	
1CHO	E	1	245	HFCGGSLI-N	---	E--	NWVVTAAH	C-G	-----	V	TT	
1H4W	A	16	246	HFCGGSLI-S	---	E--	QWVVSAAH	C-Y	-----	K	T	
1HJ9	A	16	245	HFCGGSLI-N	---	S--	QWVVSAAH	C-Y	-----	K	S	
2A31	A	16	245	HFCGGSLI-N	---	S--	QWVVSAAH	C-Y	-----	K	S	
1TRN	A	16	246	HFCGGSLI-N	---	E--	QWVVSAGH	C-Y	-----	K	S	
1FY8	E	14	245	HFCGGSLI-N	---	D--	QWVVSAAH	C-Y	-----	K	S	
1A0J	A	16	245	HFCGGSLI-S	---	S--	TWVVSAAH	C-Y	-----	K	S	
1HJ8	A	16	245	HFCGGSLV-N	---	E--	NWVVSAAH	C-Y	-----	K	S	
2F91	A	16	243	HFCGASIY-N	---	E--	NYAITAGH	C-VYG	-----	DDYEN	P	
1NPM	A	16	243	LICGGVLV-G	---	D--	RWVLTAAH	C-K	-----	K	Q	
1EUF	A	16	243	HICGGFLV-R	---	E--	DFVLTAAH	C-L	-----	G	S	
3RP2	A	16	243	VICGGFLI-S	---	R--	QFVLTAAH	C-K	-----	G	R	
1IAU	A	16	244	KRCGGFLI-R	---	D--	DFVLTAAH	C-W	-----	G	S	
1GVK	B	16	245	HTCGGTLI-R	---	Q--	NWVMTAAH	C-V	-----	D	REL	
1ELT	A	16	242	HTCGGSLI-R	---	Q--	GWVMTAAH	C-V	-----	D	SAR	
1BRU	P	16	245	HTCGGTLV-D	---	Q--	SWVLTAAH	C-I	-----	S	SSR	
1EKB	B	16	243	QVCGASLV-S	---	R--	DWLVSAAH	C-VYG	-----	RNM	EPS	
1EAX	A	16	244	HICGASLI-S	---	P--	NWLVSAAH	C-Y	DDRGFR	YS	DPT	
1H8D	H	16	246	LLCGASLI-S	---	D--	RWVLTAAH	C-LLYPPWDK	NF	-----	TEN	
1VR1	H	16	243	FLCGASLI-S	---	D--	RWVLTAAH	C-LLYPPWDK	NF	-----	TEN	



1C5L	H	16	247	LLCGASLI-S	-----D---	--RWVLTAAH	C-LLYPPWDK	NF-----TEN
2PUX	B	16	246	LLCGASLI-S	-----D---	--RWVLTAAH	C-ILYPPWDK	NF-----TEN
1ETR	H	16	247	LLCGASLI-S	-----D---	--RWVLTAAH	C-LLYPPWDK	NF-----TVD
1HNE	E	16	243	HFCGATLI-A	-----P---	--NFVMSAAH	C-VAN-----	-V-----NVR
1FUJ	A	16	243	HFCGGTLI-H	-----P---	--SFVLTAAH	C-LRD-----	-I-----PQR
1A7S	A	1	225	HFCGGALI-H	-----A---	--RFVMTAAS	C-F-----	-P-----G--
1AZZ	A	16	246	YFCGGSLI-S	-----P---	--EWILTAAH	C-MDG-----	-A-----G--
1LO6	A	16	243	LLCGGVLI-H	-----P---	--LWVLTAAH	C-K-----	-K-----P--
2PKA	A	16	246	FQCGGVLV-N	-----P---	--KWVLTAAH	C-K-----	-N-----D--
1T32	A	16	244	SRCGGFLV-R	-----E---	--DFVLTAAH	C-W-----	-G-----S--
1Q3X	A362	686	TTAAGALL-Y	-----D---	--NWVLTAAH	A-VYE-----	-Q-----KHD	
1EQ9	A	16	244	HRCGASIL-D	-----N---	--NNVLTAAH	C-VDG-----	-L-----SN-
1FIW	A	16	254	HVCGGSLI-N	-----S---	--QWLLTAAH	C-FRI-----	-KK-----KVT
1BIO	A	16	243	HLCGGVLV-A	-----E---	--QWVLSAAH	C-LED-----	-A-----ADG
1OP0	A	16	245	FFCGGTLI-N	-----P---	--EWWVTAAH	C-D-----	-S-----T--
1BQY	A	16	245	FLCGGTLI-N	-----Q---	--DWVVTAAH	C-D-----	-S-----N--
1PQ7	A	16	239	PWCGGSLI-N	-----A---	--NTVLTAAH	C-VSG-----	-YA-----Q--
1OS8	A	16	245	-GCGGALY-A	-----Q---	--DIVLTAAH	C-V-----	-SGSGNNT--
2FPZ	A	16	244	HFCGGSLI-H	-----P---	--QWVLTAAH	C-V-----	-GPD----VKD
1LTO	A	16	244	HFCGGSLI-H	-----P---	--QWVLTAAH	C-L-----	-GPD----VKD
1C5Y	B	16	245	YVCGGSLM-S	-----P---	--CWVISATH	C-FIDY-----	-PK-----K--
1TON	A	16	246	YLCGGVLI-D	-----P---	--SWVITAAH	C-Y-----	-S-----N--
1M9U	A	16	242	HSCGASLL-S	-----S---	--TSALSASH	C-VDG-----	-V-----LPN
2BZ6	H	16	257	QLCGGTLI-N	-----T---	--IWVVSAAH	C-FDK-----	-IKN----WR-
1RTF	B	16	244	FLCGGILI-S	-----S---	--CWILSAAH	C-FQE-----	-R-----FP-
2HLC	A	16	246	WVCGGSLI-D	-----N---	--KWILTAAH	C-VLD-----	-A-----V--
1NN6	A	18	241	KFCGGFLI-R	-----R---	--NFVLTAAH	C-A-----	-G-----R--
1GVZ	A	16	246	FQCGGVLV-H	-----P---	--QWVLTAAH	C-M-----	-S-----D--
1DLE	A	1K250	ESCMGAVV-S	-----E---	--YFVLTAAH	C-F-----	-T-----VDD	
1PYT	D701	944	HTCGGTLI-T	-----P---	--NHVLTAAH	C-I-----	-S-----NTL	
2O8L	A	1	216	FIASGVVV-G	-----K---	--DTLLTNKH	V-VDA----TH	-G-----DPH
1P3C	A	1	215	SSCTGTLI-A	-----P---	--NKILTNGH	C-VYNTASRS	YS-----A--
1AGJ	A	1	242	TSATGVLI-G	-----K---	--NTVLTNRH	I-AKF----AN	-G-----DPS
1QTF	A	1	246	TLATGVLI-G	-----K---	--NTIVTNYH	V-ARE----AA	-K-----NPS
1FON	A	9	240	HTCGGSLI-A	-----P---	--DWVVTAGH	C-I-----	-STS----R--
2SGA	A	16	242	SRCSLGFN-V	-----SVN-	GVAHALTAGH	C-TNI----S-	-----AS
2SFA	A	1	191	GRCSLGFN-V	-----RSSS	GATYALTAGH	C-TEI----A-	-----ST
1SGP	E	16	242	GRCSLGFN-V	-----RS-G	STYYFLTAGH	C-TDG----A-	-----TT
1HPG	A	16	242	SRCSAAFN-V	-----TK-G	GARYFVTAGH	C-TNI----S-	-----AN
2H5C	A	15A245	SLCSVGFS-V	-----TR-G	ATKGFVTAGH	C-GTV----N-	-----AT	
1Y8T	A	6	314	EEGSGIIL-S	A-----E-	--GLILTNNH	V-IA-----	-A-----AA
1SOT	A	43	353	TLGSGVI-D	-----QR-	--GYIITNKH	VIN-----	-D-----AD-
1DY9	A	1	175	QSFLATCV-N	-----G---	--VCWTVYH	G-AG-----	-S-----KT
1ARB	A	1	263	LACTGSLV-N	NTANDRK---	--MYFLTAAH	C-G-----	-MGTAS-TAA
1MBM	A	7	204	SYGTGSVWTR	NNE---V---	--VVLTASH	V-VGRA----	-N-----MA
2FOM	B	18	167	QIGAGVYK-E	-----G---	--TFHTMWH	V-TR-----	-G-----AV
2FP7	B	19	170	QAGAGVMV-E	-----G---	--VFHTLWH	T-TK-----	-G-----AA
1EP5	A	120	275	INGYACVV-G	-----G---	--KLFRRPMH	V-----	-E-----
1GCT	A	1	245	---	S--DV-V	VAGEFD---	---	Q--G--S-SSE-K
1K2I	1	1	245	---	S--DV-V	VAGEFD---	---	Q--G--S-SSE-K
1CHO	E	1	245	---	S--DV-V	VAGEFD---	---	Q--G--S-SSE-K
1H4W	A	16	246	---	R--IQ-V	RLGEHN---	---	I--K--V-L-E-G
1HJ9	A	16	245	---	G--IQ-V	RLGEDN---	---	I--N--V-V-E-G
2A31	A	16	245	---	R--IQ-V	RLGEHN---	---	I--D--V-L-E-G
1TRN	A	16	246	---	R--IQ-V	RLGEHN---	---	I--E--V-L-E-G
1FY8	E	14	245	---	R--IQ-V	RLGEHN---	---	I--N--V-L-E-G
1A0J	A	16	245	---	R--IQ-V	RLGEHN---	---	I--A--V-N-E-G
1HJ8	A	16	245	---	R--VE-V	RLGEHN---	---	I--K--V-T-E-G
2F91	A	16	243	---	SG-LQ-I	VAGELD---	---	M--S--V-N-E-G
1NPM	A	16	243	---	K--YS-V	RLGDHS---	---	L-----Q-S---

1EUF	A	16	243	---	S	--	IN	-V	TLGAHN	---	---	I	---	M	---	E	-R	-E	-R							
3RP2	A	16	243	---	E	--	IT	-V	ILGAHD	---	---	V	---	R	---	K	-R	-E	-S							
1IAU	A	16	244	---	S	--	IN	-V	TLGAHN	---	---	I	---	K	---	E	-Q	-E	-P							
1GVK	B	16	245	---	T	--	FR	-V	VVGEHN	---	---	L	---	N	---	Q	-N	-N	-G							
1ELT	A	16	242	---	T	--	WR	-V	VLGEHN	---	---	L	---	N	---	T	-N	-E	-G							
1BRU	P	16	245	---	T	--	YR	-V	VLGRHS	---	---	L	---	S	---	T	-N	-E	-P							
1EKB	B	16	243	---	K	--	WK	-A	VLGLHM	---	---	A	---	S	---	NLT	-S	-P	---							
1EAX	A	16	244	---	Q	--	WT	-A	FLGLHD	---	---	Q	---	S	---	QRS	-A	-P	---							
1H8D	H	16	246	---	D	--	LL	-V	RIGKHS	---	---	R	---	T	---	R	-Y	-ERN	---							
1VR1	H	16	243	---	D	--	LL	-V	RIGKHS	---	---	R	---	T	---	R	-Y	-ERN	---							
1C5L	H	16	247	---	D	--	LL	-V	RIGKHS	---	---	R	---	T	---	R	-Y	-ERN	---							
2PUX	B	16	246	---	D	--	LL	-V	RIGKHS	---	---	R	---	T	---	R	-Y	-ERN	---							
1ETR	H	16	247	---	D	--	LL	-V	RIGKHS	---	---	R	---	T	---	R	-Y	-ERK	---							
1HNE	E	16	243	---	A	--	VR	-V	VLGAHN	---	---	L	---	---	---	S	-RREP	---	---							
1FUJ	A	16	243	---	L	--	VN	-V	VLGAHN	---	---	V	---	---	---	R	-TQEP	---	---							
1A7S	A	1	225	---	V	--	ST	-V	VLGAYD	---	---	L	---	---	---	R	-RRERQ	---	---							
1AZZ	A	16	246	---	F	--	VD	-V	VLGAHN	---	---	I	---	R	---	E	-D	-E	-A							
1LO6	A	16	243	---	N	--	LQ	-V	FLGKHN	---	---	L	---	R	---	Q	-R	-E	-S							
2PKA	A	16	246	---	N	--	YE	-V	WLGRHN	---	---	L	---	F	---	E	-N	-E	-N							
1T32	A	16	244	---	N	--	IN	-V	TLGAHN	---	---	I	---	Q	---	R	-R	-E	-N							
1Q3X	A362	686		AS	-A	--	LD	-I	RMGTLK	---	---	R	---	L	---	S	-P	-H	---							
1EQ9	A	16	244	---	LNRLK	-V	HVGTNY	---	---	---	L	---	S	---	E	-S	---	---	---							
1FIW	A	16	254	---	D	--	WR	-L	IFGAKE	---	---	V	---	EW	GTNKP	V	-K	-P	---							
1BIO	A	16	243	---	K	--	VQ	-V	LLGAHS	---	---	L	---	S	---	Q	-P	-E	-P							
1OP0	A	16	245	---	D	--	FQ	-M	QLGVHS	---	---	K	---	---	---	K	VLN	-E	---							
1BQY	A	16	245	---	N	--	FQ	-L	LFGVHS	---	---	K	---	---	---	K	ILN	-E	---							
1PQ7	A	16	239	---	SG	-FQ	-I	RAGSLS	---	---	R	---	T	---	---	---	S	-G	---							
1OS8	A	16	245	---	S	--	IT	-A	TGGVVD	---	---	L	---	---	---	Q	-S	-S	-S							
2FPZ	A	16	244	LA	-A	--	LR	-V	QLREQH	---	---	L	---	Y	---	Y	-Q	---	---							
1LTO	A	16	244	LA	-T	--	LR	-V	QLREQH	---	---	L	---	Y	---	Y	-Q	---	---							
1C5Y	B	16	245	---	ED	-YI	-V	YLGRSR	---	---	L	---	N	---	S	-N	-T	-Q	---							
1TON	A	16	246	---	N	--	YQ	-V	LLGRNN	---	---	L	---	F	---	K	-D	-E	-P							
1M9U	A	16	242	---	N	--	IR	-V	IAGLWQ	---	---	Q	---	S	---	D	-T	-S	-G							
2BZ6	H	16	257	---	N	--	LI	-A	VLGEHD	---	---	L	---	S	---	E	-H	-D	-G							
1RTF	B	16	244	---	PHHLT	-V	ILGRTY	---	---	---	RV	---	V	---	P	-GEE	---	---	---							
2HLC	A	16	246	---	S	--	VV	-V	YLGSAV	---	---	Q	---	Y	---	E	---	---	---							
1NN6	A	18	241	---	S	--	IT	-V	TLGAHN	---	---	I	---	T	---	E	-E	-E	-D							
1GVZ	A	16	246	---	D	--	YQ	-I	WLGRHN	---	---	L	---	S	---	K	-D	-E	-D							
1DLE	A	1K250		KEHS	--	IK	-V	SVGG	---	---	---	---	---	---	---	E	-K	---	---							
1PYT	D701	944		---	T	--	YR	-V	ALGKNN	---	---	L	---	E	---	V	-EDE	-A	---							
2O8L	A	1	216	---	A	--	LK	-A	FPSAIN	---	---	---	---	Q	---	DN	-Y	---	---							
1P3C	A	1	215	---	K	--	GS	-V	YPMND	---	---	---	---	S	---	T	-A	---	---							
1AGJ	A	1	242	---	K	--	VS	-F	RPSINT	---	---	---	---	DDN	GN	---	TE	-T	---							
1QTF	A	1	246	---	N	--	II	-F	TPAQNR	---	---	---	---	DAEKNE	---	FP	-T	---	---							
1FON	A	9	240	---	T	--	YQ	-V	VLGEYD	---	---	R	---	S	---	V	-L	-E	-G							
2SGA	A	16	242	---	---	---	WS	---	---	---	---	---	---	---	---	---	---	---	---							
2SFA	A	1	191	---	---	---	WY	-T	NSGQTS	---	---	---	---	L	---	---	---	---	---							
1SGP	E	16	242	---	---	---	WW	-A	NSARTT	---	---	---	---	V	---	---	---	---	---							
1HPG	A	16	242	---	---	---	WS	-A	SS	GG	---	---	---	V	---	---	---	---	---							
2H5C	A	15A245		---	---	---	AR	-I	G	GA	---	---	---	V	---	---	---	---	---							
1Y8T	A	6	314	---	P	--	KT	-T	VTFS	D	---	---	---	---	---	---	---	G	---							
1SOT	A	43	353	---	Q	--	II	-V	ALQ	---	---	---	---	---	---	---	---	---	---							
1DY9	A	1	175	---	---	---	LA	-G	---	---	---	---	---	---	---	---	---	---	---							
1ARB	A	1	263	---	S	--	IV	-V	YW-NYQNSTC	RAPNTPASGA	NGDG	---	S	---	M	-S	---	---	---							
1MBM	A	7	204	---	---	---	TLKI	GD	---	---	---	---	---	---	---	---	---	---	---							
2FOM	B	18	167	---	---	---	LM	-H	---	---	---	---	---	---	---	---	---	---	---							
2FP7	B	19	170	---	---	---	LM	-S	---	---	---	---	---	---	---	---	---	---	---							
1EP5	A120	275		---	---	---	GK	-I	DN	---	---	---	---	---	---	---	---	---	---							
1GCT	A	1	245	-	IQK	LK	-IA-	-	K	-VF	-K	-NS	KY	---	N	---	S	---	---	---	---	L	---	TI	---	NNDI
1K2I	1	1	245	-	IQK	LK	-IA-	-	K	-VF	-K	-NS	KY	---	N	---	S	---	---	---	---	L	---	TI	---	NNDI

1CHO	E	1	245	-IQK <b>LK</b> -IA-	-K-VF-K-NS	KY	---	N	---	S	-----	L	---	TI	---	NNDI	
1H4W	A	16	246	NEQFIN-AV-	-K-II-R-HP	KY	---	N	---	R	-----	D	---	TL	---	DNDI	
1HJ9	A	16	245	NEQFIS-AS-	-K-SI-V-HP	SY	---	N	---	S	-----	N	---	TL	---	NNDI	
2A31	A	16	245	NEQFIN-AA-	-K-II-T-HP	NF	---	N	---	G	-----	N	---	TL	---	DNDI	
1TRN	A	16	246	NEQFIN-AA-	-K-II-R-HP	QY	---	D	---	R	-----	K	---	TL	---	NNDI	
1FY8	E	14	245	NEQFVN-AA-	-K-II-K-HP	NF	---	D	---	R	-----	K	---	TL	---	NNDI	
1A0J	A	16	245	TEQFID-SV-	-K-VI-M-HP	SY	---	N	---	S	-----	R	---	NL	---	DNDI	
1HJ8	A	16	245	SEQFIS-SS-	-R-VI-R-HP	NY	---	S	---	S	-----	Y	---	NI	---	DNDI	
2F91	A	16	243	SEQIIT-VS-	-K-II-L-HE	NF	---	D	---	Y	-----	N	---	LL	---	DNDI	
1NPM	A	16	243	PEQEIQ-VA-	-Q-SI-Q-HP	CY	---	N	---	S	-----	E	---	DI	---	SHDI	
1EUF	A	16	243	TQQVIP-VR-	-R-PI-P-HP	DY	---	N	---	D	-----	E	---	TL	---	ANDI	
3RP2	A	16	243	TQQKIK-VE-	-K-QI-I-HE	SY	---	N	---	S	-----	V	---	PN	---	LHDI	
1IAU	A	16	244	TQQFIP-VK-	-R-PI-P-HP	AY	---	N	---	P	-----	K	---	NE	---	SNDI	
1GVK	B	16	245	TEQYVG-VQ-	-K-IV-V-HP	YW	---	N	---	T	-----	D	---	DVAAGYDI			
1ELT	A	16	242	KEQIMT-VN-	-S-VF-I-HS	GW	---	N	---	S	-----	D	---	DVAGGYDI			
1BRU	P	16	245	GSLAVK-VS-	-K-LV-V-HQ	DW	---	N	---	S	-----	N	---	QLSNGNDI			
1EKB	B	16	243	QIETRL-ID-	-Q-IV-I-NP	HY	---	N	---	K	-----	R	---	RK	---	NNDI	
1EAX	A	16	244	GVQERR-LK-	-R-II-S-HP	FF	---	N	---	D	-----	F	---	TF	---	DYDI	
1H8D	H	16	246	IEKISM-LE-	-K-IY-I-HP	RY	---	N	---	W	-----	R	E	NL	---	DRDI	
1VR1	H	16	243	IEKISM-LE-	-K-IY-I-HP	RY	---	N	---	W	-----	R	E	NL	---	DRDI	
1C5L	H	16	247	IEKISM-LE-	-K-IY-I-HP	RY	---	N	---	W	-----	R	E	NL	---	DRDI	
2PUX	B	16	246	VEKISM-LE-	-K-IY-V-HP	RY	---	N	---	W	-----	R	E	NL	---	DRDI	
1ETR	H	16	247	VEKISM-LD-	-K-IY-I-HP	RY	---	N	---	W	-----	K	E	NL	---	DRDI	
1HNE	E	16	243	TRQVFA-VQ-	-R-IF-E-D-	GY	---	D	---	P	-----	V	---	NL	---	LNDI	
1FUJ	A	16	243	TQQHFS-VA-	-Q-VF-L-N-	NY	---	D	---	A	-----	E	---	NK	---	LNDI	
1A7S	A	1	225	SQQTFS-IS-	-S-MS-E-N-	GY	---	D	---	P	-----	Q	---	QN	---	LNDL	
1AZZ	A	16	246	TQVTIQ-ST-	-D-FT-V-HE	NY	---	N	---	S	-----	F	---	VI	---	SNDI	
1L06	A	16	243	SQEQSS-VV-	-R-AV-I-HP	DY	---	D	---	A	-----	A	---	SH	---	DQDI	
2PKA	A	16	246	TAQFFG-VT-	-A-DF-P-HP	GF	---	N	---	L	SADG	---	K	---	DY	---	SHDL
1T32	A	16	244	TQQHIT-AR-	-R-AI-R-HP	QY	---	N	---	Q	-----	R	---	TI	---	QNDI	
1Q3X	A362	686		-YTQAW-SE-	-A-VF-I-HE	GY	---	T	---	H	D	---	A	---	GF	---	DNDI
1EQ9	A	16	244	-GDVYD-VE-	-D-AV-V-NK	NY	---	D	---	D	-----	F	---	LL	---	RNDV	
1FIW	A	16	254	PLQERY-VE-	-K-II-I-HE	KY	---	S	---	A	-----	S	---	SE	---	ANDI	
1BIO	A	16	243	SKRLYD-VL-	-R-AV-P-HP	DS	---	Q	---	P	-----	D	---	TI	---	DHDL	
1OP0	A	16	245	DEQTRN-PK-	-E-KF-I-CP	NK	---	N	---	NNE	-----	V	---	L	---	DKDI	
1BQY	A	16	245	DEQTRD-PK-	-E-KF-F-CP	NR	---	K	---	KDD	-----	E	---	V	---	DKDI	
1PQ7	A	16	239	-GITSS-LS-	-S-VR-V-HP	SY	---	S	---		-----		---	GN	---	NNDL	
1OS8	A	16	245	-AVKVR-ST-	-K-VL-Q-AP	GY	---	N	---	G	-----		---	T	---	GKDW	
2FPZ	A	16	244	-DQLLP-VS-	-R-II-V-HP	QF	---	Y	---	T	-----		---	AQ	---	TGADI	
1LTO	A	16	244	-DQLLP-VS-	-R-II-V-HP	QF	---	Y	---	I	-----		---	IQ	---	TGADI	
1C5Y	B	16	245	GEMKFE-VE-	-N-LI-L-HK	DY	---	S	---	A	-----	D	TLAN	---	HNDI		
1TON	A	16	246	FAQRRL-VR-	-Q-SF-R-HP	DY	---	I	---	P	LPV	---	H	---	DI	---	SNDL
1M9U	A	16	242	-TQTAN-VD-	-S-YT-M-HE	NY	---	G	---	A	GT	---	A	---	SY	---	SNDI
2BZ6	H	16	257	DEQSRR-VA-	-Q-VI-I-PS	TY	---	V	---	P	-----	G	---	TT	---	NHDI	
1RTF	B	16	244	-EQKFE-VE-	-K-YI-V-HK	EF	---	D	---	D	-----	D	---	TY	---	DNDI	
2HLC	A	16	246	GEAVVN-SE-	-R-II-S-HS	MF	---	N	---	P	-----	D	---	TY	---	LNDV	
1NN6	A	18	241	TWQKLE-VI-	-K-QF-R-HP	KY	---	N	---	T	-----	S	---	TL	---	HHDI	
1GVZ	A	16	246	TAQFHQ-VS-	-D-SF-L-DP	QF	---	DLS	---	L	LKKKYLRPYD	---	D	---	DI	---	SHDL
1DLE	A	1K	250	--RDLE-IE-	-V-VL-F-HP	NY	---	N	---	I	NGKKEAGIPE	---	F	---	FY	---	DYDV
1PYT	D701	944		GSLYVG-VD-	-T-IF-V-HE	KW	---	N	---	S	-----	F	---	LV	---	RNDI	
2O8L	A	1	216	PNGGFT-AE-	-Q-IT-K-YS	G	---		---		-----		---		---	EGDL	
1P3C	A	1	215	VNGSAN-MT-	-E-FY-V-PS	GYINTG	---		---		-----		---	AS	---	QYDF	
1AGJ	A	1	242	PYGKFE-VK-	-E-ILQEPFG	A	---		---		-----		---		---	GVDL	
1QTF	A	1	246	PYGKFE-AE-	-E-IKESPYG	Q	---		---		-----		---		---	GLDL	
1FON	A	9	240	SEQVIP-INA	GD-LF-V-HP	LW	---	N	---	S	-----	N	---	CVAC	---	CGNDI	
2SGA	A	16	242	----I--GT-	-R-TG-T-SF	P	---		---		-----		---		---	NNDY	
2SFA	A	1	191	----L--GT-	-R-AG-T-SF	P	---		---		-----		---		---	GNDY	
1SGP	E	16	242	----L--GT-	-T-SG-S-SF	P	---		---		-----		---		---	NNDY	
1HPG	A	16	242	----V--GV-	-R-EG-T-SF	P	---		---		-----		---		---	TNDY	
2H5C	A	15A	245	----V--GT-	-F-AA-R-VF	P	---		---		-----		---		---	GNDR	
1Y8T	A	6	314	-RTAPF-T--	-V-VG-A-D-	PT	---		---		-----		---		---	SDI	



1P3C	A	1	215	AVIK <b>KT</b> DT---	--N---	IGN	T---	VGY--	-----	R-	SIR--	Q----	
1AGJ	A	1	242	ALIR <b>LK</b> P <b>DQ</b> N	GVS---	L-G	DK---	ISP--	-----	A-	KIG--	T----	
1QTF	A	1	246	AI <b>IKL</b> K <b>PNE</b> K	GES---	A-G	DL---	IQP--	-----	A-	NIP--	D----	
1FON	A	9	240	ALV <b>KL</b> S <b>R</b> ---	--S---	A-Q	LG <b>DK</b> V--	QL--	-----	A-	NLP--	P----	
2SGA	A	16	242	<b>GI</b> IR <b>H</b> S <b>N</b> ---	--PA--	A-A	---	<b>DGR</b> VY	LY <b>NG</b> S <b>Y</b> Q <b>D</b> IT	TAG	---	---	
2SFA	A	1	191	<b>GL</b> IR <b>H</b> S <b>N</b> ---	--AS--	A-A	---	<b>DGR</b> VY	LY <b>NG</b> S <b>Y</b> R <b>D</b> IT	GAG	---	---	
1SGP	E	16	242	<b>GIV</b> R <b>Y</b> T <b>N</b> ---	--TT--	IPK	---	<b>DGT</b> VG	---	<b>G</b> Q <b>D</b> IT	SAA	---	
1HPG	A	16	242	<b>GIV</b> R <b>Y</b> T <b>D</b> ---	--GS--	S-P	---	<b>AGT</b> VD	LY <b>NG</b> S <b>T</b> Q <b>D</b> IS	SAA	---	---	
2H5C	A	15A	245	<b>AW</b> V <b>S</b> L <b>T</b> S---	--AQ--	T-L	---	L <b>P</b> R <b>V</b> A	NG- <b>S</b> S <b>F</b> V <b>T</b> V <b>R</b>	<b>G</b> S <b>T</b> ---	---	---	
1Y8T	A	6	314	<b>AV</b> V <b>R</b> V-- <b>Q</b> ---	--G---	V-S	-G---	L <b>T</b> P--	-----	I-	SLG--	S----	
1SOT	A	43	353	<b>AV</b> L <b>K</b> I-- <b>N</b> ---	--A---	T-G	G---	L <b>P</b> T--	-----	I-	PIN--	A----	
1DY9	A	1	175	<b>L</b> V <b>G</b> W <b>P</b> -----	--A---	P--	PG--	<b>A</b> R <b>S</b> --	-----	M-	T <b>P</b> C--	T----	
1ARB	A	1	263	<b>T</b> L <b>L</b> E <b>L</b> N <b>N</b> ---	--A---	A-N	PA-F <b>N</b> L <b>F</b> W--	---	-----	A-	<b>G</b> W <b>D</b> --	<b>R</b> ----	
1MBM	A	7	204	<b>F</b> A <b>E</b> A <b>V</b> -----	--TTQS <b>E</b> L--	PG--	<b>N</b> W <b>P</b> Q--	---	-----	L-	<b>H</b> F <b>A</b> --	<b>Q</b> ----	
2FOM	B	18	167	<b>L</b> I <b>S</b> Y <b>G</b> -----	--G---	G--	---	W--	-----	K-	<b>L</b> E <b>G</b> --	<b>E</b> ----	
2FP7	B	19	170	<b>R</b> L <b>C</b> Y <b>G</b> -----	--G---	P--	---	W--	-----	K-	<b>L</b> Q <b>H</b> --	<b>K</b> ----	
1EP5	A	120	275	<b>L</b> E <b>Y</b> A <b>D</b> -- <b>V</b> ---	--P <b>Q</b> --	<b>N</b> - <b>M</b>	<b>R</b> A---	<b>D</b> T--	-----	F-	<b>K</b> Y <b>T</b> --	<b>H</b> ----	
1GCT	A	1	245	-----	--AS--	<b>D</b> -- <b>D</b> --	--F-	<b>A</b> A <b>G</b> T <b>T</b> --	<b>C</b> V <b>T</b> T <b>G</b> -- <b>W</b> ---	--G	-----	<b>L</b>	
1K2I	I	1	245	-----	--AS--	<b>D</b> -- <b>D</b> --	--F-	<b>A</b> A <b>G</b> T <b>T</b> --	<b>C</b> V <b>T</b> T <b>G</b> -- <b>W</b> ---	--G	-----	<b>L</b>	
1CHO	E	1	245	-----	--AS--	<b>D</b> -- <b>D</b> --	--F-	<b>A</b> A <b>G</b> T <b>T</b> --	<b>C</b> V <b>T</b> T <b>G</b> -- <b>W</b> ---	--G	-----	<b>L</b>	
1H4W	A	16	246	-----	----A-	P-	--P-	<b>A</b> A <b>G</b> T <b>E</b> --	<b>C</b> L <b>I</b> S <b>G</b> -- <b>W</b> ---	--G	-----	<b>N</b>	
1HJ9	A	16	245	-----	----S-	<b>C</b> --	--A-	<b>S</b> A <b>G</b> T <b>Q</b> --	<b>C</b> L <b>I</b> S <b>G</b> -- <b>W</b> ---	--G	-----	<b>N</b>	
2A31	A	16	245	-----	----S-	<b>C</b> --	--A-	<b>A</b> A <b>G</b> T <b>E</b> --	<b>C</b> L <b>I</b> S <b>G</b> -- <b>W</b> ---	--G	-----	<b>N</b>	
1TRN	A	16	246	-----	----A-	P-	--P-	<b>A</b> T <b>G</b> T <b>K</b> --	<b>C</b> L <b>I</b> S <b>G</b> -- <b>W</b> ---	--G	-----	<b>N</b>	
1FY8	E	14	245	-----	----S-	<b>C</b> --	--A-	<b>P</b> A <b>G</b> T <b>Q</b> --	<b>C</b> L <b>I</b> S <b>G</b> -- <b>W</b> ---	--G	-----	<b>N</b>	
1A0J	A	16	245	-----	----S-	<b>C</b> --	--A-	<b>S</b> S <b>G</b> T <b>R</b> --	<b>C</b> L <b>V</b> S <b>G</b> -- <b>W</b> ---	--G	-----	<b>N</b>	
1HJ8	A	16	245	-----	----S-	<b>C</b> --	--A-	<b>P</b> A <b>G</b> T <b>M</b> --	<b>C</b> T <b>V</b> S <b>G</b> -- <b>W</b> ---	--G	-----	<b>N</b>	
2F91	A	16	243	-----	<b>Q</b> G--	<b>H</b> -- <b>T</b> --	--A-	<b>T</b> -- <b>G</b> <b>D</b> --	<b>V</b> I <b>V</b> T <b>G</b> -- <b>W</b> ---	--G	-----	<b>T</b>	
1NPM	A	16	243	-----	----L-	<b>C</b> --	--P-	<b>K</b> V <b>G</b> Q <b>K</b> --	<b>C</b> I <b>I</b> S <b>G</b> -- <b>W</b> ---	--G	-----	<b>T</b>	
1EUF	A	16	243	-----	<b>S</b> L--	<b>A</b> -- <b>E</b> --	--V-	<b>K</b> P <b>G</b> M <b>M</b> --	<b>C</b> S <b>V</b> A <b>G</b> -- <b>W</b> ---	--G	-----	<b>R</b>	
3RP2	A	16	243	-----	<b>P</b> S--	<b>D</b> -- <b>F</b> --	--I-	<b>H</b> P <b>G</b> A <b>M</b> --	<b>C</b> W <b>A</b> A <b>G</b> -- <b>W</b> ---	--G	-----	<b>K</b>	
1IAU	A	16	244	-----	<b>N</b> K--	<b>A</b> -- <b>Q</b> --	--V-	<b>K</b> P <b>G</b> Q <b>T</b> --	<b>C</b> S <b>V</b> A <b>G</b> -- <b>W</b> ---	--G	-----	<b>Q</b>	
1GVK	B	16	245	-----	<b>A</b> G--	<b>T</b> -- <b>I</b> --	--L-	<b>A</b> N <b>N</b> S <b>P</b> --	<b>C</b> Y <b>I</b> T <b>G</b> -- <b>W</b> ---	--G	-----	<b>L</b>	
1ELT	A	16	242	-----	<b>S</b> N--	<b>Q</b> -- <b>I</b> --	--L-	<b>P</b> N <b>N</b> N <b>P</b> --	<b>C</b> Y <b>I</b> T <b>G</b> -- <b>W</b> ---	--G	-----	<b>K</b>	
1BRU	P	16	245	-----	<b>A</b> G--	<b>T</b> -- <b>I</b> --	--L-	<b>P</b> N <b>N</b> Y <b>V</b> --	<b>C</b> Y <b>V</b> T <b>G</b> -- <b>W</b> ---	--G	-----	<b>R</b>	
1EKB	B	16	243	-----	<b>E</b> N--	<b>Q</b> -- <b>V</b> --	--F-	<b>P</b> P <b>G</b> R <b>I</b> --	<b>C</b> S <b>I</b> A <b>G</b> -- <b>W</b> ---	--G	-----	<b>A</b>	
1EAX	A	16	244	-----	<b>A</b> S--	<b>H</b> -- <b>V</b> --	--F-	<b>P</b> A <b>G</b> K <b>A</b> --	<b>I</b> W <b>V</b> T <b>G</b> -- <b>W</b> ---	--G	-----	<b>H</b>	
1H8D	H	16	246	-----	---- <b>R</b> E <b>T</b> A <b>A</b>	<b>S</b> L <b>L</b> --	<b>Q</b> A <b>G</b> Y <b>K</b> --	<b>G</b> R <b>V</b> T <b>G</b> --	<b>W</b> ---	--G	-----	<b>N</b>	
1VR1	H	16	243	-----	---- <b>R</b> E <b>T</b> A <b>A</b>	<b>S</b> L <b>L</b> --	<b>Q</b> A <b>G</b> Y <b>K</b> --	<b>G</b> R <b>V</b> T <b>G</b> --	<b>W</b> ---	--G	-----	<b>N</b>	
1C5L	H	16	247	-----	---- <b>R</b> E <b>T</b> A <b>A</b>	<b>S</b> L <b>L</b> --	<b>Q</b> A <b>G</b> Y <b>K</b> --	<b>G</b> R <b>V</b> T <b>G</b> --	<b>W</b> ---	--G	-----	<b>N</b>	
2PUX	B	16	246	-----	---- <b>K</b> Q <b>T</b> V <b>T</b>	<b>S</b> L <b>L</b> --	<b>R</b> A <b>G</b> Y <b>K</b> --	<b>G</b> R <b>V</b> T <b>G</b> --	<b>W</b> ---	--G	-----	<b>N</b>	
1ETR	H	16	247	-----	---- <b>K</b> Q <b>T</b> A <b>A</b>	<b>K</b> L <b>L</b> --	<b>H</b> A <b>G</b> F <b>K</b> --	<b>G</b> R <b>V</b> T <b>G</b> --	<b>W</b> ---	--G	-----	<b>N</b>	
1HNE	E	16	243	-----	<b>Q</b> G--	<b>R</b> -- <b>R</b> --	--L-	<b>G</b> N <b>G</b> V <b>Q</b> --	<b>C</b> L <b>A</b> M <b>G</b> -- <b>W</b> ---	--G	-----	<b>L</b>	
1FUJ	A	16	243	-----	<b>Q</b> D--	<b>Q</b> -- <b>P</b> --	--V-	<b>P</b> H <b>G</b> T <b>Q</b> --	<b>C</b> L <b>A</b> M <b>G</b> -- <b>W</b> ---	--G	-----	<b>R</b>	
1A7S	A	1	225	-----	<b>Q</b> N--	<b>A</b> -- <b>T</b> --	--V-	<b>E</b> A <b>G</b> T <b>R</b> --	<b>C</b> Q <b>V</b> A <b>G</b> -- <b>W</b> ---	--G	-----	<b>S</b>	
1AZZ	A	16	246	-----	---- <b>T</b> --	<b>D</b> --	--V-	<b>G</b> V <b>G</b> T <b>V</b> --	<b>V</b> T <b>P</b> T <b>G</b> -- <b>W</b> ---	--G	-----	<b>L</b>	
1LO6	A	16	243	-----	---- <b>D</b> C--	<b>S</b> --	--A-	<b>N</b> - <b>T</b> T <b>S</b> --	<b>C</b> H <b>I</b> L <b>G</b> -- <b>W</b> ---	--G	-----	<b>K</b>	
2PKA	A	16	246	-----	---- <b>Q</b> --	<b>E</b> --	--P-	<b>E</b> L <b>G</b> S <b>T</b> --	<b>C</b> E <b>A</b> S <b>G</b> -- <b>W</b> ---	--G	-----	<b>S</b>	
1T32	A	16	244	-----	<b>A</b> Q--	<b>E</b> -- <b>G</b> --	--L-	<b>R</b> P <b>G</b> T <b>L</b> --	<b>C</b> T <b>V</b> A <b>G</b> -- <b>W</b> ---	--G	-----	<b>R</b>	
1Q3X	A	362	686	-----	<b>K</b> E--	<b>A</b> -- <b>E</b> S	F-M-	<b>R</b> T <b>D</b> D <b>I</b> --	<b>G</b> T <b>A</b> S <b>G</b> -- <b>W</b> ---	--G	-----	<b>L</b>	
1EQ9	A	16	244	-----	---- <b>N</b> --	<b>D</b> --	-- <b>E</b> DL <b>E</b> S <b>N</b> P--	<b>C</b> T <b>L</b> T <b>G</b> -- <b>W</b> ---	--G	-----	<b>S</b>		
1FIW	A	16	254	-----	<b>F</b> R--	<b>A</b> G--	P-	--P-	<b>R</b> V <b>P</b> Q <b>T</b> --	<b>C</b> W <b>V</b> A <b>G</b> -- <b>W</b> ---	--G	-----	<b>F</b>
1BIO	A	16	243	-----	---- <b>V</b> D <b>R</b> D--	---	--V-	<b>A</b> P <b>G</b> T <b>L</b> --	<b>C</b> D <b>V</b> A <b>G</b> -- <b>W</b> ---	--G	-----	<b>I</b>	
1OP0	A	16	245	-----	---- <b>S</b> --	P-	--P-	<b>S</b> V <b>G</b> S <b>V</b> --	<b>C</b> R <b>I</b> M <b>G</b> -- <b>W</b> ---	--G	-----	<b>S</b>	
1BQY	A	16	245	-----	---- <b>S</b> --	P-	--P-	<b>S</b> V <b>G</b> S <b>V</b> --	<b>C</b> R <b>I</b> M <b>G</b> -- <b>W</b> ---	--G	-----	<b>K</b>	
1PQ7	A	16	239	-----	<b>S</b> G--	<b>S</b> -- <b>D</b> --	--P-	<b>V</b> A <b>G</b> S <b>S</b> --	<b>A</b> T <b>V</b> A <b>G</b> -- <b>W</b> ---	--G	-----	<b>A</b>	
1OS8	A	16	245	-----	---- <b>T</b> --	<b>T</b> A	<b>Y</b> - <b>N</b> - <b>Q</b> --	<b>G</b> T--	<b>F</b> T <b>V</b> A <b>G</b> -- <b>W</b> ---	--G	-----	<b>A</b>	
2FPZ	A	16	244	-----	<b>A</b> S--	<b>E</b> -- <b>T</b> --	--F-	<b>P</b> P <b>G</b> M <b>P</b> --	<b>C</b> W <b>V</b> T <b>G</b> -- <b>W</b> ---	--G	-----	<b>D</b>	
1LTO	A	16	244	-----	<b>A</b> S--	<b>E</b> -- <b>T</b> --	--F-	<b>P</b> P <b>G</b> M <b>P</b> --	<b>C</b> W <b>V</b> T <b>G</b> -- <b>W</b> ---	--G	-----	<b>D</b>	
1C5Y	B	16	245	-----	<b>M</b> Y--	<b>N</b> -- <b>D</b> --	--P-	<b>Q</b> F <b>G</b> T <b>S</b> --	<b>C</b> E <b>I</b> T <b>G</b> -- <b>F</b> ---	--G	-----	<b>K</b>	

1TON	A	16	246	-----	----	K--E-	--P-KVGST-	CLASG-W---	--G-----	S	
1M9U	A	16	242	-----	-----	-----	----NNNN-	--D-YAGTT-	CVISG-W---	--G-----	R
2BZ6	H	16	257	-----	-----	RTF-S--ER	TLA-F-VRF-	SLVSG-W---	--G-----	Q	
1RTF	B	16	244	-----	-----	AD-L----	QLPDWTE-	CELSG-Y---	--G-----	K	
2HLC	A	16	246	-----	-----	GE-ELNN-	-K-FENIW-	ATVSG-W---	--G-----	Q	
1NN6	A	18	241	-----	-----	-----	-N-F----	-V-PPGRM-	CRVAG-W---	--GRTGVLP	P
1GVZ	A	16	246	-----	-----	-----	----Q-E-	--P-KLGST-	CYTSG-W---	--G-----	L
1DLE	A	1K250		TRALRLPPTT	TCQQQK	--EE	L-LPAQDIK-	ALFVS-E---	-----	-----	
1PYT	D701	944		-----	-----	-----	----G-SL-	--L-PQDYP-	CFVTG-W---	--G-----	R
2O8L	A	1	216	-----	-----	-N--A-E-	-T-QTNQN-	ITVTG-Y---	--P-----	G	
1P3C	A	1	215	-----	-----	-----	----V-T-	--N-LTGTT-	IKISG-Y---	--PG-----	D
1AGJ	A	1	242	-----	-----	-S--N-D-	--L-KDGDK-	LELIG-Y---	--P-----	F	
1QTF	A	1	246	-----	-----	-H--I-D-	--I-AKGDK-	YSLIG-Y---	--P-----	Y	
1FON	A	9	240	-----	-----	-----	----AG-DI	--L-PNEAP-	CYISG-W---	--G-----	R
2SGA	A	16	242	-----	-----	-N-----	--A-FVGQA-	VQRSG-S---	--T-----		
2SFA	A	1	191	-----	-----	-N-----	--A-YVGQT-	VQRSG-S---	--T-----		
1SGP	E	16	242	-----	-----	-N-----	--A-TVGMA-	VTRRG-S---	--T-----		
1HPG	A	16	242	-----	-----	-N-----	--A-VVGQA-	IKKSG-S---	--T-----		
2H5C	A	15A245		-----	-----	-E-----	--A-AVGAA-	VCRSG-R---	--T-----		
1Y8T	A	6	314	-----	-----	-S--S-D-	--L-RVGQP-	VLAIG-S---	--P-----	L	
1SOT	A	43	353	-----	-----	-R--R-V-	--P-HIGDV-	VLAIGNP---	--Y-----		
1DY9	A	1	175	-----	-----	-----	-C--G-	-S-S--D-	LYLVT-R---	--H-----	
1ARB	A	1	263	-----	-----	-R--D-Q-	-N-Y--PG-	AIAIH-H---	--P-----	N	
1MBM	A	7	204	-----	-----	-----	-P--T-	-T-G--P-	ASWCT-A---	--T-----	
2FOM	B	18	167	-----	-----	-----	-----	-W--K-EG--EE	VQVLA-LEPG	KNP	-----
2FP7	B	19	170	-----	-----	-----	-----	-W--N-GH--DE	VQMIV-VEPG	KNV	-----
1EP5	A120	275		-----	-----	-E-----	--K-PQ-GY-	YSWH-----	-----	-----	

1GCT	A	1	245	TR--Y----	-----	TP	--D-RLQQ	ASLPLL---	S NT	-----	
1K2I	1	1	245	TR--Y----	-----	TP	--D-RLQQ	ASLPLL---	S NT	-----	
1CHO	E	1	245	TR--Y--AN-	-----	TP	--D-RLQQ	ASLPLL---	S NT	-----	
1H4W	A	16	246	TL-SF--GA	-----	DYP	--D-ELKC	LDAPVL---	T QA	-----	
1HJ9	A	16	245	TK-SS--GT	-----	SYP	--D-VLKC	LKAPIL---	S DS	-----	
2A31	A	16	245	TK-SS--GS	-----	SYP	--S-LLQC	LKAPVL---	S DS	-----	
1TRN	A	16	246	TA-SS--GA	-----	D-P	--D-ELQC	LDAPVL---	S QA	-----	
1FY8	E	14	245	-----	-----	P	--D-LLQC	LDAPLL---	P QA	-----	
1A0J	A	16	245	LS-GS--SS	-----	NYP	--D-TLRC	LDPIL---	S SS	-----	
1HJ8	A	16	245	TM-SS--TA	-----	D-S	--N-KLQC	LNIPIL---	S YS	-----	
2F91	A	16	243	TS-E--GG	-----	NTP	--D-VLQK	VTVPLV---	S DE	-----	
1NPM	A	16	243	VT-SP--QE	-----	NFP	--N-TLNC	AEVKIY---	S QN	-----	
1EUF	A	16	243	LG-V--NM	-----	PST	--D-KLQE	VDLEVQ---	S EE	-----	
3RP2	A	16	243	TG-V--RD	-----	PTS	--Y-TLRE	VELRIM---	D EK	-----	
1IAU	A	16	244	TA-P--LG	-----	KHS	--H-TLQE	VKMTVQ---	E DR	-----	
1GVK	B	16	245	TR-T--NG	-----	QLA	--Q-TLQQ	AYLPTV---	D YA	-----	
1ELT	A	16	242	TS-T--GG	-----	PLS	--D-SLKQ	AWLPSV---	D HA	-----	
1BRU	P	16	245	LQ-T--NG	-----	ASP	--D-ILQQ	GQLLVV---	D YA	-----	
1EKB	B	16	243	LI-Y--QG	-----	STA	--D-VLQE	ADVPLL---	S NE	-----	
1EAX	A	16	244	TQ-Y--GG	-----	TGA	--L-ILQK	GEIRVI---	N QT	-----	
1H8D	H	16	246	LK-E--T	-----	GQP	--S-VLQV	VNLPIV---	E RP	-----	
1VR1	H	16	243	LK-E--T	-----	GQP	--S-VLQV	VNLPIV---	E RP	-----	
1C5L	H	16	247	LK-E--T	-----	GQP	--S-VLQV	VNLPIV---	E RP	-----	
2PUX	B	16	246	LR-E--TWT	TNINEIQP	-----	S--VLQV	VNLPIV---	E RP	-----	
1ETR	H	16	247	RR-E--TWT	TSVAEVQP	-----	S--VLQV	VNLPLV---	E RP	-----	
1HNE	E	16	243	L--GR--NR	-----	GIA	--S-VLQE	LNVTVV---	TS	-----	
1FUJ	A	16	243	V--GA--HD	-----	PPA	--Q-VLQE	LNVTVV---	TF	-----	
1A7S	A	1	225	Q--RS--GG	-----	RLS	--R-FPRF	VNVTVT---	P ED	-----	
1AZZ	A	16	246	PS-D--SAL	-----	GIS	--D-VLRQ	VDVPIM---	S NA	-----	
1LO6	A	16	243	TA-DG--D	-----	FP	--D-TIQC	AYIHLV---	S RE	-----	
2PKA	A	16	246	IE-PGPDDF	-----	EFP	--D-EIQC	VQLTLL---	Q NT	-----	
1T32	A	16	244	VS-----MR	-----	RGT	--D-TLRE	VQLRVQ---	R DR	-----	
1Q3X	A362	686		TQ-----RG	-----	FLA	--R-NLMY	VDIPIV---	D HQ	-----	

1EQ9	A	16	244	TR-L-GG	----	NTP	---	N-ALQE	IELIVH	---	P	QK	-----	
1FIW	A	16	254	LQ-E-NAR	----	RTS	---	P-MLQE	ARVDLI	---	D	LG	-----	
1BIO	A	16	243	VN-H-AG	----	RRP	---	D-SLQH	VLLPVL	---	D	RA	-----	
1OP0	A	16	245	IT-PV-KE	----	TFP	---	D-VPYC	ANINLL	---	D	HA	-----	
1BQY	A	16	245	TI-PT-KE	----	IYP	---	D-VPHC	ANINIL	---	D	HA	-----	
1PQ7	A	16	239	TS-E-GGS	----	STP	---	V-NLLK	VTVPIV	---	S	RA	-----	
1OS8	A	16	245	NR-E-GG	----	SQQ	---	R-YLLK	ANVPFV	---	S	DA	-----	
2FPZ	A	16	244	VD-N-DE	----	RLPPP	---	F-PLKQ	VKVPIM	---	E	NH	-----	
1LTO	A	16	244	VD-N-DE	----	PLPPP	---	F-PLKQ	VKVPIM	---	E	NH	-----	
1C5Y	B	16	245	EA-S-TDY	----	LYP	---	E-QLKM	TVVKLI	---	S	HR	-----	
1TON	A	16	246	TNPS-EM	----	VVS	---	H-DLQC	VNIHLL	---	S	NE	-----	
1M9U	A	16	242	TD-G-TN	----	NLP	---	D-ILQK	SSIPVI	---	T	TA	-----	
2BZ6	H	16	257	LL-D-RG	----	ATA	---	L-ELMV	LNVPRL	---	M	TQ	-----	
1RTF	B	16	244	HE-A-LSP	----	FYS	---	E-RLKE	AHVRLY	---	P	SS	-----	
2HLC	A	16	246	SN- -TD	----	T	---	V-ILQY	TYNLVI	---	D	ND	-----	
1NN6	A	18	241	G- - - -	----	S	---	D-TLQE	VKLRIM	---	D	PQ	-----	
1GVZ	A	16	246	ISTFTNRG	----	S	---	G-TLQC	VELRLQ	---	S	NE	-----	
1DLE	A	1K250			----	EE	---	K-KLTR	KEVYIKNGDK	---	KGS	-----		
1PYT	D701	944		LYTNGPI	----	A	---	A-ELQQ	GLQPVV	---	D	YA	-----	
208L	A	1	216		----	DK	---	PVATMWE	SKGKIT	---	Y	L	-----	
1P3C	A	1	215	K- - - -	----	MRS	TGKV	SQWE	MSGSVT	---	R	E	-----	
1AGJ	A	1	242	D- - - -	----	HK	---	VNQMR	SEIELT	---	T	L	-----	
1QTF	A	1	246	N- - - -	----	YS	---	AYSLYQ	SQIEMF	---	N	D	-----	
1FON	A	9	240	L- - - -	----	YTG	GPLPD	KLQQ	ALLPTV	---	D	YE	-----	
2SGA	A	16	242		----		---	TGL	RSGSVT	---	G	L	-----	
2SFA	A	1	191		----		---	TGL	HSGRVT	---	G	L	-----	
1SGP	E	16	242		----		---	TGT	HSGSVT	---	A	L	-----	
1HPG	A	16	242		----		---	TKV	TSGTVT	---	A	V	-----	
2H5C	A	15A245			----		---	TGY	QCGTIT	---	A	K	-----	
1Y8T	A	6	314		----	GL	---	EGTV	TTGIVS	---	AL	---	-----	
1SOT	A	43	353		----	N	---	L-GQTI	TQGIIS	---	A	TGRIGLNPTG	-----	
1DY9	A	1	175		----		---	AD	VIPVRR	---	R	---	-----	
1ARB	A	1	263		----	VA	---	EKRI	SNSTSP	---	T	-S	-----	
1MBM	A	7	204		----		---	GD	EEGLLS	---	---	---	-----	
2FOM	B	18	167		----		---	RA	VQTKPG	---	L	FKTNTG	-----	
2FP7	B	19	170		----		---	KN	VQTKPG	---	V	FKTPEG	-----	
1EP5	A120	275			----		---	H	GAVQYE	---	N	G	-----	
1GCT	A	1	245	NCK-K-YW	----	G	---	T	---	---	K	I	K	-----
1K2I	1	1	245	NCK-K-YW	----	G	---	T	---	---	K	I	K	-----
1CHO	E	1	245	NCK-K-YW	----	G	---	T	---	---	K	I	K	-----
1H4W	A	16	246	ECK-A-SY	----	P	---	G	---	---	K	I	T	-----
1HJ9	A	16	245	SCK-S-AY	----	P	---	G	---	---	Q	I	T	-----
2A31	A	16	245	SCK-S-SY	----	P	---	G	---	---	Q	I	T	-----
1TRN	A	16	246	KCE-A-SY	----	P	---	G	---	---	K	I	T	-----
1FY8	E	14	245	DCE-A-SY	----	P	---	G	---	---	K	I	T	-----
1A0J	A	16	245	SCN-S-AY	----	P	---	G	---	---	Q	I	T	-----
1HJ8	A	16	245	DCN-N-SY	----	P	---	G	---	---	M	I	T	-----
2F91	A	16	243	DCR-A-DY	----	GA	---	D	---	---	E	I	L	-----
1NPM	A	16	243	KCE-R-AY	----	P	---	G	---	---	K	I	T	-----
1EUF	A	16	243	KCI-A-RF	----		---	K	---	---	N	Y	IP	-----
3RP2	A	16	243	ACVDY-RY	----		---		---	---	---	Y	EY	-----
1IAU	A	16	244	KCE-S-DL	----	R	---	H	---	---	Y	Y	DS	-----
1GVK	B	16	245	ICS-SSSYW	----	G	---	S	---	---	T	V	K	-----
1ELT	A	16	242	TCS-SSGWW	----	G	---	S	---	---	T	V	K	-----
1BRU	P	16	245	TCS-KPGWW	----	G	---	S	---	---	T	V	K	-----
1EKB	B	16	243	KCQ-Q-QM	----	P	---	E	---	---	YN	I	T	-----
1EAX	A	16	244	TCE-N-LL	----	P	---	Q	---	---	Q	I	T	-----
1H8D	H	16	246	VCK-D-ST	----	R	---	I	---	---	R	I	T	-----
1VR1	H	16	243	VCK-D-ST	----	R	---	I	---	---	R	I	T	-----
1C5L	H	16	247	VCK-D-ST	----	R	---	I	---	---	R	I	T	-----



2PUX	B	16	246	VCK-A--ST-	---	R	---	I	---	---	---	R	I	---	T	---
1ETR	H	16	247	VCK-A--ST-	---	R	---	I	---	---	---	R	I	---	T	---
1HNE	E	16	243	LC	-----								R	-----		
1FUJ	A	16	243	FC	-----								R	-----		
1A7S	A	1	225	QC	-----								R	-----		
1AZZ	A	16	246	DCD-A--VY-	---			G	---			I	V	---	T	---
1LO6	A	16	243	ECE-H--AY-	---	P	---	G	---			Q	I	---	T	---
2PKA	A	16	246	FCA-D--AH-	---	P	---	D	---			K	V	---	T	---
1T32	A	16	244	QCL-R--IF-	---			G	---			S	Y	---	DP	---
1Q3X	A362	686		KCT-A--AY-	---	EKPPYP	R	G	---			S	V	---	T	---
1EQ9	A	16	244	QCE-R--DQ-	---	W	---					R	V	---	I	---
1FIW	A	16	254	LCN-STRWY-	---	N	---	G	---			R	I	---	R	---
1BIO	A	16	243	TCN-RRTHH-	---	D	---	G	---			A	I	---	T	---
1OP0	A	16	245	VCQ-A--GY-	---	PE	---	L	---			L	A	---	E	---
1BQY	A	16	245	VCR-T--AY-	---	SW	---	R	---			Q	V	---	A	---
1PQ7	A	16	239	TCR-A--QY-	---	GT	---	S	---			A	I	---	T	---
1OS8	A	16	245	ACR-S--AY-	---	G	---	N	---			E	L	---	VA	---
2FPZ	A	16	244	ICD-A--KY-	---	HLGAYT	GDDVR					I	V	---	R	---
1LTO	A	16	244	ICD-A--KY-	---	HLGAYT	GDDVR					I	I	---	R	---
1C5Y	B	16	245	ECQ-QPHY-	---	G	---	S	---			E	V	---	T	---
1TON	A	16	246	KC	---	IE	---								TYKDNVT	---
1M9U	A	16	242	QCT-A--AMV	GV--GG	---		A	---			N	I	---	W	---
2BZ6	H	16	257	DCL-Q--QSR	KVGD	S	---	P	---			N	I	---	T	---
1RTF	B	16	244	RCTSQ--HLL	N--R	---						T	V	---	T	---
2HLC	A	16	246	RCA-Q--EY-	---			PPG	---			I	I	---	V	---
1NN6	A	18	241	A-C-S--H-	---	F	---	R	---			D	F	---	DH	---
1GVZ	A	16	246	KCA-R--AY-	---	P	---	E	---			K	M	---	T	---
1DLE	A	1K250		CER	---	DA	---	QYAPGY	DKVKDISEV	---		V	V	---	T	---
1PYT	D701	944		TCS-QRDWW	---	G	---	T	---			T	V	---	K	---
2O8L	A	1	216												K	---
1P3C	A	1	215												D	---
1AGJ	A	1	242												S	---
1QTF	A	1	246													
1FON	A	9	240	HCS-QWDWW	---	G	---	I	---			T	V	---	K	---
2SGA	A	16	242												NATVN	---
2SFA	A	1	191												NATVN	---
1SGP	E	16	242												NATVN	---
1HPG	A	16	242												NVTVN	---
2H5C	A	15A245													NVTAN	---
1Y8T	A	6	314	NR	---										P	---
1SOT	A	43	353	RQ	---										N	---
1DY9	A	1	175												G	---
1ARB	A	1	263	FVA	---	W	---	G	---			G	G	---	AG	---
1MBM	A	7	204												G	---
2FOM	B	18	167												T	---
2FP7	B	19	170												E	---
1EP5	A120	275														
1GCT	A	1	245			D	---	AMICAG	A	---		S-G-V-S	---	S	CMGDSGG	
1K2I	1	1	245			D	---	AMICAG	A	---		S-G-V-S	---	S	CMGDSGG	
1CHO	E	1	245			D	---	AMICAG	A	---		S-G-V-S	---	S	CMGDSGG	
1H4W	A	16	246			N	---	SMFCVG	F-L-E	---		G-G-K-D	---	S	CQRDSGG	
1HJ9	A	16	245			S	---	NMFCAG	YG-L-E	---		G-KG-D	---	S	CQGDSSG	
2A31	A	16	245			G	---	NMICVG	F-L-E	---		G-G-K-D	---	S	CQGDSSG	
1TRN	A	16	246			S	---	NMFCVG	F-L-E	---		G-G-K-D	---	S	CQGDSSG	
1FY8	E	14	245			D	---	NMVCVG	F-L-E	---		G-G-K-D	---	S	CQGDSSG	
1A0J	A	16	245			S	---	NMFCAG	F-M-E	---		G-G-K-D	---	S	CQGDSSG	
1HJ8	A	16	245			N	---	AMFCAG	Y-L-E	---		G-G-K-D	---	S	CQGDSSG	
2F91	A	16	243			D	---	SMICA	-GVP-E	---		G-G-K-D	---	S	CQGDSSG	
1NPM	A	16	243			E	---	GMVCAG	S-S	---		N-G-A-D	---	T	CQGDSSG	
1EUF	A	16	243			F	---	TQICAG	D-P-S	---		K-R-K-N	---	S	FSGDSGG	



3RP2	A	16	243	-----K--	----FQVCVG	S--P-T----	T-L-R--A--	-A-FMGDSGG
1IAU	A	16	244	-----T--	----IELCVG	D--P-E----	I-K-K--T--	-S-FKGDSSG
1GVK	B	16	245	-----N--	----SMVCAG	G----D----	G-V-R--S--	-G-CQGDSSG
1ELT	A	16	242	-----T--	----TMVCAG	G----G----	--A-N--S--	-G-CNGDSSG
1BRU	P	16	245	-----T--	----NMICAG	G----D----	G-I-I--S--	-S-CNGDSSG
1EKB	B	16	243	-----E--	----NMVCAG	Y--E-A----	G-G-V--D--	-S-CQGDSSG
1EAX	A	16	244	-----P--	----RMMC VG	F--L-S----	G-G-V--D--	-S-CQGDSSG
1H8D	H	16	246	-----D--	----NMFCA-	-GYK-PDEGK	R---G--D--	-A-CEGDSSG
1VR1	H	16	243	-----D--	----NMFCA-	-YYK-PDEGK	R---G--D--	-A-CEGDSSG
1C5L	H	16	247	-----D--	----NMFCA-	-GYK-PDEGK	R---G--D--	-A-CEGDSSG
2PUX	B	16	246	-----D--	----NMFCA-	-GFK-VNDTK	R---G--D--	-A-CEGDAGG
1ETR	H	16	247	-----D--	----NMFCA-	-GYK-PGEGK	R---G--D--	-A-CEGDSSG
1HNE	E	16	243	-----R--	----SNVCTL	V---R----	GRQ-A--G--	-V-CFGDSSG
1FUJ	A	16	243	-----P--	----HNICTF	V---P----	RRK-A--G--	-I-CFGDSSG
1A7S	A	1	225	-----P--	----NNVCTG	V---L----	TRR-G--G--	-I-CNGDGGT
1AZZ	A	16	246	-----D--	----GNICID	S---T----	G-G-K--G--	-T-CNGDSSG
1LO6	A	16	243	-----Q--	----NMLCAG	D-----	EKYGK--D--	-S-CQGDSSG
2PKA	A	16	246	-----E--	----SMLCAG	Y--L-P----	G-G-K--D--	-T-CMGDSSG
1T32	A	16	244	-----R--	----RQICVG	D--R-R----	E-R-K--A--	-A-EKGDSSG
1Q3X	A362	686		-----A--	----NMLCAG	L--E-S----	G-G-K--D--	-S-CRGDSSG
1EQ9	A	16	244	-----D--	----SHICTL	T---K----	R-G-E--G--	-A-CHGDSSG
1FIW	A	16	254	-----S--	----TNVCAG	Y--P-E----	G-K-I--D--	-T-CQGDSSG
1BIO	A	16	243	-----E--	----RLMCAE	S-----	N-R-R--D--	-S-CKGDSSG
1OP0	A	16	245	-----Y--	----RTL CAG	I---V----	QGG-K--D--	-T-CGGDSSG
1BQY	A	16	245	-----N--	----TTL CAG	I---L----	QGG-R--D--	-T-CHFDSGG
1PQ7	A	16	239	-----N--	----QMFC A-	-GVS-S----	G-G-K--D--	-S-CQGDSSG
1OS8	A	16	245	-----N--	----EEICAG	Y--PDT----	G-G-V--D--	-T-CQGDSSG
2FPZ	A	16	244	-----D--	----DMLCAG	N---T----	--R-R--D--	-S-CQGDSSG
1LTO	A	16	244	-----D--	----DMLCAG	N---S----	--Q-R--D--	-S-CKGDSSG
1C5Y	B	16	245	-----T--	----KMLCA-	-ADP-Q----	W-K-T--D--	-S-CQGDSSG
1TON	A	16	246	-----D--	----VMLCAG	E--M-E----	G-G-K--D--	-T-CAGDSSG
1M9U	A	16	242	-----D--	----NHICVQ	D--P-A----	G-N-T--G--	-A-CNGDSSG
2BZ6	H	16	257	-----E--	----YMFCA-	-GYS-D----	GS--K--D--	-S-CKGDSSG
1RTF	B	16	244	-----D--	----NMLCAG	DT--R----	N-L-H--D--	-A-CQGDSSG
2HLC	A	16	246	-----E--	----STICGD	T---S----	D-G-K--S--	-P-CFGDSSG
1NN6	A	18	241	-----N--	----LQLCVG	N--P-R----	K-T-A--F--	-K-G--DSSG
1GVZ	A	16	246	-----E--	----FVLCAT	H--R-D----	D-S-G--S--	-I-CLGDSSG
1DLE	A	1K250		-----P--	----RFLCTG	--G-VSPY-	A-D-P--N--	-T-CRGDSSG
1PYT	D701	944		-----E--	----TMVCAG	G----D----	G-V-I--S--	-A-CNGDSSG
2O8L	A	1	216	-----G--	----EAMQYD	-----	--L--S--	-T-TGGNSGS
1P3C	A	1	215	-----T--	----NLAYYT	-----	--I-D--T--	-FSG--NSGS
1AGJ	A	1	242	-----R--	----GLRYY	-----	--G--F--	-T-VPGNSGS
1QTF	A	1	246	-----	----SQYF	-----	--G--Y--	-T-EVGNSSGS
1FON	A	9	240	-----K--	----TMVCAG	G----D----	T-R-S--G--	-C-NG-DSSG
2SGA	A	16	242	YGSSGIVY--	----GMIQTN	-----	--V--C--	-A-QPGDSSG
2SFA	A	1	191	YGGGDIVS--	----GLIQTN	-----	--V--C--	-A-EPGDSSG
1SGP	E	16	242	YGGGDVVY--	----GMIRTN	-----	--V--C--	-A-EPGDSSG
1HPG	A	16	242	YGD-GPVY--	----NMVRTT	-----	--A--C--	-S-AGGDSSG
2H5C	A	15A245		YAE-GAVR--	----GLTQGN	-----	--A--C--	-M-GRGDSSG
1Y8T	A	6	314	-----VST	NTVLDAIQTD	A-----	-----A--	-I-NPGNSGG
1SOT	A	43	353	-----	----FLQTD	A-----	-----S--	-I-NMGNSGG
1DY9	A	1	175	-----D--	----SRGSL	-----	----LSPRPV	SY-LKGSSSG
1ARB	A	1	263	-----T--	----THLN VQ	W--Q----	P-S-GG-V--	-T-EFGSSGS
1MBM	A	7	204	-----	----EVCLA	-----	-----W--	-T-TSGDSSGS
2FOM	B	18	167	-----I--	----GAVS	-----	-----L-D	-F-SFGTSSGS
2FP7	B	19	170	-----I--	----GAVT	-----	-----L-D	-Y-PTGTSSGS
1EP5	A120	275		-----	----RFTV	-----	-PK-G--V--	-G-AKGDSSGR
1GCT	A	1	245	PLVC--KK-N	GA--WTLVGI	V--SWG-SS-T	-----C--	--STS--TPG
1K2I	1	1	245	PLVC--KK-N	GA--WTLVGI	V--SWG-SS-T	-----C--	--STS--TPG
1CHO	E	1	245	PLVC--KK-N	GA--WTLVGI	V--SWG-SS-T	-----C--	--STS--TPG

1H4W	A	16	246	PVVC--N---	----GQLQGV	V-SWG-N--G	-----C-A-	--WKN--RPG
1HJ9	A	16	245	PVVC--S---	----GKLQGI	V-SWG-S--G	-----CQA-	--KN--KPG
2A31	A	16	245	PVVC--N---	----GQLQGI	V-SWG-Y--G	-----C-A-	--QKN--KPG
1TRN	A	16	246	PVVC--N---	----GQLQGV	V-SWG-D--G	-----C-A-	--QKN--KPG
1FY8	E	14	245	PVVC--N---	----GELQGI	V-SWG-Y--G	-----C-A-	--LPD--NPG
1A0J	A	16	245	PVVC--N---	----GQLQGV	V-SWG-Y--G	-----C-A-	--QRN--KPG
1HJ8	A	16	245	PVVC--N---	----GELQGV	V-SWG-Y--G	-----C-A-	--EPG--NPG
2F91	A	16	243	PLAA--SD-T	GS--TYLAGI	V-SWG-Y--G	-----C-A-	--RPG--YPG
1NPM	A	16	243	PLVC--D---	----GMLQGI	T-SWG-S--D	-----PC-G-	--KPE--KPG
1EUF	A	16	243	PLVC--N---	----GVAQGI	V-SYG-RN-D	-----G---	--T---TPD
3RP2	A	16	243	PLLC--A---	----GVAHGI	V-SYG-HP-D	-----A---	--K---PPA
1IAU	A	16	244	PLVC--N---	----KVAQGI	V-SYG-RN-N	-----G---	--M---PPR
1GVK	B	16	245	PLHC--LV-N	-GQ-YAVHGV	T-SFV-SRLG	-----C-N-	--VTR--KPT
1ELT	A	16	242	PLNC--QV-N	-GS-YYVHGV	T-SFV-SSSG	-----C-N-	--ASK--KPT
1BRU	P	16	245	PLNC--QG-A	NGQ-WQVHGI	V-SFG-SSLG	-----C-N-	--YYH--KPS
1EKB	B	16	243	PLMC--QE-N	-NR-WLLAGV	T-SFG-Y--Q	-----C-A-	--LPN--RPG
1EAX	A	16	244	PLSS--VE-A	DGR-IFQAGV	V-SWG-D--G	-----C-A-	--QRN--KPG
1H8D	H	16	246	PFVM--KS-P	FNNRWYQMGI	V-SWG-E--G	-----C-D-	--RDG--KYG
1VR1	H	16	243	PFVM--KS-P	FNNRWYQMGI	V-SWG-E--G	-----C-D-	--RDG--KYG
1C5L	H	16	247	PFVM--KS-P	FNNRWYQMGI	V-SWG-E--G	-----C-D-	--RDG--KYG
2PUX	B	16	246	PFVM--KS-P	FNNRWYQMGI	V-SWG-E--G	-----C-D-	--RKG--KYG
1ETR	H	16	247	PFVM--KS-P	YNNRWYQMGI	V-SWG-E--G	-----C-D-	--RDG--KYG
1HNE	E	16	243	PLVC--N---	----GLIHGI	A-SFV-R--G	-----GC-A-	--SGL--YPD
1FUJ	A	16	243	PLIC--D---	----GIIQGI	D-SFV-I--W	-----GC-A-	--TRL--FPD
1A7S	A	1	225	PLVC--E---	----GLAHGV	A-SFS-L--G	-----PC-G-	--R---GPD
1AZZ	A	16	246	PLNY--N---	----GLTYGI	T-SFG-AAAG	-----CEA-	--G---YPD
1L06	A	16	243	PLVC--G---	----DHLRGL	V-SWG-NI-P	-----C---	--GSKE-KPG
2PKA	A	16	246	PLIC--N---	----GMWQGI	T-SWG-H--T	-----PC-G-	--SAN--KPS
1T32	A	16	244	PLLC--N---	----NVAHGI	V-SYG-K--S	-----SG---	--V---PPE
1Q3X	A362	686	ALVF--LD-S	ETERWVFGGI	V-SWG-SM-N	-----C-G-	--EAG--QYG	
1EQ9	A	16	244	PLVA--N---	----GAQIGI	V-SFG-S--P	-----C-A-	--LG--EPD
1FIW	A	16	254	PLMC--KDSA	ENS-YVVVGI	T-SWG-V--G	-----C-A-	--RAK--RPG
1BIO	A	16	243	PLVC--G---	----GVLEGV	V-TSGSR--V	-----C-G-	--NRK--KPG
1OP0	A	16	245	PLIC--N---	----GQFQGI	V-SYG-A--H	-----PC-G-	--QGP--KPG
1BQY	A	16	245	PLIC--N---	----GIFQGI	V-SWG-G--H	-----PC-G-	--QPG--EPG
1PQ7	A	16	239	PIVD--S---	-S--NTLIGA	V-SWG-N--G	-----C-A-	--RPN--YSG
1OS8	A	16	245	PMFR--KDNA	DE--WIQVGI	V-SWG-Y--G	-----C-A-	--RPG--YPG
2FPZ	A	16	244	PLVC--KV-N	-GT-WLQAGV	V-SWG-E--G	-----C-A-	--QPN--RPG
1LTO	A	16	244	PLVC--KV-N	-GT-WLQAGV	V-SWD-E--G	-----C-A-	--QPN--RPG
1C5Y	B	16	245	PLVC--SL-Q	GR--MTLTGI	V-SWG-R--G	-----C-A-	--LKD--KPG
1TON	A	16	246	PLIC--D---	----GVLQGI	T-SGG-AT-P	-----CA-	--KPK--TPA
1M9U	A	16	242	PLNC--PDGG	----TRVVGV	T-SWV-V--S	S-GLGAC-L	--PD--YPS
2BZ6	H	16	257	PHAT--HYRG	T--WYLTGI	V-SWG-Q--G	-----C-A-	--TVG--HFG
1RTF	B	16	244	PLVC--LNDG	R--MTLVGI	I-SWG-L--G	-----C-G-	--QKD--VPG
2HLC	A	16	246	PFVLSDK---	----NLLIGV	V-SFV-SGAG	-----CES-	--G--KPV
1NN6	A	18	241	PLLC--A---	----GVAQGI	V-SYG-RS-D	-----A---	--K---PPA
1GVZ	A	16	246	ALICD-G---	----VFQGI	T-SWG-YS-E	-----C-A-	--DFN--DNF
1DLE	A	1K250	PLIVH-KRSR	----FIQVGV	I-SWG-VV-D	-----VCKQ	V-PAH--ARD	
1PYT	D701	944	PLNC--QADG	Q--WDVIRGI	V-SFG-SGLS	-----C-N-	--TFK--KPT	
2O8L	A	1	216	PVFN--EK--	----NEVIGI	H-WGG-V--P	-----	--N---EFN
1P3C	A	1	215	AMLD--QN--	----QQIVGV	H-NAG-YS-N	-----	--G---TIN
1AGJ	A	1	242	GIFN--SN--	----GELVGI	H-SSK-V--S	H-LDRE	--H---QIN
1QTF	A	1	246	GIFN--LK--	----GELIGI	H-SGK-G--G	Q-----	--H---NLP
1FON	A	9	240	PLNC--PAAD	GS--WQVHGV	T-SFV-SAFG	-----C-N-	--TIK--KPT
2SGA	A	16	242	SLFA--G--	----STALGL	T-SGG-S--G	-----NCRT	--G---GTT
2SFA	A	1	191	ALFA--G--	----STALGL	T-SGG-S--G	-----NCRT	--G---GTT
1SGP	E	16	242	PLYS--G--	----TRAIGL	T-SGG-S--G	-----NCSS	--G---GTT
1HPG	A	16	242	AHFA--G--	----SVALGI	H-SGS-S--G	-----C-S	GTA--GSA
2H5C	A	15A245	SWIT--SA--	----GQAQGV	M-SGG-N--V	Q-SNGMNC-G	IPASQ--RSS	
1Y8T	A	6	314	ALVN--N---	----AQLVGV	N-SAI-A--T	L-Q--SG-S-	--IG---LG
1SOT	A	43	353	ALVNS-L---	----GEL-GI	N-TLS-F--D	-----	--E---GIG

1DY9	A	1	175	PLLC--PS--	---	GHVVGI	F	RAA-V--	C	T-RG-----	--VA---KAV
1ARB	A	1	263	PIYSP-E---	---	KRVLGQ	L	HGG-P--	S	-----SCS--	--ATGTMRS
1MBM	A	7	204	AVVQ---G--	---	DAVVGV	H	TGS-N--	T	-----S---	GVA
2FOM	B	18	167	PIVD--KK--	---	GKVVGL	Y	GNG-V--	V	TRSG-----	--AY---VS
2FP7	B	19	170	PIVD--KN--	---	GDVIGL	Y	GNG-V--	I	MPNG-----	--SY---IS
1EP5	A	120	275	PILD--NQ--	---	GRVVAI	VLGGV-N-	E	-----	--GS---RTA	
1GCT	A	1	245	VYA-RV-T--	-----	---	A-LVN--	---	WVQQ-	T-L--A--AN	
1K2I	1	1	245	VYA-RV-T--	-----	---	A-LVN--	---	WVQQ-	T-L--A--AN	
1CHO	E	1	245	VYA-RV-T--	-----	---	A-LVN--	---	WVQQ-	T-L--A--AN	
1H4W	A	16	246	VYT-KV-Y--	-----	---	N-YVD--	---	WIKD-	T-I--A--AN	
1HJ9	A	16	245	VYT-KV-C--	-----	---	N-YVS--	---	WIKQ-	T-I--A--SN	
2A31	A	16	245	VYT-KV-C--	-----	---	N-YVN--	---	WIQQ-	T-I--A--AN	
1TRN	A	16	246	VYT-KV-Y--	-----	---	N-YVK--	---	WIKN-	T-I--A--AN	
1FY8	E	14	245	VYT-KV-C--	-----	---	N-YVD--	---	WIQD-	T-I--A--AN	
1A0J	A	16	245	VYT-KV-C--	-----	---	N-YRS--	---	WISS-	T-M--S--SN	
1HJ8	A	16	245	VYA-KV-C--	-----	---	I-FND--	---	WLTS-	T-M--A--SY	
2F91	A	16	243	VYT-EV-S--	-----	---	Y-HVD--	---	WIKA-	N-A-----	
1NPM	A	16	243	VYT-KI-C--	-----	---	R-YTT--	---	WIKK-	T-M--D----	
1EUF	A	16	243	VYT-RI-S--	-----	---	S-FLS--	---	WIHS-	T-M--R----	
3RP2	A	16	243	IFT-RV-S--	-----	---	T-YVP--	---	WINA-	V-I--N----	
1IAU	A	16	244	ACT-KV-S--	-----	---	S-FVH--	---	WIKK-	T-M--K--R-	
1GVK	B	16	245	VFT-RV-S--	-----	---	A-YIS--	---	WINN-	V-I--A--SN	
1ELT	A	16	242	VFT-RV-S--	-----	---	A-YIS--	---	WMNG-	I-M-----	
1BRU	P	16	245	VFT-RV-S--	-----	---	N-YID--	---	WINS-	V-I--A--NN	
1EKB	B	16	243	VYA-RV-P--	-----	---	R-FTE--	---	WIQS-	--F--L--H-	
1EAX	A	16	244	VYT-RL-P--	-----	---	L-FRD--	---	WIKI-	N-T--G--V-	
1H8D	H	16	246	FYT-HV-F--	-----	---	R-LKK--	---	WIQK-	V-I--D--QF	
1VR1	H	16	243	FYT-HV-F--	-----	---	R-LKK--	---	WIQK-	V-I--D----	
1C5L	H	16	247	FYT-HV-F--	-----	---	R-LKK--	---	WIQK-	V-I--D--QF	
2PUX	B	16	246	FYT-HV-F--	-----	---	R-LKR--	---	WIQK-	V-I--D--QF	
1ETR	H	16	247	FYT-HV-F--	-----	---	R-LKK--	---	WIQK-	V-I--D--RL	
1HNE	E	16	243	AFA-PV-A--	-----	---	Q-FVN--	---	WIDS-	I-I--Q----	
1FUJ	A	16	243	FFT-RV-A--	-----	---	L-YVD--	---	WIRS-	T-L--R----	
1A7S	A	1	225	FFT-RV-A--	-----	---	L-FRD--	---	WIDG-	V-L--N--NP	
1AZZ	A	16	246	AFT-RV-T--	-----	---	Y-FLD--	---	WIQT-	Q-T--G--IT	
1LO6	A	16	243	VYT-NV-C--	-----	---	R-YTN--	---	WIQK-	T-I--Q----	
2PKA	A	16	246	IYT-KL-I--	-----	---	F-YLD--	---	WIDD-	T-I--T--EN	
1T32	A	16	244	VFT-RV-S--	-----	---	S-FLP--	---	WIRT-	T-M--R--S-	
1Q3X	A	362	686	VYT-KV-I--	-----	---	N-YIP--	---	WIEN-	I-I--S--DF	
1EQ9	A	16	244	VYT-RV-S--	-----	---	S-FVS--	---	WINA-	N-L-----K	
1FIW	A	16	254	VYT-ST-W--	-----	---	S-YLN--	---	WIAS-	K-I--G-STAV	
1BIO	A	16	243	IYT-RV-A--	-----	---	S-YAA--	---	WIDS-	V-L--A-----	
1OP0	A	16	245	IYT-NV-F--	-----	---	D-YTD--	---	WIQR-	N-I--AGNTD	
1BQY	A	16	245	VYT-KV-F--	-----	---	D-YLD--	---	WIKS-	I-I--AGNKD	
1PQ7	A	16	239	VYA-SV-G--	-----	---	A-LRS--	---	FIDT-	Y-A-----	
1OS8	A	16	245	VYT-EV-S--	-----	---	T-FAS--	---	AIAS-	A-A--R--TL	
2FPZ	A	16	244	IYT-RV-T--	-----	---	Y-YLD--	---	WIHH-	Y-V--PK---	
1LTO	A	16	244	IYT-RV-T--	-----	---	Y-YLD--	---	WIHH-	Y-V--PK---	
1C5Y	B	16	245	VYT-RV-S--	-----	---	H-FLP--	---	WIRS-	H-T--KEE--	
1TON	A	16	246	IYA-KL-I--	-----	---	K-FTS--	---	WIKK-	V-M--KENP-	
1M9U	A	16	242	VYT-RV-S--	-----	---	A-YLG--	---	WIGD-	N-S-----	
2BZ6	H	16	257	VYT-RV-S--	-----	---	Q-YIE--	---	WLQK-	L-M--R--SE	
1RTF	B	16	244	VYT-KV-T--	-----	---	N-YLD--	---	WIRD-	N-M-----R	
2HLC	A	16	246	GFS-RV-T--	-----	---	S-YMD--	---	WIQQ-	N-TG-I--KF	
1NN6	A	18	241	VFT-RI-S--	-----	---	H-YRP--	---	WINQ-	I-L--Q--AN	
1GVZ	A	16	246	VFT-KV-M--	-----	---	P-HKK--	---	WIKE-	T-I--E--KN	
1DLE	A	1K	250	FHI-NL-F--	-----	---	Q-VLP--	---	WLKE-	K-L--Q-DEDL	
1PYT	D	701	944	VFT-RV-S--	-----	---	A-YID--	---	WINQ-	K-L--Q---L	
2O8L	A	1	216	GAV-FI-N--	-----	---	ENVRN--	---	FLKQ-	N-I--E--DI	
1P3C	A	1	215	GGP-KATA--	-----	---	A-FVE--	---	FINY-	A-K--A---Q	

1AGJ	A	1	242	YGV-GI-G--	-----	---NYVKR--	----IINE-	K-N--E----
1QTF	A	1	246	IGV-FF-NRK	ISSLYSDNT	FGDTLGND--	----LKKR-	AKL--D--K-
1FON	A	9	240	VFT-RV-S--	-----	---A-FID--	----WIDE-	T-I--A--SN
2SGA	A	16	242	FYQ-PV----	-----	---TE----	----ALSA-	Y-G----AT
2SFA	A	1	191	FFQ-PV----	-----	---TE----	----ALSA-	Y-G----VS
1SGP	E	16	242	FFQ-PV----	-----	---TE----	----ALVA-	Y-G----VS
1HPG	A	16	242	IHQ-PV----	-----	---TE----	----ALSA-	Y-G----VT
2H5C	A	15A	245	LFE-RL----	-----	---QP----	----ILSQ-	Y-G----LS
1Y8T	A	6	314	FAI-PV-D--	-----	---Q-AKR--	----IADEL	I-S--TGKAS
1SOT	A	43	353	FAI-PF-Q--	-----	---L-AT---	----KIDK-	L-I--R--DG
1DY9	A	1	175	DFI-PV-E--	-----	---SM-----	-----	-----
1ARB	A	1	263	QYG-RV-F--	-----	---T-SWTGG	GAAASRLSD-	--WLDPASTG
1MBM	A	7	204	YVT-TP----	-----	---SGKLLGA	DTVTLSSLK	HFTGPLTSIP
2FOM	B	18	167	AIA-N-----	-----	-----	-----	-----
2FP7	B	19	170	AIV-QG-E--	-----	---R-----	-----	-----
1EP5	A	120	275	LSVVMW-N--	-----	---E-KGV--	----TVKY-	T-P--EN•EQ
1GCT	A	1	245	-----	-----	-----	-----	-----
1K2I	I	1	245	-----	-----	-----	-----	-----
1CHO	E	1	245	-----	-----	-----	-----	-----
1H4W	A	16	246	S-----	-----	-----	-----	-----
1HJ9	A	16	245	-----	-----	-----	-----	-----
2A31	A	16	245	-----	-----	-----	-----	-----
1TRN	A	16	246	S-----	-----	-----	-----	-----
1FY8	E	14	245	-----	-----	-----	-----	-----
1A0J	A	16	245	-----	-----	-----	-----	-----
1HJ8	A	16	245	-----	-----	-----	-----	-----
2F91	A	16	243	-----	-----	-----	-----	-----
1NPM	A	16	243	-----	-----	-----	-----	-----
1EUF	A	16	243	-----	-----	-----	-----	-----
3RP2	A	16	243	-----	-----	-----	-----	-----
1IAU	A	16	244	-----	-----	-----	-----	-----
1GVK	B	16	245	-----	-----	-----	-----	-----
1ELT	A	16	242	-----	-----	-----	-----	-----
1BRU	P	16	245	-----	-----	-----	-----	-----
1EKB	B	16	243	-----	-----	-----	-----	-----
1EAX	A	16	244	-----	-----	-----	-----	-----
1H8D	H	16	246	G-----	-----	-----	-----	-----
1VR1	H	16	243	-----	-----	-----	-----	-----
1C5L	H	16	247	GE-----	-----	-----	-----	-----
2PUX	B	16	246	G-----	-----	-----	-----	-----
1ETR	H	16	247	GS-----	-----	-----	-----	-----
1HNE	E	16	243	-----	-----	-----	-----	-----
1FUJ	A	16	243	-----	-----	-----	-----	-----
1A7S	A	1	225	GPGPA-----	-----	-----	-----	-----
1AZZ	A	16	246	P-----	-----	-----	-----	-----
1LO6	A	16	243	-----	-----	-----	-----	-----
2PKA	A	16	246	P-----	-----	-----	-----	-----
1T32	A	16	244	-----	-----	-----	-----	-----
1Q3X	A	362	686	-----	-----	-----	-----	-----
1EQ9	A	16	244	K-----	-----	-----	-----	-----
1FIW	A	16	254	HMIQLPT---	-----	-----	-----	-----
1BIO	A	16	243	-----	-----	-----	-----	-----
1OP0	A	16	245	ATCPP-----	-----	-----	-----	-----
1BQY	A	16	245	ATCPP-----	-----	-----	-----	-----
1PQ7	A	16	239	-----	-----	-----	-----	-----
1OS8	A	16	245	-----	-----	-----	-----	-----
2FPZ	A	16	244	-----	-----	-----	-----	-----
1LTO	A	16	244	-----	-----	-----	-----	-----
1C5Y	B	16	245	-----	-----	-----	-----	-----
1TON	A	16	246	-----	-----	-----	-----	-----

1M9U	A	16	242	-----	-----	-----	-----	-----
2BZ6	H	16	257	PRPGVLLRAP	FP-----	-----	-----	-----
1RTF	B	16	244	P-----	-----	-----	-----	-----
2HLC	A	16	246	-----	-----	-----	-----	-----
1NN6	A	18	241	-----	-----	-----	-----	-----
1GVZ	A	16	246	S-----	-----	-----	-----	-----
1DLE	A	1K250		GFL-----	-----	-----	-----	-----
1PYT	D701		944	-----	-----	-----	-----	-----
2O8L	A	1	216	NFA-----	-----	-----	-----	-----
1P3C	A	1	215	-----	-----	-----	-----	-----
1AGJ	A	1	242	-----	-----	-----	-----	-----
1QTF	A	1	246	-----	-----	-----	-----	-----
1FON	A	9	240	-----	-----	-----	-----	-----
2SGA	A	16	242	VL-----	-----	-----	-----	-----
2SFA	A	1	191	IL-----	-----	-----	-----	-----
1SGP	E	16	242	VY-----	-----	-----	-----	-----
1HPG	A	16	242	VY-----	-----	-----	-----	-----
2H5C	A	15A245		LVTG-----	-----	-----	-----	-----
1Y8T	A	6	314	HASLGVQVTN	DKDTILGAKIV	EVVAGGAAAN	AGVPKGVVVT	KVDDRPINSA
1SOT	A	43	353	RVIRGYIGIG	GRGIVVNEVS	PDGPAANAGI	QVNDLIISVD	NKPAISALET
1DY9	A	1	175	-----	-----	-----	-----	-----
1ARB	A	1	263	AQFIDGLDS-	-----	-----	-----	-----
1MBM	A	7	204	KDIPDNIIAD	VDAVPRSLAM	LIDGLSNRE-	-----	-----
2FOM	B	18	167	-----	-----	-----	-----	-----
2FP7	B	19	170	-----	-----	-----	-----	-----
1EP5	A120		275	W-----	-----	-----	-----	-----
1GCT	A	1	245	-----	-----	-----	-----	-----
1K2I	I	1	245	-----	-----	-----	-----	-----
1CHO	E	1	245	-----	-----	-----	-----	-----
1H4W	A	16	246	-----	-----	-----	-----	-----
1HJ9	A	16	245	-----	-----	-----	-----	-----
2A31	A	16	245	-----	-----	-----	-----	-----
1TRN	A	16	246	-----	-----	-----	-----	-----
1FY8	E	14	245	-----	-----	-----	-----	-----
1A0J	A	16	245	-----	-----	-----	-----	-----
1HJ8	A	16	245	-----	-----	-----	-----	-----
2F91	A	16	243	-----	-----	-----	-----	-----
1NPM	A	16	243	-----	-----	-----	-----	-----
1EUF	A	16	243	-----	-----	-----	-----	-----
3RP2	A	16	243	-----	-----	-----	-----	-----
1IAU	A	16	244	-----	-----	-----	-----	-----
1GVK	B	16	245	-----	-----	-----	-----	-----
1ELT	A	16	242	-----	-----	-----	-----	-----
1BRU	P	16	245	-----	-----	-----	-----	-----
1EKB	B	16	243	-----	-----	-----	-----	-----
1EAX	A	16	244	-----	-----	-----	-----	-----
1H8D	H	16	246	-----	-----	-----	-----	-----
1VR1	H	16	243	-----	-----	-----	-----	-----
1C5L	H	16	247	-----	-----	-----	-----	-----
2PUX	B	16	246	-----	-----	-----	-----	-----
1ETR	H	16	247	-----	-----	-----	-----	-----
1HNE	E	16	243	-----	-----	-----	-----	-----
1FUJ	A	16	243	-----	-----	-----	-----	-----
1A7S	A	1	225	-----	-----	-----	-----	-----
1AZZ	A	16	246	-----	-----	-----	-----	-----
1LO6	A	16	243	-----	-----	-----	-----	-----
2PKA	A	16	246	-----	-----	-----	-----	-----
1T32	A	16	244	-----	-----	-----	-----	-----
1Q3X	A362		686	-----	-----	-----	-----	-----
1EQ9	A	16	244	-----	-----	-----	-----	-----

1FIW	A	16	254	-----	-----	-----	-----
1BIO	A	16	243	-----	-----	-----	-----
1OP0	A	16	245	-----	-----	-----	-----
1BQY	A	16	245	-----	-----	-----	-----
1PQ7	A	16	239	-----	-----	-----	-----
1OS8	A	16	245	-----	-----	-----	-----
2FPZ	A	16	244	-----	-----	-----	-----
1LTO	A	16	244	-----	-----	-----	-----
1C5Y	B	16	245	-----	-----	-----	-----
1TON	A	16	246	-----	-----	-----	-----
1M9U	A	16	242	-----	-----	-----	-----
2BZ6	H	16	257	-----	-----	-----	-----
1RTF	B	16	244	-----	-----	-----	-----
2HLC	A	16	246	-----	-----	-----	-----
1NN6	A	18	241	-----	-----	-----	-----
1GVZ	A	16	246	-----	-----	-----	-----
1DLE	A	1K	250	-----	-----	-----	-----
1PYT	D701		944	-----	-----	-----	-----
2O8L	A	1	216	-----	-----	-----	-----
1P3C	A	1	215	-----	-----	-----	-----
1AGJ	A	1	242	-----	-----	-----	-----
1QTF	A	1	246	-----	-----	-----	-----
1FON	A	9	240	-----	-----	-----	-----
2SGA	A	16	242	-----	-----	-----	-----
2SFA	A	1	191	-----	-----	-----	-----
1SGP	E	16	242	-----	-----	-----	-----
1HPG	A	16	242	-----	-----	-----	-----
2H5C	A	15A	245	-----	-----	-----	-----
1Y8T	A	6	314	DALVAAVRSK	APGATVALTF	QDPSSGSRIV	QVTLGKA
1SOT	A	43	353	DQVAEIRPGS	VIPVVVQLTL	QVTIQEYPA-	-----
1DY9	A	1	175	-----	-----	-----	-----
1ARB	A	1	263	-----	-----	-----	-----
1MBM	A	7	204	-----	-----	-----	-----
2FOM	B	18	167	-----	-----	-----	-----
2FP7	B	19	170	-----	-----	-----	-----
1EP5	A	120	275	-----	-----	-----	-----

IFRs das quatro subfamílias de serino proteases: tripsinas (verde), quimotripsinas (vermelho), elastases (azul) e trombinas (amarelo); com os inibidores ecotina, BPTI e ovomucoide.

### Ecotina – 1AZZ

	37	37C	39	40	41	42	57	58	59	59B	59D	59E	59G	60	94	96	97	97A	98	99	143	143D	143H	143Q	
1PQ7	R	-	-	P	W	C	H	-	-	-	-	-	-	Y	Y	-	-	-	G	N	-	E	G	T	
1TRN	-	-	-	H	F	C	H	C	Y	-	-	-	-	K	Y	R	K	-	T	L	-	-	-	-	
1HJ9	-	-	Y	H	F	C	H	C	-	-	-	-	-	K	Y	S	N	-	T	L	-	-	-	Y	
1HJ8	-	-	Y	H	F	C	H	C	-	-	-	-	-	K	Y	S	Y	-	N	I	N	-	-	-	
2F91	E	-	F	H	F	C	H	C	-	-	-	-	-	-	-	Y	N	-	L	L	T	E	G	T	
1H4W	-	-	-	H	F	C	H	C	Y	-	-	-	-	K	Y	R	D	-	T	L	-	-	-	Y	
2A31	-	-	-	H	F	C	H	C	-	-	-	-	-	K	F	G	N	-	T	L	-	-	-	Y	
1A0J	-	-	Y	H	F	-	H	-	-	-	-	-	-	-	Y	S	R	-	N	L	-	-	-	Y	
1OS8	M	-	-	-	G	C	H	C	-	-	-	-	-	-	Y	G	-	-	T	-	-	E	G	Q	
1K21	D	-	F	H	F	C	H	C	G	-	-	-	-	-	Y	S	L	-	T	I	L	-	-	T	
1GCT	-	-	-	H	F	C	-	C	G	-	-	-	-	-	-	-	L	-	T	I	-	-	-	T	
1EQ9	-	-	-	H	R	C	H	C	-	-	-	-	-	-	Y	D	F	-	L	L	S	-	G	T	
1PYT	Y	-	R	-	T	C	H	C	-	-	-	-	-	S	W	S	F	-	L	V	R	N	-	-	
1ELT	Y	-	Y	H	T	C	H	C	-	-	-	-	-	W	S	D	-	D	V	K	T	G	L		
1M9U	R	-	-	H	S	C	H	C	-	-	-	-	-	Y	-	-	-	S	Y	R	-	-	L		
1HNE	-	-	-	H	F	C	H	C	-	N	-	-	-	V	Y	P	V	-	N	L	L	-	-	I	
1GVK	Y	-	-	H	T	C	H	-	-	-	-	-	-	D	W	T	D	-	D	V	L	T	N	L	
1BRU	Y	-	R	H	T	C	H	C	-	-	-	-	-	W	S	N	-	Q	L	R	T	N	S		
1C5L	R	P	E	L	L	C	H	-	-	Y	-	-	W	K	-	-	W	R	E	N	L	N	E	T	Q
1VR1	-	-	R	F	L	C	H	-	-	Y	P	W	K	-	-	W	R	E	N	L	N	E	T	Q	
2PUX	R	P	E	L	L	C	H	C	-	Y	P	W	K	-	-	W	R	E	N	L	N	E	T	Q	
1H8D	R	P	E	L	L	C	H	C	-	Y	P	W	K	-	-	W	R	E	N	L	N	E	T	-	
1ETR	R	-	E	L	L	C	H	-	-	Y	-	W	K	-	-	W	K	E	N	L	N	E	T	Q	
1FY8	-	-	Y	H	F	C	H	C	Y	-	-	-	-	K	F	R	K	-	T	L	-	-	-	-	
1CHO	-	-	-	H	F	C	H	C	-	-	-	-	-	-	Y	S	L	-	T	I	L	-	-	T	
1AZZ	-	-	M	Y	F	C	H	C	-	-	-	-	-	-	Y	S	F	-	V	I	L	-	-	I	

	171	172	173	174	175	177	189	190	191	192	193	195	213	214	215	216	217	217A	219	220	221A	224	226
Q	Y	G	S	A	-	D	S	C	Q	G	S	V	S	W	G	N	-	G	C	R	Y	-	-
-	Y	P	G	K	-	D	S	C	Q	G	S	V	S	W	G	D	-	G	C	-	K	-	-
-	Y	P	G	Q	-	D	S	C	Q	G	S	V	S	W	G	S	-	G	C	-	K	G	-
S	Y	P	G	M	-	D	S	C	Q	G	S	V	S	W	G	Y	-	G	C	-	N	G	-
D	Y	G	D	E	-	D	S	C	Q	G	S	V	S	W	G	Y	-	G	C	-	Y	G	-
S	Y	P	G	K	-	D	S	C	Q	R	S	V	S	W	G	H	-	G	C	-	R	G	-
-	Y	P	G	Q	-	D	S	C	Q	G	S	V	S	W	G	Y	-	G	C	-	K	G	-
-	Y	P	G	Q	-	D	S	C	Q	G	S	V	S	W	G	Y	-	G	C	-	K	G	-
A	Y	G	N	E	-	T	C	Q	G	S	V	S	W	G	Y	-	G	C	-	Y	-	-	-
Y	W	G	T	K	K	-	S	C	M	G	S	V	S	W	G	S	S	-	C	-	T	-	-
-	W	G	T	K	-	-	S	C	M	G	S	-	S	W	G	S	S	-	T	-	S	-	-
D	Q	W	-	R	-	A	C	H	G	S	V	S	F	G	S	-	P	C	-	E	D	-	-
-	W	G	T	K	-	-	C	-	-	S	V	S	F	G	S	-	G	-	C	-	K	T	-
-	W	G	S	T	K	-	-	C	N	G	S	T	S	F	V	S	S	-	-	-	K	-	-
-	M	G	A	N	-	-	-	C	N	G	S	T	S	W	V	V	-	S	-	-	Y	-	-
-	-	-	-	-	-	-	V	C	F	G	S	A	S	F	V	R	-	G	-	-	Y	-	-
-	W	G	S	T	K	-	-	C	Q	G	S	T	S	F	V	S	R	-	-	-	K	-	-
-	W	G	S	T	K	-	S	C	N	G	S	V	S	F	G	S	S	-	C	-	K	-	-
S	T	R	I	R	-	D	A	C	E	G	S	V	S	W	G	E	-	G	C	-	K	-	-
S	T	R	I	R	-	D	A	C	E	G	S	V	S	W	G	E	-	G	C	R	K	-	-
S	T	R	I	R	-	D	A	C	E	G	A	V	S	W	G	E	-	G	C	R	K	-	-
-	T	R	I	R	-	D	A	C	E	G	S	V	S	W	G	E	-	G	C	-	K	-	-
S	T	R	I	R	-	D	A	C	E	G	S	V	S	W	G	E	-	G	C	R	K	-	-
S	-	P	G	K	-	D	S	C	Q	G	S	V	S	W	G	Y	-	G	C	-	N	G	-
Y	W	G	T	K	K	-	S	C	M	G	S	V	S	W	G	S	S	-	C	-	T	-	-
Y	Y	-	G	I	-	-	T	C	N	G	S	T	S	F	G	A	A	-	C	-	Y	D	-

### BPTI – 1FY8

	37	37C	39	40	41	42	57	58	59B	59D	59E	59F	59G	59H	60	73	94	96	97	97A	99	141	142
1PQ7	R	-	G	P	W	C	H	-	-	-	-	-	-	-	Y	-	Y	-	-	-	N	-	-
1TRN	-	-	Y	H	F	C	H	C	-	-	-	-	-	-	K	-	Y	R	K	-	L	W	G
1HJ9	-	-	Y	H	F	C	H	C	-	-	-	-	-	-	K	I	Y	S	N	-	L	-	-
1HJ8	-	-	Y	H	F	C	H	C	-	-	-	-	-	-	K	I	Y	S	Y	-	I	W	G
2F91	E	-	F	H	F	C	H	C	-	-	-	-	-	-	-	-	-	Y	N	-	L	W	G
1H4W	-	-	S	H	F	C	H	C	-	-	-	-	-	-	K	I	Y	R	D	-	L	W	-
2A31	-	-	S	H	F	C	H	C	-	-	-	-	-	-	K	I	-	G	N	-	L	W	-
1A0J	-	-	Y	H	F	-	H	-	-	-	-	-	-	-	K	I	Y	S	R	-	L	W	-
1OS8	M	-	-	G	C	H	C	-	-	-	-	-	-	-	-	-	Y	G	-	-	-	-	-
1K2I	-	-	F	H	F	C	H	C	-	-	-	-	-	-	-	Q	Y	S	L	-	I	-	-
1GCT	-	0	-	H	F	C	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1EQ9	-	-	S	H	R	C	H	C	G	-	-	-	-	-	-	L	-	D	F	-	L	-	-
1PYT	Y	-	R	H	T	C	H	C	-	-	-	-	-	-	S	L	W	S	F	-	V	-	-
1ELT	Y	-	Y	H	T	C	H	C	-	-	-	-	-	-	D	L	W	S	D	-	V	-	-
1M9U	R	-	S	H	S	C	H	C	G	-	-	-	-	-	V	Q	Y	-	-	-	Y	-	-
1HNE	L	-	G	H	F	C	H	C	N	-	-	-	-	-	V	L	Y	P	V	-	L	-	-
1GVK	Y	-	A	H	T	C	H	-	-	-	-	-	-	-	D	L	W	T	D	-	V	-	-
1BRU	Y	-	R	H	T	C	H	C	-	-	-	-	-	-	S	L	W	S	N	-	L	-	-
1C5L	R	P	E	L	L	C	H	-	Y	P	W	-	K	N	F	R	-	W	R	E	L	-	G
1VR1	-	-	R	F	L	C	H	-	Y	P	W	D	K	N	F	R	-	W	R	E	L	-	G
2PUX	R	P	E	L	L	C	H	C	Y	P	W	D	K	-	F	R	-	W	R	E	L	W	G
1H8D	R	P	E	L	L	C	H	C	Y	P	W	D	K	N	F	R	-	W	R	E	L	-	G
1ETR	R	-	E	L	L	C	H	-	Y	-	W	D	K	-	F	R	-	W	-	E	L	-	G
1FY8	-	-	Y	H	F	C	H	C	-	-	-	-	-	-	K	I	F	R	K	-	L	-	-
1CHO	-	-	F	H	F	C	H	C	-	-	-	-	-	-	-	Q	Y	S	L	-	I	W	-
1AZZ	-	-	M	Y	F	C	H	C	-	-	-	-	-	-	-	-	Y	S	F	-	I	-	-

	143	143Q	172	174	189	190	191	192	193	195	213	214	215	216	217	217A	219	220	226	227	228	
-	T	-	-	D	S	C	Q	G	S	V	S	W	G	N	-	G	C	G	V	-	-	-
-	-	-	-	D	S	C	Q	G	S	V	S	W	G	D	-	G	C	G	V	-	-	-
-	Y	-	-	D	S	C	Q	G	S	V	S	W	G	S	-	G	C	G	V	-	-	-
N	-	-	-	D	S	C	Q	G	S	V	S	W	G	Y	-	G	C	G	V	-	-	-
T	T	-	-	D	S	C	Q	G	S	V	S	W	G	Y	-	G	C	G	V	-	-	-
-	Y	-	-	D	S	C	Q	R	S	V	S	W	G	H	-	G	C	G	V	-	-	-
-	Y	-	-	D	S	C	Q	G	S	V	S	W	G	Y	-	G	C	G	V	-	-	-
-	Y	-	-	D	S	C	Q	G	S	V	S	W	G	Y	-	G	C	G	V	Y	-	-
-	Q	-	-	D	T	C	Q	G	S	V	S	W	G	Y	-	G	C	G	V	-	-	-
L	T	W	-	S	S	C	M	G	S	V	S	W	G	S	S	-	C	G	-	Y	-	-
-	T	-	-	S	C	M	G	S	-	S	W	G	S	S	T	-	C	D	-	V	Y	-
S	T	-	-	A	C	H	G	S	V	S	F	G	-	-	P	C	D	-	-	-	-	-
R	-	W	-	-	C	-	-	S	V	S	F	G	S	G	-	C	T	-	-	-	-	-
K	L	W	-	-	C	N	G	S	T	S	F	V	S	S	-	-	-	-	-	-	-	-
R	L	-	-	-	A	C	N	G	S	T	S	W	V	V	-	S	-	S	-	-	-	-
L	I	-	-	-	V	C	F	G	S	A	S	F	V	R	-	-	-	-	-	-	-	-
L	L	W	-	-	C	Q	G	S	T	S	F	V	S	R	-	-	-	-	-	-	-	-
R	S	W	-	-	S	C	N	G	S	V	S	F	G	S	S	-	C	S	-	F	-	-
N	Q	-	I	D	A	C	E	G	S	V	S	W	G	E	-	G	C	G	-	Y	-	-
N	Q	-	I	D	A	C	E	G	S	V	S	W	G	E	-	G	C	G	-	-	-	-
N	Q	-	I	D	A	C	E	G	A	V	S	W	G	E	-	G	C	G	F	Y	-	-
N	Q	-	I	D	A	C	E	G	S	V	S	W	G	E	-	G	C	G	-	-	-	-
N	Q	-	I	D	A	C	E	G	S	V	S	W	G	E	-	G	C	G	F	Y	-	-
-	-	-	-	D	S	C	Q	G	S	V	S	W	G	Y	-	G	C	G	V	-	-	-
L	T	W	-	S	S	C	M	G	S	V	S	W	G	S	S	-	C	G	-	Y	-	-
L	I	-	-	-	T	C	N	G	S	T	S	F	G	A	A	-	C	D	-	-	-	-



### Ovomucoide – 1CHO

	37	37C	38	39	40	41	42	57	58	59	59B	59D	59E	59F	59G	60	73	74	94	97	99	142
1PQ7	R	-	N	G	P	W	C	H	-	-	-	-	-	-	-	Y	R	-	-	-	N	-
1TRN	-	-	-	Y	H	F	C	H	C	Y	-	-	-	-	-	K	-	E	-	K	L	G
1HJ9	-	-	-	Y	H	F	C	H	C	Y	-	-	-	-	-	K	-	N	-	N	L	-
1HJ8	-	-	-	Y	H	F	C	H	C	Y	-	-	-	-	-	K	-	K	Y	Y	I	G
2F91	E	-	-	F	H	F	C	H	C	V	-	-	-	-	-	D	-	-	-	N	L	G
1H4W	-	-	-	S	H	F	C	H	C	Y	-	-	-	-	-	K	-	K	-	D	L	-
2A31	S	-	-	S	H	F	C	H	C	-	-	-	-	-	-	K	-	-	-	N	L	-
1A0J	S	-	-	Y	H	F	-	H	-	Y	-	-	-	-	-	K	-	-	-	R	L	-
1OS8	M	-	-	-	G	C	H	C	V	-	-	-	-	-	-	-	-	-	Y	-	-	-
1K2I	D	-	-	F	H	F	C	H	C	G	-	-	-	-	-	-	-	-	-	-	I	-
1GCT	-	-	-	F	H	F	C	-	C	-	-	-	-	-	-	V	-	-	-	-	I	-
1EQ9	L	-	-	S	H	R	C	H	C	-	-	-	-	-	-	L	-	-	Y	F	L	-
1PYT	Y	-	-	R	H	T	C	H	C	-	-	-	-	-	-	S	-	-	W	F	V	-
1ELT	Y	-	-	Y	H	T	C	H	C	-	-	-	-	-	-	-	-	-	W	D	V	-
1M9U	R	-	-	S	H	S	C	H	C	-	-	-	-	-	-	V	-	-	-	-	Y	-
1HNE	L	-	R	G	H	F	C	H	C	-	N	-	-	-	-	V	-	-	Y	-	L	-
1GVK	Y	-	-	A	H	T	C	H	-	-	-	-	-	-	-	-	-	-	W	D	V	-
1BRU	Y	-	-	R	H	T	C	H	C	-	-	-	-	-	-	S	-	-	W	N	L	-
1C5L	R	P	Q	E	L	L	C	H	-	-	Y	-	W	D	K	F	R	T	-	-	L	-
1VR1	A	-	E	R	F	L	C	H	-	-	Y	P	W	D	K	F	R	T	-	-	L	-
2PUX	R	P	Q	E	L	L	C	H	C	-	Y	P	W	D	K	F	R	T	-	-	L	-
1H8D	R	P	Q	E	L	L	C	H	C	-	Y	P	W	D	K	F	R	T	-	-	L	G
1ETR	R	-	-	E	L	L	C	H	-	-	Y	-	W	D	K	F	R	-	-	-	L	G
1FY8	-	-	-	Y	H	F	C	H	C	Y	-	-	-	-	-	K	-	-	-	K	L	-
1CHO	D	-	-	F	H	F	C	H	C	G	-	-	-	-	-	-	-	-	-	-	I	-
1AZZ	I	-	D	M	Y	F	C	H	C	-	G	-	-	-	-	-	-	-	R	Y	F	I

	143	143D	143H	143Q	172	173	174	175	189	190	191	192	193	195	213	214	215	216	217	217A
A	E	G	T	Y	-	-	-	-	D	S	C	Q	G	S	V	S	W	G	N	-
N	-	-	-	Y	-	-	-	K	D	S	C	Q	G	S	V	S	W	G	D	-
N	-	-	Y	Y	-	-	-	Q	D	S	C	Q	G	S	V	S	W	G	S	-
N	-	-	-	-	-	-	-	M	D	S	C	Q	G	S	V	S	W	G	Y	-
T	E	G	T	Y	-	-	-	E	D	S	C	Q	G	S	V	S	W	G	Y	-
N	-	G	Y	-	-	-	-	K	D	S	C	Q	R	S	V	S	W	G	H	-
N	-	-	Y	Y	-	-	-	Q	D	S	C	Q	G	S	V	S	W	G	Y	-
N	-	-	Y	-	-	-	-	Q	D	S	C	Q	G	S	V	S	W	G	Y	-
A	-	G	Q	Y	-	-	-	E	D	T	C	Q	G	S	V	S	W	G	Y	-
L	-	-	T	W	-	-	-	K	-	S	C	M	G	S	V	S	W	G	S	S
-	-	-	T	G	-	-	-	-	S	C	M	G	S	-	S	W	G	S	S	S
S	-	G	T	Q	W	-	-	-	A	C	H	G	S	V	S	F	G	S	-	-
R	N	-	-	W	-	-	-	T	-	A	C	-	-	S	V	S	F	G	S	G
K	T	G	L	W	-	-	T	-	-	C	N	G	S	T	S	F	V	S	S	S
R	-	-	L	-	G	A	-	-	-	C	N	G	S	T	S	W	V	V	-	-
L	-	-	I	-	-	-	-	-	V	C	F	G	S	A	S	F	V	R	-	-
L	-	N	L	W	-	-	-	-	-	C	Q	G	S	T	S	F	V	S	R	-
R	-	N	S	W	-	-	T	-	S	C	N	G	S	V	S	F	G	S	S	S
N	E	T	Q	-	R	I	-	-	D	A	C	E	G	S	V	S	W	G	E	-
N	E	T	Q	-	R	I	-	-	D	A	C	E	G	S	V	S	W	G	E	-
N	E	T	Q	-	R	I	-	-	D	A	C	E	G	S	V	S	W	G	E	-
N	-	T	Q	-	R	I	-	-	D	A	C	E	G	S	V	S	W	G	E	-
N	-	-	-	-	-	-	-	K	D	S	C	Q	G	S	V	S	W	G	Y	-
L	-	A	T	W	-	-	-	K	-	S	C	M	G	S	V	S	W	G	S	S
L	-	-	I	Y	-	-	-	I	-	T	C	N	G	S	T	S	F	G	A	A

## Anexo B

Artigo publicado na GMR- Genetic and Molecular Research.



### Finding protein-protein interaction patterns by contact map matching

R.C. Melo<sup>1,2</sup>, C. Ribeiro<sup>1,2</sup>, C.S. Murray<sup>3</sup>, C.J.M. Veloso<sup>1,2</sup>,  
C.H. da Silveira<sup>1,2</sup>, G. Neshich<sup>3</sup>, W. Meira Jr.<sup>3</sup>, R.L. Carceroni<sup>2</sup>  
and M.M. Santoro<sup>1</sup>

<sup>1</sup>Departamento de Bioquímica e Imunologia,  
Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

<sup>2</sup>Departamento de Ciência da Computação,  
Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

<sup>3</sup>Embrapa Informação Tecnológica, Campinas, SP, Brasil

Corresponding author: R.C. Melo

E-mail: raquelcm@gmail.com/neshich@cbi.cnptia.embrapa.br

Genet. Mol. Res. 6 (4): 946-963 (2007)

Received August 03, 2007

Accepted September 25, 2007

Published October 05, 2007

**ABSTRACT.** We propose a novel method for defining patterns of contacts present in protein-protein complexes. A new use of the traditional contact maps (more frequently used for representation of the intra-chain contacts) is presented for analysis of inter-chain contacts. Using an algorithm based on image processing techniques, we can compare protein-protein interaction maps and also obtain a dissimilarity score between them. The same algorithm used to compare the maps can align the contacts of all the complexes and be helpful in the determination of a pattern of conserved interactions at the interfaces. We present an example for the application of this method by analyzing

**A seguir temos o artigo referente a esta tese que foi publicado na BMC – Structural Biology**

METHODOLOGY ARTICLE

Open Access

# Analysis of binding properties and specificity through identification of the interface forming residues (IFR) for serine proteases *in silico* docked to different inhibitors

Cristina Ribeiro<sup>1†</sup>, Roberto C Togawa<sup>2†</sup>, Izabella AP Neshich<sup>3</sup>, Ivan Mazoni<sup>3</sup>, Aduino L Mancini<sup>3</sup>, Raquel C de Melo Minardi<sup>5</sup>, Carlos H da Silveira<sup>4</sup>, José G Jardine<sup>3</sup>, Marcelo M Santoro<sup>1</sup>, Goran Neshich<sup>3\*</sup>

## Abstract

**Background:** Enzymes belonging to the same super family of proteins in general operate on variety of substrates and are inhibited by wide selection of inhibitors. In this work our main objective was to expand the scope of studies that consider only the catalytic and binding pocket amino acids while analyzing enzyme specificity and instead, include a wider category which we have named the Interface Forming Residues (IFR). We were motivated to identify those amino acids with decreased accessibility to solvent after docking of different types of inhibitors to sub classes of serine proteases and then create a table (matrix) of all amino acid positions at the interface as well as their respective occupancies. Our goal is to establish a platform for analysis of the relationship between IFR characteristics and binding properties/specificity for bi-molecular complexes.

**Results:** We propose a novel method for describing binding properties and delineating serine proteases specificity by compiling an exhaustive table of interface forming residues (IFR) for serine proteases and their inhibitors. Currently, the Protein Data Bank (PDB) does not contain all the data that our analysis would require. Therefore, an *in silico* approach was designed for building corresponding complexes

The IFRs are obtained by "rigid body docking" among 70 structurally aligned, sequence wise non-redundant, serine protease structures with 3 inhibitors: bovine pancreatic trypsin inhibitor (BPTI), ecotine and ovomucoid third domain inhibitor. The table (matrix) of all amino acid positions at the interface and their respective occupancy is created. We also developed a new computational protocol for predicting IFRs for those complexes which were not deciphered experimentally so far, achieving accuracy of at least 0.97.

**Conclusions:** The serine proteases interfaces prefer polar (including glycine) residues (with some exceptions). Charged residues were found to be uniquely prevalent at the interfaces between the "miscellaneous-virus" subfamily and the three inhibitors. This prompts speculation about how important this difference in IFR characteristics is for maintaining virulence of those organisms.

Our work here provides a unique tool for both structure/function relationship analysis as well as a compilation of indicators detailing how the specificity of various serine proteases may have been achieved and/or could be altered. It also indicates that the interface forming residues which also determine specificity of serine protease sub-family can not be presented in a canonical way but rather as a matrix of alternative populations of amino acids occupying variety of IFR positions.

\* Correspondence: neshich@cbi.cnptia.embrapa.br

† Contributed equally

<sup>3</sup>Embrapa Information Technologies, Campinas, SP, Brazil

Full list of author information is available at the end of the article

## Background

Serine proteases play an important role in processes such as blood clotting, digestion and in some pathways of cell development [1]. Serine proteases can hydrolyze either peptide bonds or esters. Proteases digest proteins by hydrolyzing the peptide bonds which are responsible for keeping amino acids together [2,3]. The cleavage specificity of elastase, trypsin, chymotrypsin and other serine proteases depends on the volume/size, form/shape, and polarity/charge/hydrophobicity of the specific part of a protein surface where a substrate will be docking - the specificity pocket [4,5]. There are three amino acid residues responsible for the enzymatic activity that are present in all serine proteases, which are denominated as the catalytic triad: His 57, Asp 102 and Ser 195 (chymotrypsin numbering system is used throughout - see [6]). Interestingly, out of those three amino acids, only Asp 102 does not make part of the interface (the definition of which is based on decreased solvent accessible area upon substrate/inhibitor binding). This is due to the fact that Asp 102 is already not accessible to solvent in isolated enzyme, because this amino acid is located at the very bottom of the active site cleft where solvent molecules (water) can not access it. The role of the first two amino acids (His 57 and Asp 102) during trypsin catalysis, for example, is to function as a proton shuttle. Trypsin cleaves peptides after Lys and Arg residues with the co-participation of Asp 189, which interacts with the positive charge on peptide [7,8]. Chymotrypsin, on the other hand, cleaves proteins after aromatic (and also large hydrophobic) residues [9]. To achieve such specificity, in chymotrypsin, one can easily identify the existence of the hydrophobic pocket, normally shielded by Met 192. Thrombin is a protease which cleaves peptides with more specificity than trypsin: it requires Arg on "P1" position [10,11]. Our main objective in this work is to expand the scope of studies that analyze enzyme specificity by including into observations not only the catalytic triad and binding pocket but also a wider category of amino acids which we have named the Interface Forming Residues (IFR) [12-15]. Namely, only a part of the molecular surface is shielded from solvent upon formation of a bi-molecular complex. The residues having a lower accessibility to solvent upon complex formation have an important role in the process of docking and also in defining specificity [15,16]. Therefore, we were motivated to calculate which amino acids are shielded from solvent in the bi-molecular complexes involving serine proteases and different types of inhibitors and then create a table (matrix) of all amino acid positions at the interface and their respective occupancies. By mapping those amino acids as a specific IFR, we are now able to analyze characteristics of each position; and by doing so, we also can make position-specific alignment among different

subfamilies of serine proteases. The key step we needed to solve during our procedural approach was to find a sufficient number of PDB [17] structures containing complexes of serine proteases with respective inhibitors. It became clear that we would need to either produce those by some novel method or abandon our work because there were not enough samples available in the PDB. The solution to this challenge is presented in details in materials and methods.

## Results

### Interface Forming Residue (IFR) tables

In order to provide a sufficient volume of data for analysis of the interfaces around active sites of serine proteases (even though there is no enough structural information available currently on complexes formed by one particular protease with any specific substrate and/or inhibitor) we needed to employ the *in silico* approach for building corresponding complexes. The key feature of our work is mapping the IFR 3D profile into a 2D matrix—from a known enzyme-inhibitor structure to those with no known structure for such complex. Mapping is done after structurally aligning all serine proteases with non-redundant sequences.

After selection of 67 serine protease structural files (in a way described in materials and methods), named here as a secondary datamart, and three prototype complexes, containing inhibitor/serine protease complexes, we ended up with total of 70 structures with non redundant enzyme sequences. The list of all corresponding PDB IDs for serine proteases used in this work is presented in the Table 1. All selected 70 serine proteases were

**Table 1 List of all PDB files containing serine protease structures used for structural alignment**

01-1LTO	02-2FPZ	03-1NN6	04-3RP2	05-1K2I
06-1EQ9	07-1GCT	08-1PYT	09-1ELT	10-1M9U
11-1HNE	12-1GVK	13-1BRU	14-1AGJ	15-1QTF
16-1HPG	17-2O8L	18-1P3C	19-2PKA	20-1GVZ
21-1LO6	22-1TON	23-2FOM	24-2FP7	25-1DY9
26-2SGA	27-1SGP	28-1PQ7	29-1TRN	30-1HJ9
31-1HJ8	32-2F91	33-1H4W	34-2A31	35-1A0J
36-1OS8	37-1OP0	38-1BQY	39-1C5L	40-1VR1
41-2PUX	42-1H8D	43-1ETR	44-1ARB	45-1Y8T
46-2SFA	47-1SOT	48-2H5C	49-1NPM	50-1EP5
51-1FIW	52-1DLE	53-1T32	54-2BZ6	55-1BIO
56-1EUF	57-1EKB	58-1IAU	59-2HLC	60-1Q3X
61-1EAX	62-1FUJ	63-1MBM	64-1FON	65-1RTF
66-1CSY	67-1A7S	<u>68-1FY8</u>	<u>69-1AZZ</u>	<u>70-1CHO</u>

The last three PDB IDs (underlined) contain a description of the native complexes of: anionic trypsinogen with BPTI, collagenase with Ecotin and alpha-chymotrypsin with Ovomuroid, respectively.

then structurally aligned. Upon rigid body docking of the three different inhibitors to 67 serine proteases from the secondary datamart, we constructed complexes which, counting also the three prototype ones, totalled to three sets of 70 two-chain macromolecules. The interfaces for those three sets were calculated using SurfV program [18] and for each *in silico* generated complex, we mapped all of the sequence positions that belonged to IFR ensemble into the table shown at Figure 1.

Analysis of the tabulated IFR data (as presented in the Figure 1) clearly indicates subfamily differences with respect to characteristics, position and frequency of occupancy of interface residues. The data compiled here can be used as an additional tool for characterizing function and specificity. This is particularly relevant for planning structure/sequence changes intended to alter protein activity, e.g., from trypsin to chymotrypsin type activity. More generally, one subfamily type of serine protease can be adjusted to behave like another if corresponding changes are made in an initial protease sequence. Attempts have been made to alter protein activity [4,19], but they have focused on a single or, at most, 5 to 10 residues. Our data, shown on Figure 1, indicate that such alterations can be achieved by eliminating key differences in IFR occupancy.

In case of trypsin, chymotrypsin, elastase and thrombin, each bound to BPTI, differences are present at 36 IFR positions (presented in the table shown at Figure 2). In the table shown at Figure 3, for instance, sequence position 40 is occupied by His (in chymotrypsin and elastase exclusively, and in trypsin mainly) or hydrophobic residues (in thrombins exclusively). Position 172 (across all 4 subfamilies) is occupied by mostly polar residues (with some hydrophobic residues present in elastase case). Positions 94, 215 and 216 are occupied by mostly polar residues in at least 3 subfamilies.

In general, positions were designated as “Ph” if they were occupied by mostly (present in at least 50% of complexes) Polar and occasionally some hydrophobic residues. Similarly, IFR positions occupied by only Charged residues were marked as “C”, those with only Polar residues as “P” and those occupied by only Hydrophobic residues as “H”. A combination of letters would mean presence of residues belonging to multiple/distinct amino acid classes. Capital letters are used to designate a predominant residue presence (again, present in at least 50% of complexes analysed) while small letters designate an occasional presence.

There are less restrictive positions which can be occupied by different classes of amino acids in either one sub family (such as position 143 and 175) or across all subfamilies (positions 37, 60, 96, 97, 192 and 224). A specific analysis of the position 97, shows that it is less

restrictive for trypsins (positively charged: K, R; polar: N, Y; negatively charged: D) and elastases (polar: N; negatively charged: D; hydrophobic: V) than for chymotrypsins (hydrophobic: L, F) and thrombins (positively charged: R).

In order to additionally facilitate analysis of the data at Figures 2, 3 and 4, we have labelled with the gray background color those IFR positions which are populated, within a subfamily, by a single residue type or are absent from the IFR ensemble (represented by “-”). It is important to realize that the presence of one type of amino acid per subfamily is treated here in a same fashion as the absence of that particular position from the IFR ensemble because we understand that the event we call “no-show” for particular position is equally information dense as the information on type of occupancy for particular IFR position. Those IFR positions occupied in each one of four sub families by one predominant type of amino acid, present in at least 50% of serine proteases showing this particular IFR, marked here with the “\*”, are labelled in light gray. Therefore, the analysis of Figures 2, 3 and 4 is aided by four distinct labels: 1) residue “amino-chromography”, 2) dark gray background, 3) light grey background and 4) “occupancy type”. The fourth label is present in the last column and indicates the two predominant residue classes present in the preceding 4 columns (four serine protease subfamilies). Similar analysis was done for two other inhibitors: BPTI with 36 identified IFR positions at the enzyme side (Figure 2), and ovomucoid third domain with 34 identified IFR positions at the enzyme side (the table shown at Figure 4).

From the data presented in Figures 2 and 4 (restrictive IFR positions of complexes formed with BPTI and ovomucoid third domain inhibitors, respectively) IFR locations 94 and 172 are shown to contain only polar residues (no presence of hydrophobic residues was found in contrast to the same locations in the Figure 3). By comparing the data from the figures 2, 3 and 4, the complexes of serine proteases with ecotine inhibitor compile 13 restrictive (gray), 10 half-restrictive (light gray) and 17 non-restrictive IFR positions (total: 40 positions). In the case of the BPTI and ovomucoid inhibitors, respective data are: 14, 7 and 15 (36); 11, 7 and 16 (34). Those numbers indicate that ecotine binding to serine proteases involves 32.5% restrictive, 25% less restrictive and 42.5% non restrictive IFR positions while BPTI and ovomucoid show respectively: 39%, 19.5% and 41.5%; 32%, 20.5% and 47%. Consequently, these numbers show that the ovomucoid inhibitor binding to serine proteases is less specific than the binding of ecotine, which in turn is less specific than BPTI. The experimental  $K_i$  [M] values reported in the literature for BPTI against trypsin is  $6 \times 10^{-14}$  while for the chymotrypsin it goes down to  $9 \times 10^{-9}$  and for elastase to  $3.5 \times 10^{-6}$

1	1LTO	V	37	
2	2FPZ	M	37C	
3	1NN6	S	39	
4	3RP2	R	40	
5	1K2I	R	41	
6	1EQ9	S	42	
7	1GCT	H	42	
8	1PYT	R	42	
9	1ELT	Y	47	
10	1M9U	R	47	
11	1HNE	L	47	
12	1GVK	Y	47	
13	1BRU	Y	47	
14	1AGJ	Q	47	
15	1QTF	S	47	
16	1HPG	G	47	
17	2O8L	T	47	
18	1P3C	G	47	
19	2PKA	H	47	
20	1GVZ	H	47	
21	1LO6	T	47	
22	1TON	E	47	
23	2FOM	S	47	
24	2FP7	Y	47	
25	1DY9	T	47	
26	2SGA	G	47	
27	1SGP	T	47	
28	1PQ7	R	47	
29	1TRN	Y	47	
30	1HJ9	Y	47	
31	1HJ8	Y	47	
32	2F91	E	47	
33	1H4W	S	47	
34	2A31	S	47	
35	1A0J	Y	47	
36	1OS8	M	47	
37	1OP0	N	47	
38	1BQY	G	47	
39	1C5L	R	47	
40	1VR1	R	47	
41	2PUX	R	47	
42	1H8D	R	47	
43	1ETR	R	47	
44	1ARB	K	47	
45	1Y8T	E	47	
46	2SFA	G	47	
47	1SOT	N	47	
48	2H5C	A	47	
49	1NPM	Q	47	
50	1EP5	L	47	
51	1FIW	I	47	
52	1DLE	H	47	
53	1T32	I	47	
54	2BZ6	V	47	
55	1BIO	L	47	
56	1EUF	F	47	
57	1EKB	Q	47	
58	1IAU	L	47	
59	2HLC	R	47	
60	1Q3X	G	47	
61	1EAX	G	47	
62	1FUJ	S	47	
63	1MBM	S	47	
64	1FON	Y	47	
65	1RTF	F	47	
66	1C5Y	R	47	
67	1A7S	N	47	
68	1FY8	Y	47	
69	1CHO	F	47	
70	1A2Z	M	47	

**Figure 1** The IFR positions and their occupancy by amino acids shown for 70 serine proteases bound to BPTI. The rows show amino acids occupying IFR positions for specific serine protease (the PDB ID given in the second column). The top row shows the sequence numbers in structurally aligned serine proteases (only IFRs are listed) for all enzymes in complex with the BPTI. The complexes were created by the procedure we named "hard docking", except for the case of 1fy8.pdb for which we have experimental data (row marked with light gray background). The color code for amino acids is as follows: Residues: **AVLIMFP** are colored grey [containing: small hydrophobic residues (**A** and **P**) and large hydrophobic residues]; Residues: **STYNQWG** are colored green [containing: polar residues]; Residues: **D** and **E** are colored red [containing: negatively charged residues]; Residues: **R** and **K** are colored blue [containing: positively charged residues]; Residue: **H** is colored cyan [containing: positively charged residue which is neutral at neutral pH]; Residue: **C** is colored yellow [containing: disulphide bridge forming residue]. [See text for details about amino acid classification] The rows are occupied by following serine protease sub-families: Trypsin(1-2), Chymase (3-4), Chymotrypsin (5-8) [red background], Elastase (9-13)[blue background], Exfoliative toxin (14-15), Glutamyl endopeptidase (16-18), Kallikrein (19-22), NS3 protease (23-25), Streptogrisin (26-27), Trypsin (28-36)[green background], Venom (37-38), Thrombin (39-43){yellow background}, Miscellaneous-Prokaryotes (44-48), Miscellaneous-Virus (49-50), Miscellaneous (51-67), Native Complex (68-70) in respective row order.

#	AA#	Trypsin	Chymotrypsin	Elastase	Thrombin	Occupancy Type
1	37)	R, E, M	Y	Y*, R, L	R*	
2	37C)	-	-	-	P*	H
3	39)	Y, G, F, S	F, S, R	Y, S, G, A, R	E*, R	
4	40)	P, H*	H*	H*	L*, F	CH
5	41)	F*, W, G	F*, R, T	T*, S, F	L*	
6	59B)	-	G	G, N	Y*	P
7	59D)	-	-	-	P*	H
8	59E)	-	-	-	W*	P
9	59F)	-	-	-	D*	C
10	59G)	-	-	-	K*	C
11	59H)	-	-	-	N*	P
12	60)	K*, Y	S	D, V, S	F*	
13	73)	I*	L*, Q	L, Q	R*	
14	94)	Y*	Y, W	W*, Y	-	P
15	96)	R, S, Y, G	S*, D	S, P, T	W*	
16	97)	K, N, Y, D, R	L*, F*	D, V, N	R*	
17	97A)	-	-	-	E	C
18	99)	N, I, L*	I, L, V	V, Y, L	L*	HP
19	141)	W*	-	-	W	P
20	142)	G	-	-	G*	P
21	143)	N, T	L, S, R	K, R, L	N*	
22	143Q)	T, Y, Q	T*	L*, I, S	Q*	PH
23	172)	-	W*	W*	-	P
24	174)	-	-	-	I*	H
25	189)	D*	S	-	D*	CP
26	190)	S*, T	S*, A	A, V, S	A*	PH
27	192)	Q*	M*, H	N*, F, Q	E*	
28	213)	V*	V*	T*, A, V	V*	Hp
29	215)	W*	W*, F*	F*, W	W*	PH
30	216)	G*	G*	V*, G	G*	PH
31	217)	G*	S*	S*, V, R	E*	
32	217A)	-	S*, G	S, R	-	PC
33	219)	G*	T, P	S	G*	PH
34	226)	G*	G, D, T	S	G*	PC
35	227)	V*	V	-	F	H
36	228)	Y	Y*	F	Y*	PH

**Figure 2** An extract from all IFR positions showing the most restrictive ones, identified for four serine proteases subfamilies, bound to BPTI. The positions shown are among the most restrictive ones in terms of the type of residues occupying them in 4 sub families of serine proteases: trypsin; chymotrypsins; elastases and thrombins, respectively. Those residue types that are present in at least 50% or more of analyzed proteases within a single subfamily, at a given IFR position, are indicated by a '\*' sign. The colors of residues correspond to STING AA color coding [see text for details as well as the legend for the Table 1]: green are Polar, gray are Hydrophobic, red are negatively Charged, blue (and cyan for His) are positively Charged residues. The gray background color is used to indicate those IFR positions which are populated, within a sub family, by a single residue type or are absent from the IFR ensemble (represented by "-"). Those IFR positions occupied in each of four sub families by one predominant class of amino acid (indicated by "\*\*") are labeled in light gray.

(ref: <http://www.gbiosciences.com/ResearchProducts/Aprotinin.aspx>). On the other side, turkey ovomucoid third domain was reported to interact with very similar association constant with eight different serine proteinases-bovine chymotrypsins A and B, porcine pancreatic

elastase I, proteinase K, Streptomyces griseus proteinases A and B, and subtilisin-Carlsberg [20]. Pál et al. [21] reported the following  $K_i$  (M) values for ecotine against trypsin, chymotrypsin and elastase, respectively:  $1 \times 10^{-12}$ ,  $4 \times 10^{-12}$  and  $1.3 \times 10^{-9}$ .



#	AA#	Trypsin	Chymotrypsin	Elastase	Thrombin	Occupancy Type
1	37)	R,E,M	D,Y	Y*, R	R*	
2	37C)	-	-	-	P*	H
3	39)	Y, F	F, R	Y, R	E*,R	
4	40)	H*, P	H*	H*	L*, F	CH
5	41)	F*, W, G	F*, R, T	T*, S, F	L*	
6	59)	Y	G*	-	-	P
7	59B)	-	-	N	Y*	P
8	59D)	-	-	-	P*	H
9	59E)	-	-	-	W*	P
10	59F)	-	-	-	D*	C
11	59G)	-	-	-	K*	C
12	60)	K*, Y	S	V, D	-	
13	94)	Y*, F	Y*, W	W*, Y	-	Ph
14	96)	R, S, Y, G	S*, D	S, P, T	W*	
15	97)	K, N, Y, D,R	L*, F*	D, V, N	R*	
16	97A)	-	-	-	E*	
17	98)	T*, G, N, L	T*, L*	D, S, N,Q	N*	
18	99)	L*, I, N	I*, L, V	V, Y, L	L*	Hp
19	143)	N, T	L, S, R	K, R, L	N*	
20	143D)	E	N	T	E*	CP
21	143H)	G	G	G, N	T*	P
22	143Q)	T, Y, Q	T*	L*, I, S	Q*	PH
23	171)	Q, S, D, A	Y, D	-	S*	
24	172)	Y*	W*, Q	W*, M	T*	Ph
25	173)	P*, G	G*, W	G*	R*	
26	174)	G*, S, D, N	T*	S*, A	I*	
27	175)	A, K, Q,M,E	K*, R, T	T*, N	R*	
28	177)	-	K*	K*	-	C
29	189)	D*	-	-	D*	C
30	190)	S*, T	S*, A	V, S	A*	PH
31	192)	Q*	M*, H	N*, F, Q	E*	
32	213)	V*	V*	T*, A, V	V*	HP
33	215)	W*	W*, F*	F*, W	W*	PH
34	216)	G*	G*	V*	G*	PH
35	217)	Y*, N,D,S,H	S*	S*, V, R	E*	
36	217A)	-	S*, G	S, R, S	-	Pc
37	219)	G	T, P	S, G	G*	Ph
38	221A)	R	S	-	R*	Cp
39	224)	Y, K, N,R	T, E, K	K*, Y	K*	CP
40	226)	G*	D, T	-	-	CP

**Figure 3** An extract from all IFR positions, identified for four serine proteases subfamilies, bound to Ecotine. Shown positions are designated in a same form as in the Figure 2 above.

#### Amino acid composition at serine protease interfaces

Additionally, the difference between the polarity and hydrophobicity of the enzyme surface and interface was investigated for each *in silico* complex formed. The buried area and total enzyme free surface area (not including the interface) was compared with respect to the type of residues present. Again, the following residue classes were used: charged, hydrophobic and polar (with glycine considered separately, although being classified as a polar residue). The data presented in Figure 5 indicate that most interfaces, in all serine protease subfamilies bound to either BPTI, ecotine or ovomucoid third domain, have

a general preference for polar residues (green and gray bars). Conversely, most of free enzyme surfaces (not including interface/buried area) show a preference for charged residues (blue bars) although some (7 out of 16) also do show a smaller (as compared to a preference for charged residues) preference for hydrophobic residues (red bars).

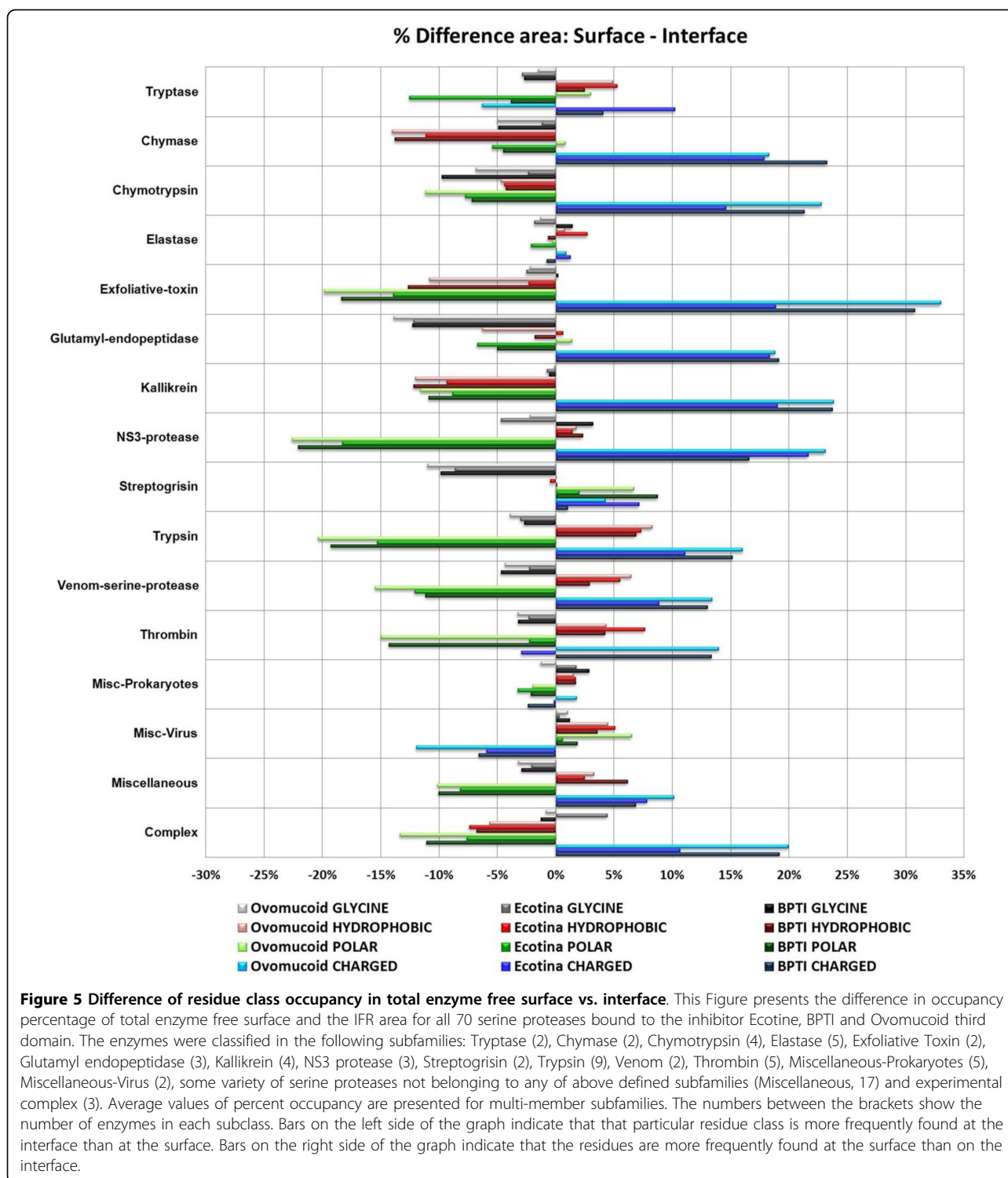
Some subfamilies do not follow these patterns. The interface of the streptogrisin subfamily shows a high presence of glycine residues for complexes with each of the three inhibitors (negative gray bars in Figure 5), while the rest of the surface shows preference for polar

#	AA#	Trypsin	Chymotrypsin	Elastase	Thrombin	Occupancy Type
1	37)	R,E,S,M	D,L,Y	Y*,R,L	R*,A	
2	37C)	-	-	-	P*	H
3	38)	N	-	R	Q*	PC
4	39)	G,Y,F,S	F,S,R	Y,S,G,A,R	E*,R	
5	40)	P,H*	H*	H*	L*	CH
6	41)	F*,W,G	F*,R,T	T*,S,F	L*	
7	59)	Y*,V	G	-	-	PH
8	59B)	-	-	N	Y*	P
9	59D)	-	-	-	P*	H
10	59E)	-	-	-	W*	P
11	59F)	-	-	-	D*	C
12	59G)	-	-	-	K*	C
13	60)	K*,D,Y	V,L,S	V,S	F*	
14	73)	R	-	-	R*	C
15	74)	E,N,K	-	-	T*	CP
16	94)	Y	Y,W	W*,Y	-	P
17	97)	K,N,Y,D,R	F*	D,N	-	
18	99)	L*,N,I	I,L,V	V,Y,L	L*	HP
19	143)	N*,A,T	L,S,R	K,R,L	N*	
20	143D)	E	N	T	E*	CP
21	143H)	G	G	G,N	T*	P
22	143Q)	T,Y,Q	T*	L*,I,S	Q*	PH
23	172)	Y*	W*,Q	W*	T	P
24	173)	-	G,W	G	R*	PC
25	174)	-	-	A	I*	H
26	175)	K,Q,M,E	K,T	T*	-	
27	189)	D*	-	-	D*	C
28	190)	S*,T	S*,A*	V,S	A*	PH
29	192)	Q*	M*,H	N*,F,Q	E*	
30	213)	V*	V*	T*,A,V	V*	PH
31	215)	W*	W*,F*	F*,W*	W*	PH
32	216)	G*	G*	V*,G	G*	PH
33	217)	Y*,N,D,S,H	S*	S*,V,R	E*	
34	217A)	-	S*,G	S,R	-	PC

**Figure 4** An extract from all IFR positions, identified for four serine protease subfamilies bound to Ovomuroid third domain. Shown positions are designated in a same form as in the Figure 2 and 3 above.

residues (green bars). Elastase shows the lowest differences among interface and surface preferences for 4 classes of amino acids defined in this work. Charged residues were found to be uniquely prevalent at the interfaces between the “miscellaneous-virus” subfamily and the three inhibitors. This prompts speculation about how important this difference in IFR characteristics is for maintaining virulence of those organisms. The upper part of the Figure 5 shows a strong presence of hydrophobic residues at the interfaces (together with polar residues including glycine), while the bottom part

of the Figure 5 shows an absence (except for the 3 experimental complexes) of the hydrophobic residues at the interface (left portion of this graph). The bottom part of the Figure 5 shows an increase in the presence of hydrophobic residues at the free enzyme surface area. The surfaces (not including interface areas) have a prevalence of charged residues while the interface contains more polar (including glycine) residues (with some exceptions as described above). Thus, the IFR pocket of serine proteases is not formed by predominantly hydrophobic residues; it is a rather polar environment. Also,



the interface contains much fewer (with one exception being “miscellaneous-virus”) charged residues (as compared to the rest of the protease surface).

If our data are compared with the ones from Janin et. al. [22], one can observe that the serine proteases are

endowed with a specific category of interfaces in terms of the contribution of the 20 amino acids types to that area: subfamilies vary in the contribution of hydrophobic residues and are almost constant in having a lower preference for the charged residues, as presented at the

Table 2. These results point to the fact that although all four serine protease subfamilies do not show a great variability with respect to the percentage of the enzyme total surface area being buried upon *in silico* complex formation with the BPTI, ecotine and ovomucoid third domain (6.7%, 9.5% and 7.2% respectively, representing value of approximately 610 to 990 Å<sup>2</sup> of buried area), the percentage of the residue classes contributing to the interface varies significantly. The percentage that charged and polar residues contribute together to the total buried area in trypsin/BPTI is 86%, in chymotrypsin/BPTI is 65%, in elastase/BPTI is 75%, and for thrombin/BPTI is 84%. For all three inhibitors examined, thrombin and trypsin have the highest percentage of charged and polar residues contributing to buried area while elastase and chymotrypsin have the lowest. The details for the difference in occupancies are presented in Figure 5 and Table 2.

Those differences may be aligned with the corresponding function that each protease performs. However, it is important to realize the boundaries we should not cross when applying what we have learned here based on inhibitor-protease binary complexes to substrate-protease complexes. Nevertheless, we assume that inhibitors of serine proteases bind to the active site clefts of their target proteases in a manner that is thought to resemble the binding mode of substrates (Laskowski et al. [23]). Trypsin and thrombin handle the positive charges of a polypeptide by imposing the (complementary) charge/polarity of their own IFR surface.

Chymotrypsin mitigates the spatial constraints imparted by large hydrophobic (and some polar) residues present in the substrate/inhibitor by accommodating glycines and hydrophobic residues into its own interface. Elastases, on the other hand, utilize hydrophobic amino acids among their IFRs to complement small, hydrophobic amino acids at the cleavage site of a polypeptide. The general trends for interfaces described above, when combined with the selectivity of IFR positions discussed earlier in this section, provide a more complete description of the specificity of serine protease subfamilies in both general and detailed way, although, as we expected from the beginning, there is no canonical code that we may convey from this analysis.

#### Benchmarking the accuracy of the "hard docking" procedure for IFR identification

The "confusion matrix" presented at the Table 3 contains 5 categories of evaluators for IFR identification. The true positives (TP) are IFR residues present in both the native and the rigid body docked complexes. The true negatives (TN) are solvent exposed residues both in native and in *in silico* complexes. The false positives (FP) are residues present in the rigid body docked IFR table but absent in natural complexes and the false negatives (FN) are not buried in native complexes but appear in the IFR of the *in silico* complexes. The accuracy was calculated based on this formula:  $ACC = (TP + TN)/(P + N)$ , where the  $P = (TP + FN)$  and  $N = (FP + TN)$ . Table 3 shows these values for 1FY8 (control)

**Table 2 The percent of the total enzyme buried area occupied by amino acid sub-classes**

Inhibitor bound to serine proteases		[%] Hydrophobic residues in buried area	[%] Charged residues in buried area	[%] Polar residues at buried area (glycine %)	[%] (Charged + Polar) residues in buried area
BPTI	Trypsin (9)	14	10	76 (10)	86
	Chymotrypsin (4)	35	11	54 (8)	65
	Elastase (5)	25	21	54 (6)	75
	Thrombins (5)	16	37	47 (6)	84
Ecotine	Trypsin (9)	13	15	72 (10)	87
	Chymotrypsin (4)	30	18	52 (6)	70
	Elastase (5)	22	19	59 (9)	78
	Thrombins (5)	12	53	35 (6)	88
Ovomucoid Third domain	Trypsin (9)	12	10	78 (11)	88
	Chymotrypsin (4)	32	10	58 (9)	68
	Elastase (5)	24	20	56 (9)	86
	Thrombins (5)	15	37	48 (7)	85

The serine proteases are bound to BPTI, ecotine and ovomucoid third domain inhibitors. Areas occupied by the three classes of amino acids, with respect to the total buried enzyme area, are expressed as a percentage. The numbers between the brackets show the number of enzyme - inhibitor complexes included in this analysis for each subfamily. The column containing polar residue participation also contains (within brackets) the participation of glycine residues.

**Table 3 Evaluation of the success rate (ACC) of predicting IFR ensemble**

PDB ID	Subfamily	Organism	TP	TN	FP	FN	ACC
1FY8	Trypsin	<i>Bos taurus</i>	44	244	0	0	1
3TGI	Trypsin	<i>Bos taurus</i>	42	239	5	0	0.98
1TPA	Trypsin	<i>Bos taurus</i>	44	240	1	3	0.98
3BTK	Trypsin	<i>Bos taurus</i>	43	241	1	4	0.98
3TPI	Trypsin	<i>Bos taurus</i>	43	242	1	2	0.98
2PTC	Trypsin	<i>Bos taurus</i>	44	241	0	3	0.98
2TGP	Trypsin	<i>Bos taurus</i>	45	240	0	3	0.98
4TPI	Trypsin	<i>Bos taurus</i>	44	239	1	4	0.98
2TPI	Trypsin	<i>Bos taurus</i>	44	241	1	2	0.98
1BZX	Trypsin	<i>Bos taurus</i>	43	242	2	1	0.98
1BRB	Trypsin	<i>Rattus norvegicus</i>	41	236	2	2	0.98
1FAK	Coagulation factor VIII	<i>Homo sapiens</i>	44	247	6	0	0.97
1CBW	Chymotrypsin	<i>Bos taurus</i>	44	238	5	1	0.97
1MTN	Chymotrypsin	<i>Bos taurus</i>	45	239	1	3	0.98
2KAI	Kallikrein	<i>Bos taurus</i>	57	224	0	4	0.98
1EAW	Matriptase MTSP1	<i>Homo sapiens</i>	42	241	4	1	0.98

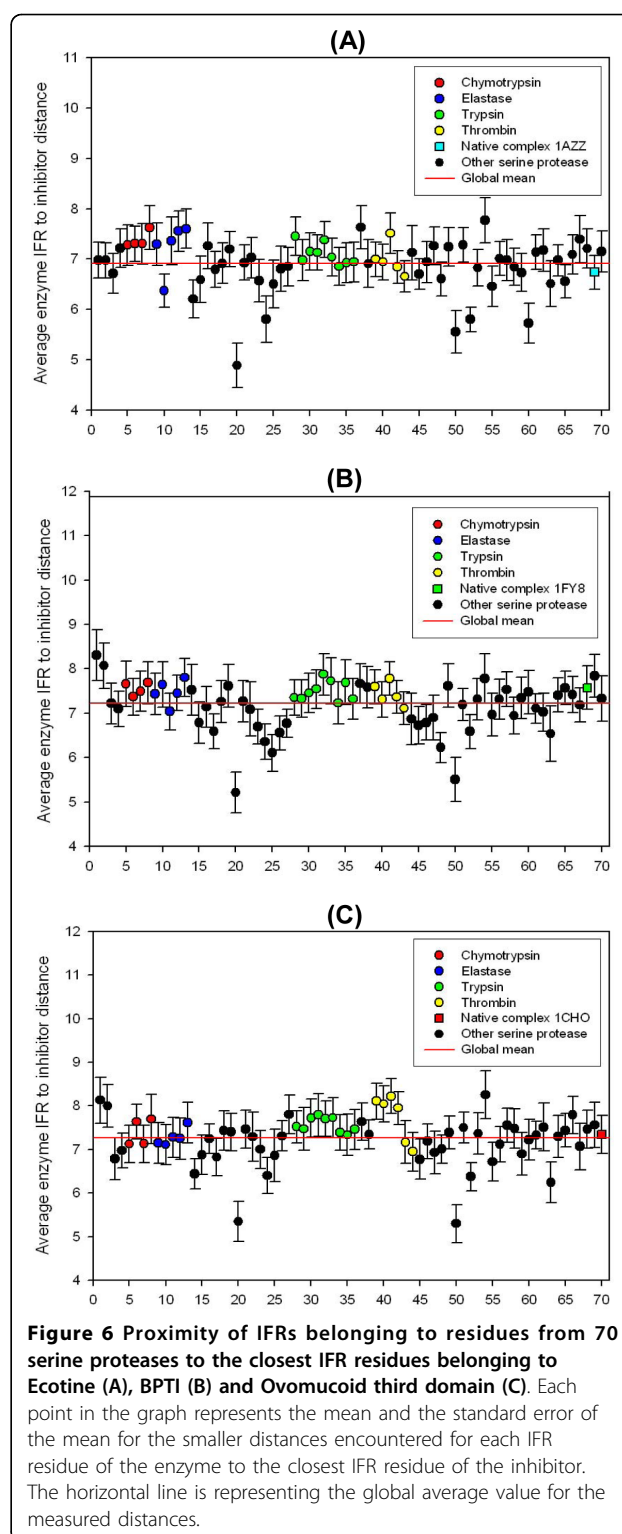
The true positives (TP) are IFR residues present in both the native and the rigid body docked complexes. The true negatives (TN) are solvent exposed residues both in native and in *in silico* complexes. The false positives (FP) are present in the rigid body docked IFR but absent in natural complexes, and the false negatives (FN) are not buried in native complexes but appear in the IFR in generated in *in silico* complexes.

and the sixteen complementary complexes. Overall, the accuracy (ACC) ranged between 0.97 and 0.98.

#### Evaluation of plausibility for *in silico* generated bi-molecular complexes

The presence of steric clashes involving enzyme and inhibitor residues was analyzed to evaluate our hard docking methodology. Steric clashes could yield unfavourable interactions within complexes that would lead to their decreased stability which might be translated to their faster de complexation *in vivo*.

We calculated the distances between enzyme and inhibitor residues at the interface for BPTI (Figure 6A), ecotine (Figure 6B) and ovomucoid third domain complexes with serine proteases (Figure 6C). All 70 serine proteases are presented using the same numbering (x-axis) as in Table 1, and the same subfamily grouping (colored symbols follow the colors of subfamilies as in Figures 2, 3 and 4). The y-axis shows the distances, measured in Ångstroms, with bars indicating a corresponding standard deviation for each point. The distances are the average values obtained by calculating the sum of the CA-CA (Alpha Carbons) distances for each amino acid belonging to the IFR ensemble from the enzyme side, to the closest amino acid from the IFR of



**Figure 6 Proximity of IFRs belonging to residues from 70 serine proteases to the closest IFR residues belonging to Ecotine (A), BPTI (B) and Ovomuroid third domain (C).** Each point in the graph represents the mean and the standard error of the mean for the smaller distances encountered for each IFR residue of the enzyme to the closest IFR residue of the inhibitor. The horizontal line is representing the global average value for the measured distances.

the inhibitor, and then divided by the number of IFR residues at the enzyme site. A horizontal line represents the average global mean which was always near 7.0 Å for all three types of inhibitor complexes. In a prior

paper [24], our research group demonstrated that 7.0 Å is the cut-off distance that optimizes the isolation, at the residue level, of first order of contacts (first coordination shell) in well-packed globular proteins. The observed 7.0 Å average global mean distance among CAs in enzyme-inhibitor complexes confirms that inhibitors and enzymes form well-packed complexes like those found in the core of globular proteins. Thus, it is plausible to conclude that complexes which present distances less than the average global mean contain spatial clashes. Those complexes can be further submitted to dynamics and minimization in order to rearrange these residues and diminish the steric clashes. We are set up to try to evaluate how many and which steric clashes would be removed by such procedure, however, this is yet to be done as a next step of our project.

## Discussion

The data collected, analyzed and presented in this work serves as a tool for understanding binding properties in general and protease specificity in particular, based on Interface Forming Residue (IFR) profile alignment.

The procedure we have developed for this work describes a): the ensemble of IFRs for any complex formed between a serine protease and an inhibitor and b): the differences in interface characteristics between various proteases with respect to specificity. From our data one may not extract a simple, concise and very straightforward rule for separation among proteases in terms of specificity (defined here mainly by IFR position occupancy and characteristics). Nature obviously prepared a mechanism for fine tuning of the activity that yields terms of specificity for each subfamily of serine proteases. There is obviously some space left for overlaps among IFR characteristics and positions that dictate the specificity for each sub family and consequently, what could have been a simple table with reduced complexity in terms of “colors” and characteristics of amino acids employed at each IFR position, became somewhat blurred. The data presented here are precise in pointing to such granular/spectral distribution in IFR occupancy that would be directly coupled to the specificity type of studied enzymes. It is our perception that the beauty of the adaptability of molecular mechanisms, expressed here in form of their specificity, can still be presented in a such simplified way (described in Figures 1 to 4) showing how specificity can be fine-tuned from different points of entry (multiple, rather than single): IFR positions and characteristics of amino acids that occupy those critical positions. The procedure we described may also open the possibility to yield some details that may account for differences in how strong might be the binding between an enzyme and an inhibitor (an issue we need to explore in

the future with more details based on elaborated IFR tables and experimental data available in literature).

It has been established that increased structural plasticity in the binding pocket increases the variation of substrate size that can fit into the critical space directed toward the catalytic triad which correspondingly decreases the specificity [25]. Consequently, it is expected that the higher the stiffness around the binding pocket, the higher the selective pressure will be on a particular substrate. Specificity then can be seen as directly proportional to structural limitations imposed first by the size of the docking space and then by the physical and chemical characteristics of this space. By focusing our attention on the type of residues occupying the enzyme and inhibitor interfaces, we have set aside the evaluation of how plasticity of binding pocket influences enzyme specificity. This factor, however, is not taken as a non important one. In fact, we will be undertaking further examination in order to understand better how plasticity may be accounted for by IFR properties - therefore we will be tackling this task in the future. As a start, the residue types occupying the interface positions implicitly contain the sort of information which relates directly to plasticity of the site.

## Conclusions

The work presented here offers insight into the structure/function relationship of serine proteases. The superposition of structurally aligned backbones of serine proteases showed that they use essentially identical scaffolding and achieve a variety of specificities by varying surface residues. Previously, there have been attempts to modify the functionality of serine proteases. These attempts usually focus on several residues and generally result in the loss of enzyme activity. Our approach explains why and corroborates with the results presented by Ma et al. [26] and Novozymes Biotech, Inc. (Davis, CA, US) which patented a technology (Microbial trypsin mutants having chymotrypsin activity - United States Patent 20050037368) by confirming necessity for several IFR amino acids to be either substituted, or deleted (“no-show” event) or to be introduced in original protease sequence in order to transform trypsin to chymotrypsin.

Based on the results we are reporting here, we are poised to assemble, in the near future, the serine protease superfamily interface data resource as an expanding collection of sequence, structural, and functional information about the serine proteases interface forming residues. A combination of graphics, images and numerical data will be used to aid in the complete analysis of structure/function relationships that still require our attention before we may fully understand the precise

role of each amino acid employed at the particular location by enzymes in order to achieve desired specificity.

## Methods

In order to provide a sufficient volume of data for analysis of the interfaces around active sites of serine proteases (even if there is no currently available information on complex formation of one particular protease with any specific substrate and/or inhibitor) we needed to employ the *in silico* approach for building corresponding complexes. The key feature of our work is mapping the IFR 3D profile into a 2D matrix—from a known enzyme-inhibitor structure to those with no known structure for such complex. Mapping is done after structurally aligning all serine proteases with non-redundant sequences.

### Selection of serine protease: structural data set

The PDB IDs for serine proteases were obtained from SCOP database [27] parseable files (release 1.73). All 1086 PDB IDs from family b.47 were selected. FASTA files relative to these IDs were obtained in the PDB. A BLAST search (blastp), using default parameters, was run to determine the percentage of identity between the selected sequences. Those sequences with lower than 95% identity were selected for further examination (meaning that there should be at least 12 to 15 different residues occupying corresponding positions among selected sequences. In general, the sequences are 240 to 260 amino acids long, which means that applied filtering eliminates structures obtained after single, or even after limited multiple, point mutations). From the resulting subset (which we named here as the “primary datamart”), further selection eliminated all but those structures showing the highest resolution (lower numerical values in Å) and best R-value. This yielded what we named as a “secondary datamart”. The “secondary datamart” at this point contains 67 serine proteases. Those structures covered the following sub families: tryptase, chymase, chymotrypsins, elastase, exfoliative toxin, glutamyl endopeptidase, kallikrein, NS3 protease, streptogrisin, trypsin, venom, thrombin, serine proteases from prokaryotes (which we named as “miscellaneous-prokaryotes”), serine proteases from viruses (which we called “miscellaneous-virus”) and serine proteases that do not fit any of the mentioned sub families which we called “miscellaneous”. The above described division resulted in 15 “sub-families” on which we applied additional manual inspection (in order to re-confirm their classification) by consulting the following sources: PDBsum [28], BRENDA [29] and SCOP parseable files.

The list of all corresponding PDB IDs for serine proteases used in this work is presented in the Table 1.

### Selection of experimentally described enzyme/inhibitor structural complexes

From the 57 available PDB files containing inhibitor/serine protease complexes (data from January 2009), comprising 11 SCOP inhibitor families, the three largest families were selected and the same approach as described above was followed to select for what we named here as the three “prototype complexes”. Complexes containing BPTI, ecotin and ovomucoid inhibitors correspond respectively to g.8, b.16 and g.68 SCOP subfamilies and are represented, respectively, by 1FY8 (inhibiting rat trypsin), 1AZZ (inhibiting crab collagenase) and 1CHO (inhibiting cattle alpha-chymotrypsin) PDB files. We named this group as “prototype complexes” because they represent the experimentally described complexes with structures deposited into the PDB.

### Structural alignment

Before the alignment procedure took place, we needed to edit the PDB files. The 1CHO.PDB file was edited to replace the three enzyme chain IDs (E, F, G) with a single “chain E”. This was done in order to easily distinguish the enzyme chains from the inhibitor chain named “I”. For the 1AZZ file, containing two chains, only chain A (enzyme) and chain C (inhibitor) were used (the PDB file contains a description of a heterodimer). The 1FY8 was not edited before further processing. In order to obtain an IFR ensemble for the serine proteases lacking an available structure for the complex with chosen inhibitor (or any inhibitor at all), we first structurally aligned all 70 selected serine proteases (67 from the “secondary datamart” plus 3 serine proteases from the “prototype complexes”) using PrISM [30] and its set of default parameters, [http://wiki.c2b2.columbia.edu/honiglab\\_public/index.php/Software:PrISM](http://wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:PrISM). The PrISM software also provided a multiple sequence alignment (being generated as an output obeying the structural alignment), which was saved and parsed to preserve both residue numbering and residue positions occupied in the structural alignment.

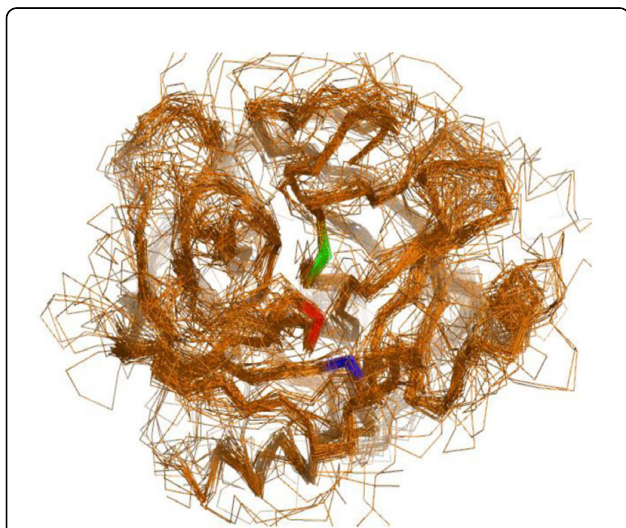
### Rigid body docking

The term “rigid body docking” is used here to describe what was done in order to get complexes of each of the 67 serine protease structures bound to 3 different inhibitors. Effectively what we did was added the inhibitor coordinates to each PDB file from the “secondary datamart”. In order to do so, we first needed to guarantee that the inhibitor molecule would be added to any of the enzyme molecules in exactly the same fashion as it did in the “prototype complex”. This guarantee was provided by aligning a pair of structures where the first was the



enzyme structure from the original “prototype complex” (e.g. 1FY8) and the second was from the PrISM output file - one of the 70 structurally aligned enzymes (e.g. 1FY8\_PrISM). This additional structural alignment was performed with PyMOL [31] software, using the command line “align (1FY8, 1FY8\_PrISM)”, which maintained the PrISM generated coordinates of the 1FY8\_PrISM structure and superimposed the enzyme coordinates from a selected “prototype complex”. Consequently, this procedure also relocated the inhibitor coordinates from the “prototype complex”. The newly obtained coordinates for the inhibitor were then extracted from the PyMOL output (PDB formatted) file and used to edit the PrISM output file for the other 69 serine proteases. By doing so, we included the inhibitor coordinates now fitting to a selected enzyme in a position exactly equivalent to the one that the inhibitor assumed in the “prototype complex”. The three sets (each containing 70 complexes) were therefore formed. The details of the procedure described above are illustrated in Figures 7, 8 and 9. Upon completing the procedure, all structures containing a complex between a serine protease (one of 70 of them) and an inhibitor (one of the three selected ones) could be described as spatially oriented in a consistent and unique manner (obeying structural alignment) and having the inhibitor in a position consistent with the one observed and described experimentally.

As the primary purpose was to delineate a general area defined as an interface, additional precision, as obtained by minimization and dynamics procedures



**Figure 7 Structural alignment of 70 different (sequence-wise non redundant) serine proteases aligned by PrISM package.**

The image was produced using PyMOL. Only the main chain is represented. The following positions in chymotrypsinogen are highlighted: Ser 195 in green, His 57 in red and Asp 102 in blue.

(generally applied to remove space-clashes), was not considered critical and therefore not applied in this stage of the work. It is important to note that many complexes are in fact “forced” to be formed *in silico* by a “rigid body docking” procedure, but due to various factors, including insufficient shape complementary between the inhibitor side and corresponding binding pocket on the enzyme side, they would form with a certain degree of instability and would therefore be short lived. A qualitative description of this scenario is described in Figure 9 (also see Figure 6. for an estimate of the degree of space clashing among enzymes and inhibitors in all *in silico* formed complexes).

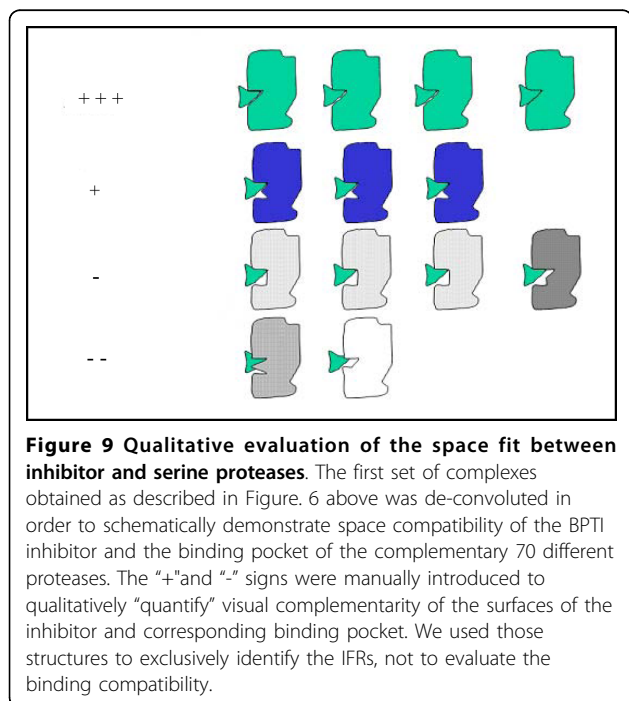
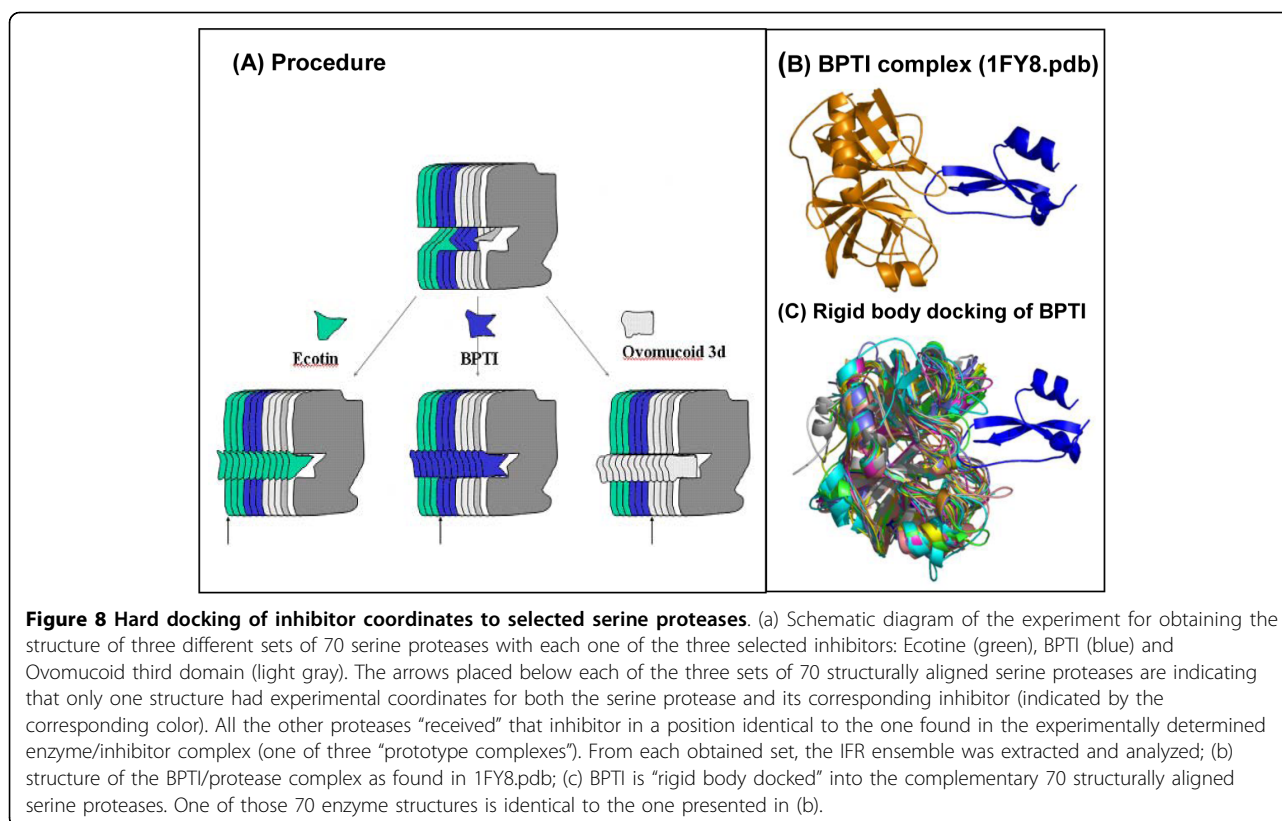
#### Surface determination in rigid body docked complexes

The three sets of 70 enzyme-inhibitor complex structures were also studied with respect to solvent accessibility of surface amino acids (calculated both before and upon inhibitor binding). The change in solvent accessibility of surface amino acids for both inhibitor and enzyme (in isolation and upon binding) was calculated by the SurfV algorithm [18]. Residues with a change in solvent accessibility were compiled into the ensemble of IFRs.

#### Interface Forming Residues Table

For each set of 70 complexes, we mapped all of the sequence positions that belonged to the IFR ensemble (table shown at Figure 1). We took the native trypsin-BPTI complex (1FY8) as a reference for sequence numbering because it follows standard serine protease sequence numbering. All residues showing a detectable loss of accessible area were labeled in the multiple structure-based sequence alignment obtained with PrISM according to the residue positions in the sequence of 1FY8 chain E. In the case of gap in the reference sequence, a letter was added to the previous number (designating the position of the particular amino acid in a given sequence), so that the label keeps following the reference sequence numbering. Columns without at least one residue labeled as an IFR were eliminated to facilitate the analysis. For the case of BPTI, 36 residues occupying IFR positions were compiled (table shown at Figure 2) and ranged from positions 37 to 228 (also see Figure 1). Corresponding numbers for the ecotine and ovomucoid third domain are 40 IFRs from 37 to 226 and 34 IFRs from 37 to 217A, respectively. The multiple alignment of IFRs for the 70 serine proteases complexed with BPTI is presented in the Figure 1 (the corresponding tabular data for ecotine and ovomucoid inhibitor are not shown here). The derivatives from the tables with 70 serine proteases bound to BPTI, ecotine, and ovomucoid third domain are presented in the tables shown at Figures 2, 3 and 4; there we have only those columns





which had some prevalent residues occupying that specific position. Particular differences among 4 principal classes of serine proteases were aimed. In other words, some positions were eliminated as they had a same residue present in each of the four sub-families of serine proteases (such as the position of HIS 57 which does not contribute to differentiation among 4 subfamilies). Those positions which were not identified as an IFR for the particular pair of serine protease and inhibitor are presented by a "-" sign. It is important to mention that the Asp\_102 is not seen in Figure 1 due to a very peculiar characteristic of the catalytic pocket (as explained previously in the "Background" session).

Once we constructed the 3 sets of 70 complexes of proteases with respective inhibitors, we made an inquiry into the relationships between structural characteristics and specificity of enzymes. We calculated the difference in residue type occupancy of the total enzyme surface and the respective interfaces. The goal was to find if the interface was different from the rest of the protein surface in terms of types of residues present. Using the SurfV algorithm, we calculated the total enzyme surface

area and determined the fraction occupied by each of four classes of amino acids. The same procedure was applied for the IFR area. Amino acid classes are as indicated: Charged, Polar, Hydrophobic and Glycine (a single member sub group of polar). The results of this inquiry are shown in Figure 5, as well as tabulated in the Table 2, and are discussed in more detail in results and discussion sections.

#### Amino acid classification and color coding

Before we can discuss our findings, it is necessary to explain the founding principles for amino acid classification (and their color coding) that we adopted for this work. In this paper we decided to pay special attention to three amino acid classes (hydrophobic, polar and charged) and also to treat, in a separate way, the three amino acids **G**, **C** and **H**. In the case of **Gly**, its occasionally given separate consideration because of its peculiarity: **Gly** has one hydrogen atom instead of the side chain (in a sense, **Gly** residues received an individual treatment here just as in the Ramachandran plot [32], for example). In case of the **Cys**, special treatment is given because it is the only residue that forms disulfide bridges. Lastly, **His** is treated separately because this residue is “neutral polar” at physiological pH but positively charged when protonated.

The amino acid color code is an adaptation of “amino-chromography” suggested by William Taylor [33]. Residues **AVLIMFP** are colored grey (small hydrophobic [**A** and **P**]) and also large hydrophobic [**VLIMF**]). Residues **STYNQWG** are colored green (polar - we are including **Y** and **W** to this ensemble although both residues have aromatic rings: phenol and indole, respectively). Those aromatic rings are often considered as another important feature for **Y** and **W** - the source of their hydrophobicity. The hydrophobic character of **W** and **Y** is often imposed above their polar characteristics. However, we consider **W** as an ambivalent residue (polar and hydrophobic) and **Y** as a polar residue. Residues **D** and **E** are colored red (negatively charged). Residues **R** and **K** are colored blue (positively charged). Residue **C** is colored yellow (disulphide bridge forming), and residue **H** [positively charged but neutral at neutral pH] is coloured cyan (light-blue).

The classification of amino acids follows the scheme given in “Structural Bioinformatics”, page 18 [34] and is supported by hydrophobicity scales by Janin [35], Kyte and Doolittle [36] and Rose [37]. It is known that some of the hydrophobicity tables cited here (and in other papers found in general literature) present an apparent contradiction with respect to classification of certain amino acids. For example **G** and **W** are often a subject of controversy in terms of being considered polar or hydrophobic. We needed to decide which option to use

in our paper, and the preceding description defines which one we opted for. It is clear therefore that an alternative classification is also possible although, we would consider it less precise.

#### Validation of Methodology

To validate our methodology for defining IFRs through “hard docking”, we selected 16 native complexes (not present in the “secondary datamart”) of serine proteases bound to BPTI—the inhibitor with the largest number of native complexes available in the PDB (Table 3 lists the PDB IDs of the selected complexes). We investigated how successful our methodology was at predicting IFR residues. Namely, by using the 1FY8 complex as a reference and transferring the inhibitor coordinates to the enzyme chains of the sixteen complexes, we compared real IFRs and ones obtained by rigid body docking. This was done by aligning the enzyme chains of the sixteen complexes using PyMOL and rigid body docking the inhibitor of 1FY8. The docking results were then compared to those reported for native bound inhibitors. Afterwards, we used the SurfV algorithm to determine the residues composing the interface (IFR) of the sixteen *in silico* complexes. The same algorithm was used to determine the IFRs in the respective native complexes. A comparison between rigid body docking and native IFR in natural complexes was then possible. From these results, a “confusion matrix” [38] was generated. Detailed analysis of corresponding data is given in results and discussion sections.

#### Acknowledgements

We would like to sincerely thank José Miguel Ortega (Biodados-UFMG) for critically reviewing the manuscript and his indispensable comments and suggestions. This work was developed as part of CR PhD thesis, which has been sponsored by the Brazilian Ministry of Education (CAPES) and Foundation for Research Support of Minas Gerais State (FAPEMIG). We would also like to express our appreciation to Marko Spasic from University of California, Berkeley for the careful reading and editing of this manuscript.

#### Author details

<sup>1</sup>Department of Biochemistry and Immunology, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil. <sup>2</sup>Embrapa Genetic Resources and Biotechnology, Brasília, DF, Brazil. <sup>3</sup>Embrapa Information Technologies, Campinas, SP, Brazil. <sup>4</sup>Department of Mathematics and Computer Science - Federal University of Itajubá, Itajubá, MG, Brazil. <sup>5</sup>Department of Computer Science, Institute of Exact Sciences, Federal University of Minas Gerais, Belo Horizonte, Brazil.

#### Authors' contributions

CR: DB and datamarts preparation, structural alignment, PyMOL alignment, IFR tables, figures, manuscript. RT: DB preparation, structural alignment, IFR tables, figures, manuscript, PRISM maintenance. IM: DB maintenance. AM: DB maintenance. RMM: Experiment validation. CH: Experimental validation. IAPN: calculating values for the tables and figures of frequency of occurrence for surface and interface residues. JGJ: project coordination. MS: project coordination and oriented this work. GN: project supervisor, elaboration of techniques for hard docking and IFR assemblies. All authors read, revised and approved the final manuscript.

Received: 8 June 2010 Accepted: 20 October 2010  
Published: 20 October 2010

## References

1. Neurath H: **Proteolytic processing and physiological regulation.** *Trends Biochem Sci* 1989, **14**:268-271.
2. Hedstrom L: **An overview of serine proteases.** *Curr Protoc Protein Sci* 2002, **Chapter 21, Unit 21 10.**
3. Hedstrom L: **Serine protease mechanism and specificity.** *Chem Rev* 2002, **102**:4501-4524.
4. Perona JJ, Craik CS: **Structural basis of substrate specificity in the serine proteases.** *Protein Sci* 1995, **4**:337-360.
5. Bartlett GJ, Porter CT, Borkakoti N, Thornton JM: **Analysis of catalytic residues in enzyme active sites.** *J Mol Biol* 2002, **324**:105-121.
6. Greer J: **Comparative modeling methods: application to the family of the mammalian serine proteases.** *Proteins* 1990, **7**:317-334.
7. Ruhlmann A, Kukla D, Schwager P, Bartels K, Huber R: **Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. Crystal structure determination and stereochemistry of the contact region.** *J Mol Biol* 1973, **77**:417-436.
8. Read RJ, James MN: **Refined crystal structure of *Streptomyces griseus* trypsin at 1.7 Å resolution.** *J Mol Biol* 1988, **200**:523-551.
9. Matthews BW, Sigler PB, Henderson R, Blow DM: **Three-dimensional structure of tosyl-alpha-chymotrypsin.** *Nature* 1967, **214**:652-656.
10. Bode W, Mayr I, Baumann U, Huber R, Stone SR, Hofsteenge J: **The refined 1.9 Å crystal structure of human alpha-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Trp insertion segment.** *Embo J* 1989, **8**:3467-3475.
11. Schechter I, Berger A: **On the size of the active site in proteases. I. Papain.** *Biochem Biophys Res Commun* 1967, **27**:157-162.
12. Neshich G, Mazoni I, Oliveira SR, Yamagishi ME, Kuser-Falcao PR, Borro LC, Morita DU, Souza KR, Almeida GV, Rodrigues DN, *et al*: **The Star STING server: a multiplatform environment for protein structure analysis.** *Genet Mol Res* 2006, **5**:717-722.
13. Neshich G, Borro LC, Higa RH, Kuser PR, Yamagishi ME, Franco EH, Krauchenco JN, Fileto R, Ribeiro AA, Bezerra GB, *et al*: **The Diamond STING server.** *Nucleic Acids Res* 2005, **33**:W29-35.
14. Neshich G, Mancini AL, Yamagishi ME, Kuser PR, Fileto R, Pinto IP, Palandrani JF, Krauchenco JN, Baudet C, Montagner AJ, Higa RH: **STING Report: convenient web-based application for graphic and tabular presentations of protein sequence, structure and function descriptors from the STING database.** *Nucleic Acids Res* 2005, **33**:D269-274.
15. Neshich G, Togawa RC, Mancini AL, Kuser PR, Yamagishi ME, Pappas G Jr, Torres WW, Fonseca e Campos T, Ferreira LL, Luna FM, *et al*: **STING Millennium: A web-based suite of programs for comprehensive and simultaneous analysis of protein structure and sequence.** *Nucleic Acids Res* 2003, **31**:3386-3392.
16. Da Silva MC, de Sa MF, Chrispeels MJ, Togawa RC, Neshich G: **Analysis of structural and physico-chemical parameters involved in the specificity of binding between alpha-amylases and their inhibitors.** *Protein Eng* 2000, **13**:167-177.
17. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The Protein Data Bank.** *Nucleic Acids Res* 2000, **28**:235-242.
18. Sridharan S, Nicholls A, Honig B: **A new vertex algorithm to calculate solvent accessible surface areas.** *Biophys J* 1992, **61**:A174.
19. Page MJ, Cera ED: **Is it possible to transform trypsin to thrombin?** *IUBMB Life* 2007, **59**:413-414.
20. Laskowski M, Ardeit W: **Turkey Ovomuroid Third Domain Inhibits Eight Different Serine Proteinases of Varied Specificity on the Same ..Leu18-Glu19...Reactive Site.** *Biochemistry* 1985, **24**:5313-30.
21. Pál G, Sprengel G, Patthy A, Gráf L: **Aleation of the specificity of ecotine, an *E coli* serine proteinase inhibitor, by site directed mutagenesis.** *FEBS Letters* 1994, **342**:57-60.
22. Janin J, Bahadur RP, Chakrabarti P: **Protein-protein interaction and quaternary structure.** *Q Rev of Biophys* 2008, **41**(2):133-180.
23. Laskowski M, Qasim MA: **What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes?** *Biochim Biophys Acta* 2000, **1477**(1-2):324-37.
24. da Silveira CH, Pires DE, Minardi RC, Ribeiro C, Veloso CJ, Lopes JC, Meira W Jr, Neshich G, Ramos CH, Habesch R, Santoro MM: **Protein cutoff scanning: A comparative analysis of cutoff dependent and cutoff free methods for prospecting contacts in proteins.** *Proteins* 2009, **74**:727-743.
25. Bone R, Silen JL, Agard DA: **Structural plasticity broadens the specificity of an engineered protease.** *Nature* 1989, **339**:191-195.
26. Ma W, Tang C, Lai L: **Specificity of trypsin and chymotrypsin: loop-motion-controlled dynamic correlation as a determinant.** *Biophys J* 2005, **89**:1183-1193.
27. Andreeva A, Howorth D, Chandonia JM, Brenner SE, Hubbard TJ, Chothia C, Murzin AG: **Data growth and its impact on the SCOP database: new developments.** *Nucleic Acids Res* 2008, **36**:D419-425.
28. Laskowski RA, Chistyakov VV, Thornton JM: **PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids.** *Nucleic Acids Res* 2005, **33**:D266-268.
29. Barthelmes J, Ebeling C, Chang A, Schomburg I, Schomburg D: **BRENDA, AMENDA and FRENDA: the enzyme information system in 2007.** *Nucleic Acids Res* 2007, **35**:D511-514.
30. Yang AS, Honig B: **Sequence to structure alignment in comparative modeling using PRISM.** *Proteins* 1999, **3**: 66-72.
31. Delano WL: **The PyMOL Molecular Graphics System.** Delano Scientific, San Carlos, CA, USA 2002 [<http://www.pymol.org>].
32. Ramachandran GN, Ramakrishnan C, Sasisekharan V: **Stereochemistry of polypeptide chain conformations.** *J Mol Biol* 1963, **7**:95-99.
33. Taylor WR: **Residual colours: a proposal for aminochromography.** *Protein Eng* 1997, **10**:743-746.
34. Philip E: **Bourne and Helge Weissig.** Structural Bioinformatics, Wiley-Liss 2003.
35. Janin J: **Surface and Inside Volumes in Globular Proteins.** *Nature* 1979, **277**:491-492.
36. Kyte J, Doolittle R: **A Simple Method for Displaying the Hydropathic Character of a Protein.** *J Mol Biol* 1982, **157**:105-132.
37. Rose G, Geselowitz A, Lesser G, Lee R, Zehfus M: **Hydrophobicity of Amino Acid Residues in Globular Proteins.** *Science* 1985, **229**:834-838.
38. Kohavi R, Provost F: **Glossary of Terms.** *Machine Learning* 1998, **30**:271-274.

doi:10.1186/1472-6807-10-36

**Cite this article as:** Ribeiro *et al.*: Analysis of binding properties and specificity through identification of the interface forming residues (IFR) for serine proteases *in silico* docked to different inhibitors. *BMC Structural Biology* 2010 **10**:36.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

