

***STRUCTURAL BACKGROUND OF THE S1 SITE SUBSTRATE  
SPECIFICITY IN PANCREATIC SERINE PROTEINASES***

Ph.D. thesis

Balázs Jelinek

**Doctoral School of Biology**

Head of the Doctoral School: Prof. Anna Erdei CMHAS

**Structural Biochemistry Doctoral Programme**

Director of Programme: Prof. László Gráf MHAS

**Supervisor**

Prof. László Gráf MHAS



Eötvös Loránd University  
Department of Biochemistry  
Budapest, 2007

# Contents

Abbreviations .....	3
1. Introduction .....	4
1.1 Classification of proteinases.....	4
1.2 Evolution of serine proteinases .....	5
1.3 Functions, structure and catalytic mechanism of the SA clan proteinases.....	6
1.4 Mechanism of the proteolytic cleavage.....	7
1.5 Activation mechanism of pancreatic serine proteinases .....	9
1.6 Specificity of pancreatic serine proteinases .....	11
1.7 The S1 site specificity of pancreatic serine proteinases.....	12
1.8 Proteases of the S1 family with trypsin-like specificity.....	16
1.9 Site specific mutagenesis studies investigating the structural basis of S1 site specificity .....	18
1.9.1 Attempts to convert trypsin to chymotrypsin.....	18
1.9.2 Attempts to convert chymotrypsin to trypsin.....	20
1.9.3 Attempts to convert trypsin to elastase .....	22
1.9.4 Attempts to convert chymotrypsin to elastase.....	22
2. Aims of the study .....	23
3. Materials and Methods .....	25
3.1 Materials.....	25
3.2 Construction of mutants .....	25
3.3 Expression and purification of the mutants.....	26
3.4 Enzyme assays.....	27
3.4.1 Introduction .....	27
3.4.2 Description of the experiments .....	29
3.5 Specificity profiling.....	31
3.6 Crystallization, structure solution and refinement .....	32
3.6.1 Introduction .....	32
3.6.2 Description of the experiments .....	33
3.7 Structure analysis .....	34
4. Results .....	36
4.1 Kinetic results.....	36
4.2 The S189D+A226G mutant structure .....	41
4.2.1 Introduction .....	41
4.2.2 The overall structure.....	41
4.2.3 The S1 specificity site .....	44
4.2.4 The activation domain.....	47
4.3 pH dependence experiments.....	49
5. Discussion .....	53
6. Summary .....	59
7. Összefoglalás.....	61
8. Acknowledgements .....	64
References .....	65

## Abbreviations

AMC	Amino-4-Methyl Coumarin
BPTI	bovine pancreatic trypsin inhibitor
CABS	4-(Cyclohexylamino)-1-butan sulfonic acid
CCP4	Collaborative Computational Project Number 4
CHES	2-(Cyclohexylamino)ethanesulfonic acid
DPI	dispersion precision indicator
EDTA	ethylenediaminetetraacetic acid
ESRF	European Synchrotron Radiation Facility
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
MES	2-(N-Morpholino)ethanesulfonic acid
MOPS	3-(N-Morpholino)propanesulfonic acid
MUGB	4-Guanidinobenzoic acid 4-methylumbelliferyl ester hydrochloride
PDB	protein data bank
PEG	Polyethylene glycol
SBTI	soybean trypsin inhibitor
SDS-PAGE	Sodium dodecyl sulfate – polyacrylamide gel electrophoresis
TLS	Translational, librational, screw
TRIS	Tris(hydroxymethyl)aminomethane

# 1. Introduction

Enzymes are macromolecules that catalyze chemical reactions. The essence of catalysis is the stabilization of the highest energy state of the reaction pathway, the transition state. Enzymes can bind their substrates in an orientation that is optimal for the reaction, and can also play a key role as a reaction partner, opening new reaction pathways. Enzymes can increase the reaction rate by numerous orders of magnitude by providing the features necessary for surmounting the activation energy barrier between the ground and transition states and, by connecting different reaction pathways, they can allow the occurrence of reactions that would not occur spontaneously at the given conditions. In the absence of enzymes, most reactions of biological systems do not occur at perceptible rates or do not occur at all, therefore their contribution is indispensable in almost all biological processes.

Another fundamental function of enzymes is to provide selectivity among the numerous different molecules generally present in biological systems. This specificity derives from the selective binding of the transition state of substrates that participate in the reaction catalyzed by the enzyme. Substrate binding is the result of a complex interplay between the enzyme and the substrate including hydrogen bonds, van der Waals and electrostatic interactions, steric hindrances and the hydrophobic effect. The specific combination of these factors defines the enzyme's substrate specificity.

My doctoral work focused on the investigation of the structural background of substrate specificity in the pancreatic serine proteinases trypsin and chymotrypsin by using the approaches of site directed mutagenesis, enzyme kinetics and protein crystallography. In the present study I first give a general description of the enzymes investigated, followed by an overview of this specific field and the aims of the study; then I present the materials and methods we used, the results we obtained and their interpretation.

## ***1.1 Classification of proteinases***

In biological systems, the hydrolysis reaction catalyzed by proteinases (enzymes also called proteases, peptidases) comprises the cleavage of peptide bonds connecting the amino

acid building blocks of peptides and proteins<sup>1</sup>. According to the position of the cleaved peptide bond, proteinases can be classified as exopeptidases that cleave the bond at the N-terminus (aminopeptidases) or at the C-terminus (carboxypeptidases) of the polypeptide chain; and endopeptidases that cleave internal bonds of the chain. The hydrolysis of the substrate is facilitated by a reactive group of the enzyme, and the classification based on the nature of this group includes serine-, threonine-, cysteine-, aspartate- glutamate- and metalloproteinases. Barrett and Rawlings (1995) introduced a further classification step that is based on amino acid sequence and tertiary structure homologies. It consists of subdividing proteinases into clans and families, where a member of a given family must show significant sequence homology with at least one other member of the family, and clans are composed of families related by structural homologies. According to the classification criteria above, the group of enzymes I focused on named pancreatic serine proteinases are serine endopeptidases that belong to the S1 family of the SA clan.

## ***1.2 Evolution of serine proteinases***

Enzymes that have evolved from a common ancestor show significant structural homologies and, depending on the extent of evolutionary divergence, might also show significant sequence homology. The overall fold of enzymes is conservative because it determines the spatial position of the also conserved catalytic residues essential to provide reaction rate enhancement. Other residues are subject to evolutionary divergence to varying extent, which can result in the appearance of novel specificities. One third of all proteinases, more than 40 enzyme families use a serine nucleophile for catalysis. These families can be classified into 7 clans according to their overall fold and to other key positions of their catalytic machinery. In the SA clan, this machinery is composed of a serine, a histidine and an aspartate, together called the catalytic triad (Huber and Bode, 1978). The spatial positions of these residues are strictly conserved throughout the members of the clan. Interestingly, similar catalytic triads can be found in two other clans of serine proteinases, SB and SC, embedded in a different overall fold, showing that the triad appeared several times independently via convergent evolution. Other parts of the catalytic machinery that stabilize the transition state also show significant similarities.

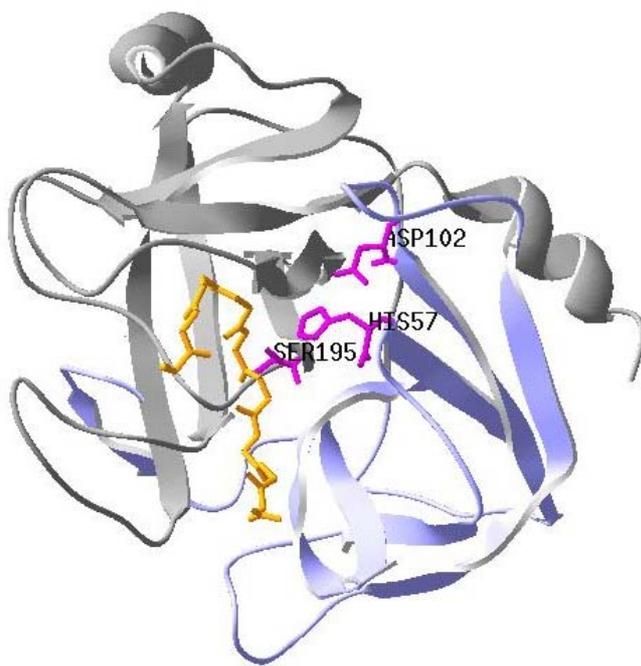
---

<sup>1</sup> Proteinases can also be assayed for esterase activity on synthetic ester substrates

### 1.3 Functions, structure and catalytic mechanism of the SA clan proteinases

The overall structural and functional concept of the SA clan is a generally widespread design. It can be found in eukaryotes, prokaryotes, archae and viruses. These enzymes play key roles in numerous physiological processes of the immune system, the blood coagulation cascade, the digestive system and various intra- and extracellular regulation mechanism. Besides the proteinase domain, many of them contain additional domains involved in regulation.

In my study I focused on the pancreatic serine proteinases trypsin, chymotrypsin and elastase that are digestive enzymes of the S1 family secreted by the pancreas. Most proteinases of the SA clan belong to the S1 family also named the chymotrypsin or the trypsin-like proteinase family, for its first discovered members. The term “enzyme” was first applied to pancreatic serine proteinases more than hundred years ago (Kuhne, 1877), and

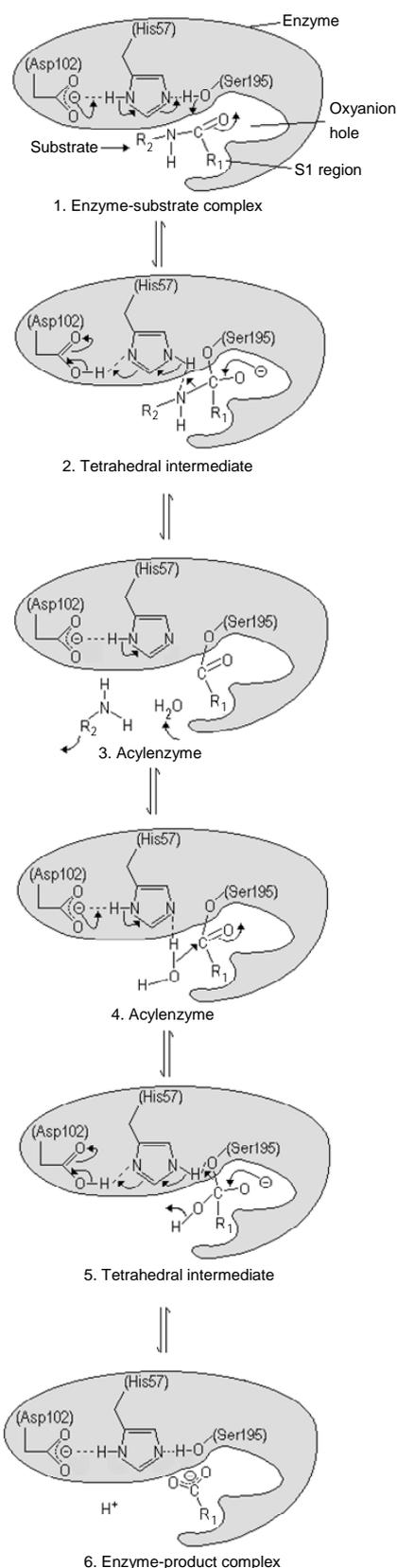


**Fig. 1** The crystal structure of trypsin

The first domain is showed in blue, the second in gray, residues of the catalytic triad in magenta and part of the backbone of the substrate indicating the binding site in orange (PDB code: 1EJM). Images that present protein structures throughout this study were generated using the softwares DeepView 3.7 (Guex and Peitsch, 1997) and PyMol (Delano, 2002).

trypsin and chymotrypsin were among the first enzymes isolated in pure form (Northrop *et al.*, 1947). Over 20 unique crystal structures of the members of this family have been determined and more than 500 sequences are available (Perona and Craik, 1997).

Here I present the general structural and functional description of trypsin, a prototype of the S1 family (**Figure 1**). The sequence numbering in the literature of the S1 family generally refers to the chymotrypsin sequence numbering, so that functional groups and regions of these enzymes can be easily identified and compared. The structure of trypsin shows a 25 kDa globular protein arranged into two  $\beta$ -barrel domains, each composed of six antiparallel  $\beta$ -strands. The second



**Fig. 2 The mechanism of the proteolytic cleavage**  
See text for details

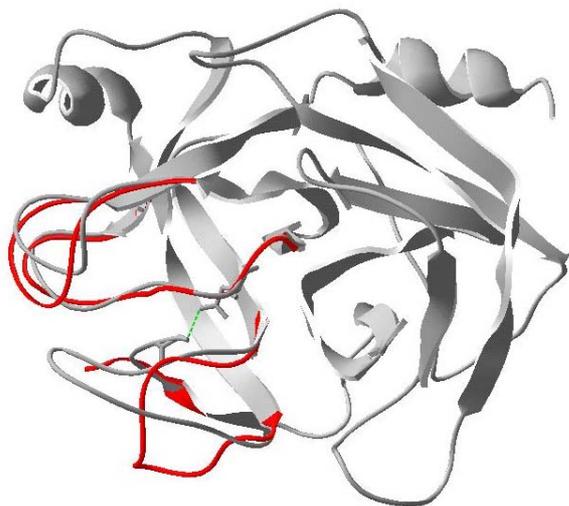
domain also contains two  $\alpha$ -helices. This structure contains six disulfide bridges, but their number can vary between different members of the family. Two straps connect the two barrels: the N-terminal region and the C-terminus form noncovalent interactions with the surface of the other barrel. The active site of the enzyme is located in the cleft between the two barrel domains. The catalytic triad is composed of amino acids from both barrels: His57 and Asp102 from barrel-1 and Ser195 from barrel-2. The substrate binding subsites are mostly composed of the surface loops surrounding Ser195 and connecting the antiparallel  $\beta$ -strands. Ser195 is called the active serine because as a nucleophilic reagent it attacks the scissile bond, with the help of His57 as a base and Asp102 that stabilizes the imidazole ring of His57 in the optimal conformation. An acidic pocket called the oxyanion hole is another important part of the active site, composed of the main chain amides of Ser195 and Gly193.

#### 1.4 Mechanism of the proteolytic cleavage

The standard free energy for the hydrolysis reaction of a Gly-Gly peptide bond is  $\Delta G^0 = -2.3$  kcal/mol at neutral pH, therefore it can occur spontaneously. However, its half-life is around seven years (Kahne and Still, 1988), due to the high stability of the conjugated peptide bond. Serine proteinases can weaken this bond by activating it with two H-bond interactions formed between their acidic oxyanion hole and the carbonyl oxygen of the conjugated bond, as the Michaelis complex is formed by the substrate bound on the surface of the enzyme (**Figure 2, step 1**). As a consequence, the activation energy for the nucleophilic attack by the active serine of the enzyme towards the carbonyl group of the substrate becomes lower.



## 1.5 Activation mechanism of pancreatic serine proteinases



**Fig. 3** *The rearrangement of the activation domain of trypsin upon activation*

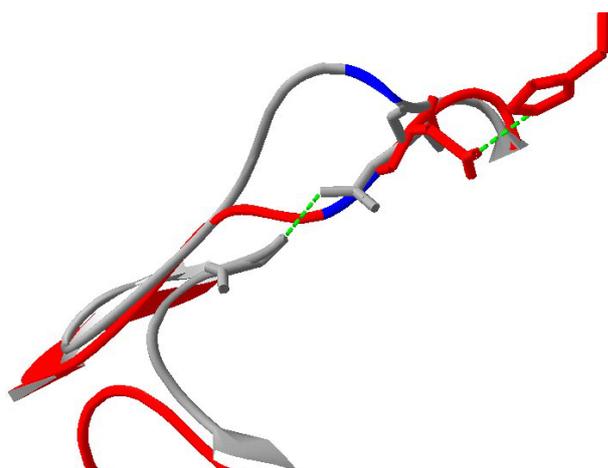
*The structure of trypsinogen (PDB code: 1TGN) is fitted over that of trypsin (1TPO). Segments composing the activation domain of trypsinogen are shown in red. The electrostatic interaction between the N-terminus and Asp194 in trypsin is shown with a green dotted line.*

Trypsin, chymotrypsin and elastase are expressed as inactive zymogens by the acinar cells of the pancreas. When excreted to the duodenum, they become active by a proteolytic cleavage at their N-terminus, resulting in a cleaved propeptide and a new N-terminus at residue 16. The propeptide of trypsin diffuses away, while the propeptide of chymotrypsin remains anchored to the enzyme by a disulfide bridge. Trypsinogen is activated by enterokinase produced by the duodenum. It can also activate itself in a process called autoactivation; chymotrypsinogen and proelastase are activated exclusively by trypsin. The positively charged  $\alpha$ -amino group of the new N-terminus turns inward to the protein, and forms a salt bridge with Asp194. The formation of this salt bridge

induces a conformational rearrangement restricted to a well defined part of the enzyme (about 15% of its residues) that is considered as an autonomous folding unit called the activation domain (Huber and Bode, 1978) (**Figure 3**). This part of the enzyme is composed of the new N-terminus (16-19), the autolysis loop (142-152) and two loops that form the substrate binding site near the scissile bond, called L1 (185-195) and L2 (217-223). Apart from autoactivation, trypsin shows another exclusive feature at this region: it is markedly disordered in trypsinogen, and becomes ordered only upon activation; while chymotrypsinogen and proelastase show a more ordered activation domain.

The rearrangements upon activation include an important conformational change that, among other structural differences of the substrate binding sites, explains the inactivity of the zymogen forms. Asp194 of the zymogen forms an electrostatic interaction with His40 and its

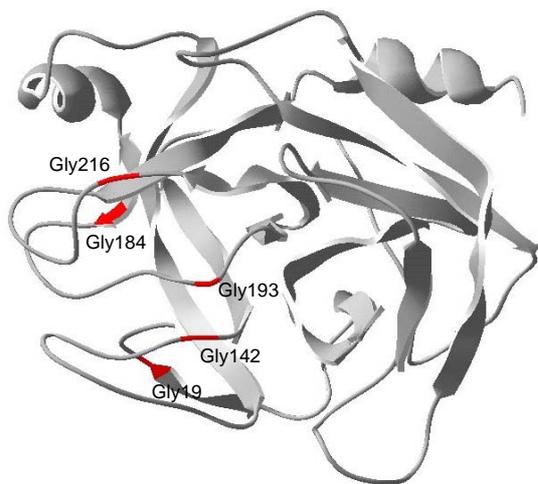
conformation does not allow the formation of the oxyanion hole that is composed of the main chain amides of the neighboring residues Gly193 and Ser195 (**Figure 4**). Following activation, the new N-terminus of the enzyme competes for the negative charge of Asp194 and forces it into a new, 180° tilted conformation, and the formation of this new salt bridge induces also the rearrangement of Gly193 and thereby the formation of the oxyanion hole indispensable for catalysis. The salt bridge stabilizes the active form of the enzyme, and its perturbation by raising the pH from the optimal 8 to 11 leads to inactivation, while lowering the pH back to 8 induces a conformational transition similar to activation (Fersht, 1972; Fersht and Renard, 1974; Stoesz and Lumry, 1978; Heremans and Heremans, 1989). This transition can be followed by detecting changes in tryptophan fluorescence of the enzyme (Verheyden *et al.*, 2004). In chymotrypsin, there is an equilibrium between the active and the zymogen forms, and even at the pH optimum, the proportion of active enzyme is only about 90% of the total amount of enzyme present in the solution, whereas at pH values as high as 12, 12% of the molecules remain active (Fersht, 1972).



**Fig. 4 The formation of the new salt bridge and the oxyanion hole in chymotrypsin**

*Asp194 of the zymogen (PDB code: 1CHG, shown in red) forms an electrostatic interaction (shown in green) with the imidazole ring of His40. Activation induces a new conformation of Asp194 (PDB code: 4CHA, shown in gray) by forming a salt bridge with the new N-terminus, and also the formation of the oxyanion hole. The position of Gly193 composing the oxyanion hole is shown in blue.*

In a targeted molecular dynamics simulation study of the trypsinogen-trypsin transition (Brünger *et al.*, 1987), some dedicated glycine residues of the activation domain showed exceptionally high changes in their  $\Phi$  and/or  $\Psi$  angle values, when compared to other parts of the activation domain loops. This suggested that conformational changes of the activation domain occur around flexible hinges formed of conserved glycines at positions 19, 142, 184, 193 and 216, where the last four positions connect the loops with the fixed  $\beta$ -sheets (**Figure 5**). Our recent study demonstrated that indeed, this setup can facilitate the activation process by decreasing the internal friction “cost” of the conformational rearrangement (Tóth. *et al.*, 2006).

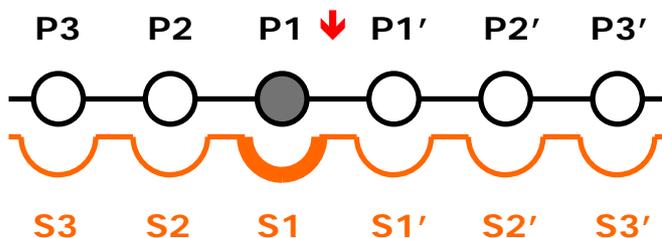


**Fig. 5 Hinges in the activation domain of trypsin**

The positions of glycines 19, 142, 184, 193 and 216 are shown in red. (PDB code: 1TPO)

## 1.6 Specificity of pancreatic serine proteinases

The various roles of the proteinases of the SA clan mentioned above suggest a diverse substrate specificity palette for these enzymes. Indeed, most of their substrate specificity determinants are composed of the surface loops that allowed rapid and varied evolutionary divergence along with the conservation of the overall enzyme fold (Perona and Craik, 1997). The stabilization and accurate positioning of the scissile bond with respect to Ser195 and the oxyanion hole are crucial for the formation of the tetrahedral intermediate. This task is accomplished by the ensemble of numerous enzyme-substrate interactions formed on both sides of the scissile bond (**Figure 6**). According to the nomenclature of Schechter and Berger (1967), the scissile bond is located between substrate residues P1-P1', and the substrate binding Sn-Sn' subsites on the surface of the enzyme are numbered according to the substrate residue they accommodate.



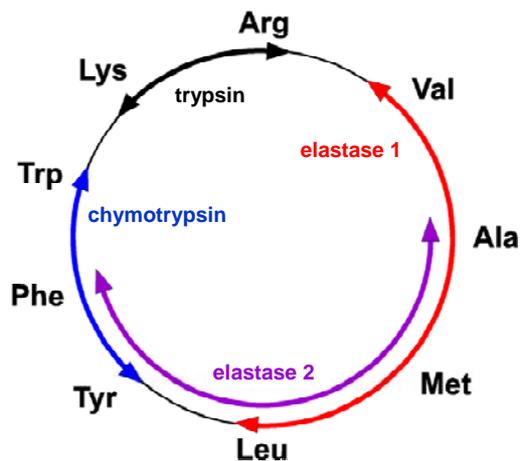
**Fig. 6 A schematic view of enzyme-substrate interactions in pancreatic serine proteinases**

The substrate is shown in black, residues P3-P3' are presented. The enzyme is shown in orange, S3-S3' denote substrate binding subsites. The scissile bond is indicated by a red arrow.

The S1 region of the enzyme includes the substrate binding pocket, a cavity formed by residues 185-195, 213-223 and 226-228. This region is the primary binding site of pancreatic

serine proteinases, its contribution to specific activity is of 3-4 orders of magnitude larger than that of other enzyme subsites. An exception is elastase, where other interactions, especially the one between P4 and S4 play also important roles (Thompson and Blout, 1973).

### 1.7 The S1 site specificity of pancreatic serine proteinases

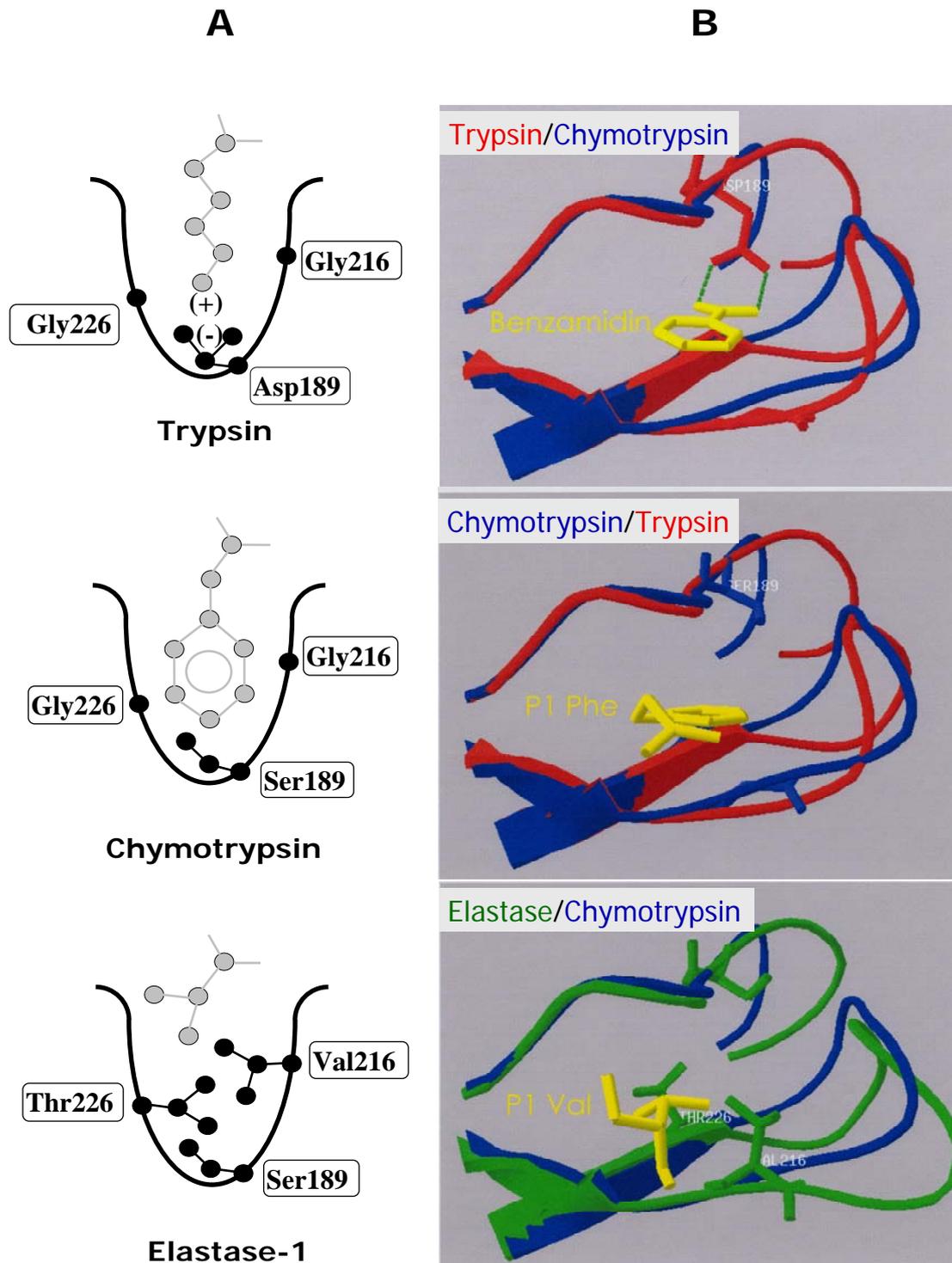


**Fig. 7** A schematic view of the S1 site specificities of pancreatic serine proteinases

The substrate binding pockets of pancreatic serine proteases evolved to accommodate different P1 substrate residues (**Figure 7**), with a specific and thereby more accurate scissile bond positioning than that of an aspecific ancestor. Trypsin, with the negative charge of Asp189 at the base of the S1 site, prefers Arg and Lys in the P1 substrate position, while the less polar S1 site of chymotrypsin with Ser189 has the highest affinity for Tyr, Phe and Trp substrate residues. The shallow pocket of elastase-1 is most suitable for Ala, Leu and Val because of the occlusion at the entrance of the pocket by Val216

and Thr226. Specificity of elastase-2 is somewhere between that of elastase-1 and chymotrypsin, with Gly216 and Ser226 at the pocket entrance (**Figure 8A**). Trypsins from different species always contain glycines at sites 216 and 226, while chymotrypsin has a Gly at site 216 and either a Gly or an Ala at site 226.

Early crystallographic studies demonstrated that the S1 pockets of these enzymes are highly complementary to their preferred P1 side chains. These studies suggested that residues 189, 216 and 226 that directly contact the substrate fully determine substrate specificity in pancreatic serine proteases (Steitz *et al.*, 1969; Watson and Shotton, 1970). This became the standard textbook model for substrate specificity, but showed to be incomplete when tested with a series of site specific mutagenesis studies described in the next chapter. Indeed, besides some striking structural similarities like the conformation of the 191-220 cystine connecting loops L1 and L2, the comparison of the sequence (**Figure 10**), the structure (**Figure 8B**) and the H-bond network (**Figure 9**) of the S1 region shows a specific composition arranged in a distinct conformation and interaction network for each type of S1 specificity.



**Fig. 8 A)** A schematic view of the S1 site of pancreatic serine proteinases. Residues that determine S1 site specificity according to Steitz, Henderson and Blow (1969), and Shotton and Watson (1970) are shown in black, P1 residues are shown in gray. **B)** Structure comparisons of the S1 sites, showing a view from the entrance of the pocket. The denoted structures were superimposed and colored as indicated. Main chain ribbons are presented. The specificity determinants according to the simple model and the P1 residues of inhibitor molecules (yellow) are shown as sticks. The trypsin structure is shown in complex with the small synthetic inhibitor benzamidine, green dotted lines show the electrostatic interaction.

The H-bond network has an important role in stabilizing the S1 region since those parts where the three S1 regions show large structural differences do not have contacts with other parts of the enzyme. A negative charge at the base of the pocket is essential for trypsin to exhibit substrate-specific activity (Evnin *et al.*, 1990). **Figure 9** shows that in trypsin, Asp189 is in the center of the H-bond network, forming cross-linking interactions with loops L1 and L2. This specific interaction system plays a key role in the stability of this region by dissipating the almost buried charge of Asp189. When its counterpart, the positively charged P1 residue of the substrate is not present, structural water molecules contribute to the dissipation of the negative charge. Apart from Asp189 that determines trypsin-like specificity, two other buried negative charges play key roles in catalysis (Asp102) and activation (Asp194) of pancreatic serine proteinases, as described above.

**Fig. 9** A schematic view of the hydrogen-bond network that stabilizes the S1 region in serine proteinases

Residues that compose the S1 pocket are shown with a shaded background. Residues of loops L1 and L2 are shown in black frames. S denotes sulfur, O is for oxygen, OG and OD denote  $\gamma$ - and  $\delta$ -oxygens, N is for nitrogen and W denotes water. Possible H-bonds are shown as arrows. (from Szabó *et al.*, 2003)

### Trypsin

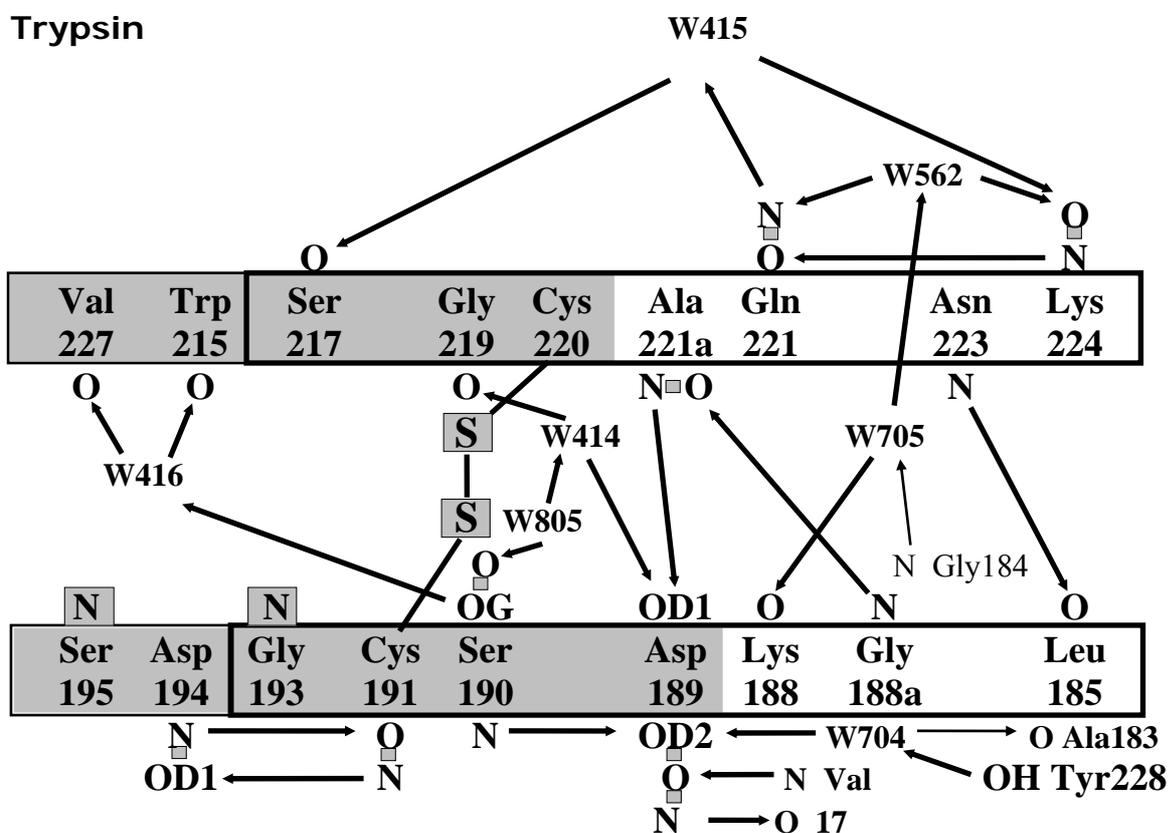
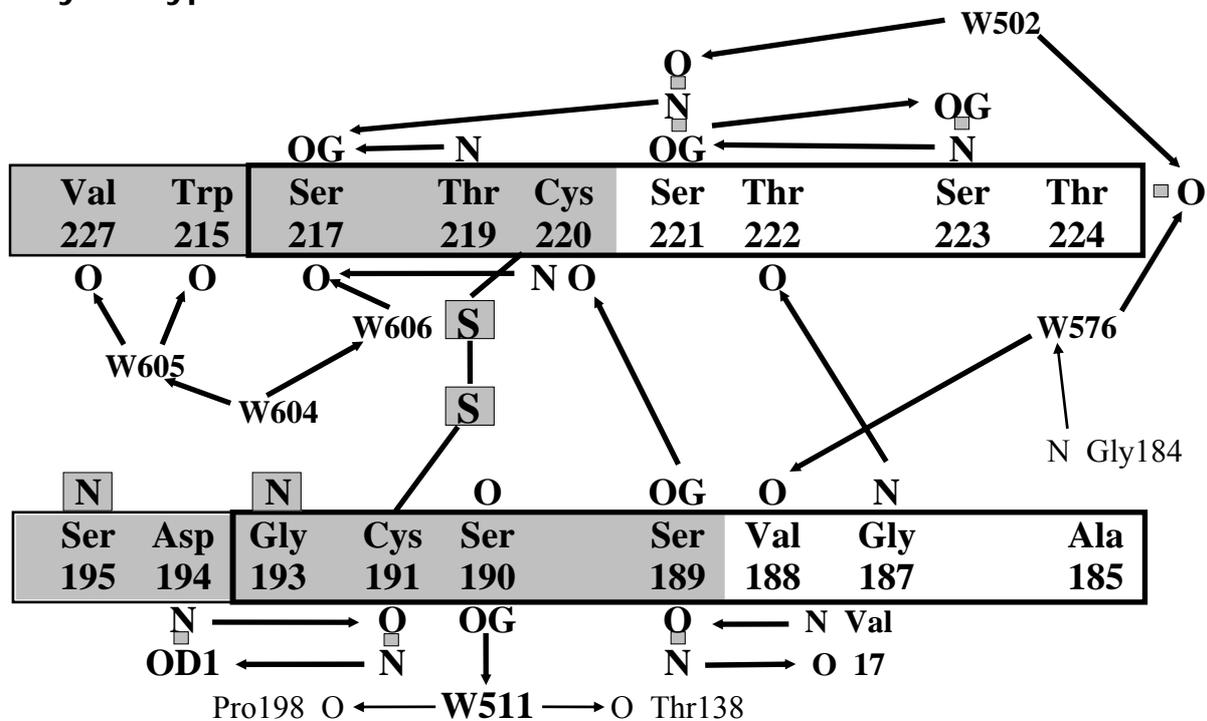
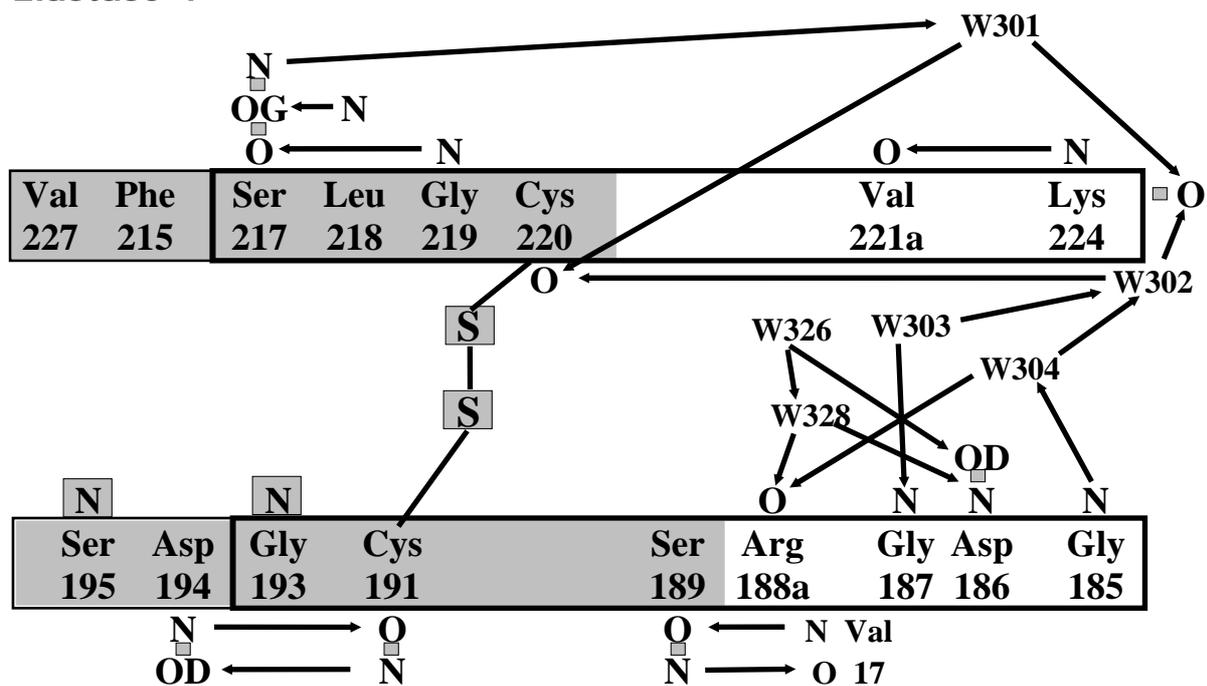


Fig. 9, continued

**Chymotrypsin**

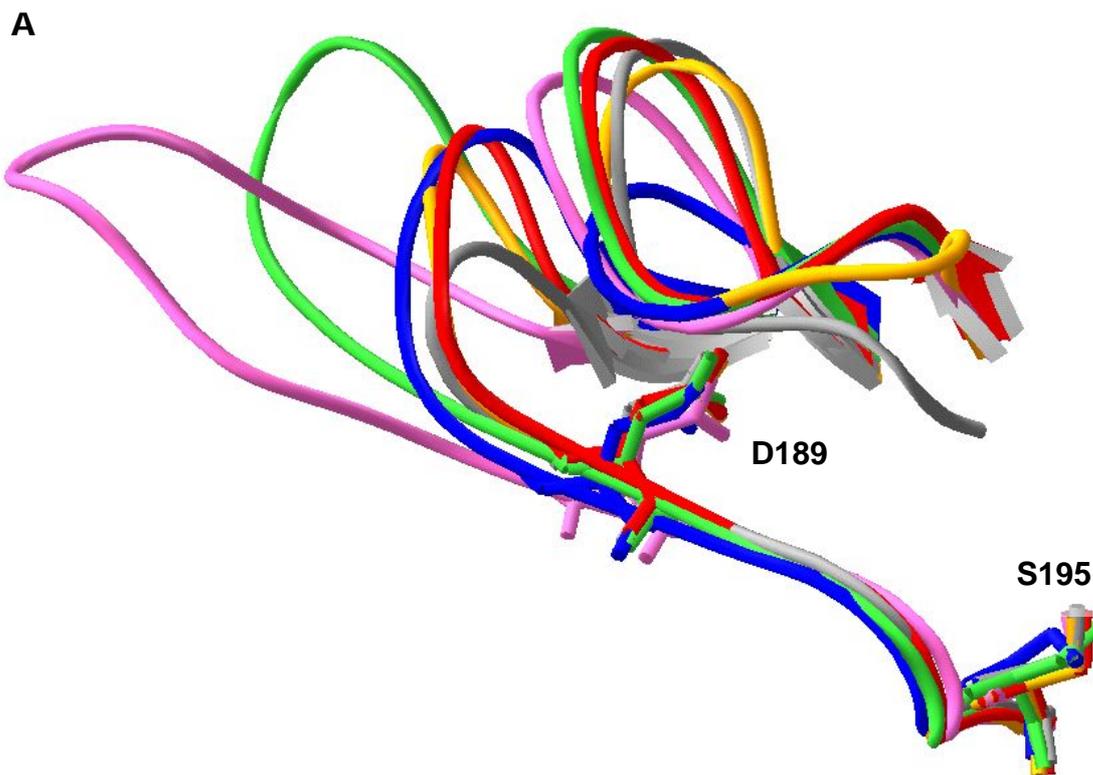


**Elastase-1**



## **1.8 Proteases of the S1 family with trypsin-like specificity**

In the S1 family of serine proteinases, the majority of enzymes have a trypsin-like S1 specificity, with an Asp189 at the base of the pocket as a counterpart of the positively charged P1 residue. The various roles of these enzymes often require very high specificity toward their natural substrates, and these interactions are usually formed by numerous factors such as additional domains responsible for specificity. Nevertheless, at the level of the S1-P1 interaction, the evolutionary concept of specificity for the positively charged residues Arg and Lys might be ideal because these residues are large, bear the distinguishing mark of the positive charge and mostly stand alone on the surface in their hydrate shell since salt bridges formed on the surface of proteins are generally not energetically favorable. In trypsin, P1 Arg is bound directly in a salt bridge to Asp189, while the binding of the smaller P1 Lys involves a water molecule. Comparison of the structure of the S1 site in these enzymes (**Fig. 10 A**) shows a practically identical position of the Asp189 side chain in each of them, with respect to the catalytic triad. The conformation of the main chain segment 189-195 that comprises also the oxyanion hole and the 191-220 cystine is also similar, while the rest of the S1 site, composed by loops L1 and L2 shows striking differences in some cases. A detailed comparison of the structural elements that stabilize the charge of Asp189 shows that different parts of the S1 site can complete this task, although they must occupy a similar steric positions. This comparison demonstrates that S1 sites with different sequence (**Fig. 10 B**), conformation and a different interaction system can stabilize the charge of Asp189 in the required position.



**B**

1EJM	GYLEG-----GKDSCQGDSSGGPVVCSG-----KLQGIVSWGSG--CAQKNKPG	224
1LTO	GNT-----RRDSCCKGDSGGPLVCK--VNGTWLQAGVVSWEDEG--CAQPNRPG	255
1OP8	GSLRG-----GRDSCNGDSSGPLLCEG-----VFRGVTSFGLNKC GDPRGPG	241
1MD8	GHP-----SLKQDACQGDSSGGVFAVRDPNTDRWVATGIVSWGIG--CSRG--YG	221
1THS	GYKPD---EGKRGDACEGDSGGPFVMKSPFNRRWYQMGIVSWGEG--CDRDGKYG	238
1BDA	GDTRSGGPQANLHDACQGDSSGGPLVCLN--DGRMTLVGIIISWGLG--CGQKDVPG	234
	.....	.....

**Fig. 10 A: Structural comparison of the substrate binding pocket of trypsin-like proteinases of the S1 family**

The crystallographic structures of the following trypsin-like enzymes were superimposed:

1. Bovine trypsin, red (PDB code: 1EJM)
2. Human alpha-1 tryptase, gray (PDB code: 1LTO)
3. Human granzyme A, orange (PDB code: 1OP8)
4. Human complement protease C1r, blue (PDB code: 1MD8)
5. Human alpha-thrombin, green (PDB code: 1THS)
6. Human single chain tissue type plasminogen activator, pink (PDB code: 1BDA)

Parts 184-195 and 216-226 are shown as ribbons, residues Asp189 and Ser195 are shown as sticks.

**B: Amino acid sequence alignment of the superimposed enzymes**

Amino acid sequence segments 184-226 (chymotrypsinogen numbering) are shown. Numbering on the right side refers to the original sequence numbering. Sites 189 and 195 are highlighted in bold; sequences of loops L1 and L2 are marked with dots. Alignment was performed with the Clustal W software (Thompson et al., 1994).

## **1.9 Site specific mutagenesis studies investigating the structural basis of S1 site specificity**

Since the tool of site specific mutagenesis became available, a series of interconversion experiments of trypsin, chymotrypsin and elastase specificities have been carried out by replacing different sets of residues at the substrate binding site with the corresponding amino acids in the other enzyme (**Figure 11**). I focus here on these interconversion experiments and their results, mostly passing by the overview of the extended literature of experiments aimed at the alteration of enzyme specificities by various mutations. In most of these experiments, variants of the substrate succinyl-Ala-Ala-Pro-**Xaa**-7-amino-4-methylcoumarin were used, where Xaa denotes the amino acid at P1 position varied according to the tested specificity. This allows the direct comparison of kinetic results; however, care must be taken because the units used in the literature are often different.

### **1.9.1 Attempts to convert trypsin to chymotrypsin**

The first attempt of specificity interconversion consisted in the removal of the negative charge from the pocket of trypsin by the D189S mutation (Gráf *et al.*, 1988). Computer graphic modeling suggested that the geometry of the pocket should remain unchanged by the mutation, as Ser189 in chymotrypsin has a similar conformation to Asp189 in trypsin. Kinetic results, however, reported a drastic loss of tryptic activity coupled to a slight rise in chymotrypsin-like activity, resulting in a poorly active and aspecific enzyme. The chymotrypsin-like activity of the mutant was four orders of magnitude below that of wild type chymotrypsin (**Figure 12**). The addition of exogenous acetate can partly reconstitute the enzymatic activity of the mutant, and also the crystal structure of the mutant in complex with the inhibitor BPTI and the acetate showed an S1 site similar to that of wild type trypsin, with Ser189 rotated outward to the solvent leaving space for the acetate (Perona *et al.*, 1994). Later, the apo structure of the mutant (Szabó *et al.*, 1999) showed that indeed, Ser189 retained the wild type position and conformation, and the Ile16-Asp194 salt bridge with the oxyanion hole remained intact. However, the pocket was substantially distorted, especially the conformation of the 191-220 cystine and loop L2 that suggested weaker substrate binding and an improper positioning of the scissile bond. As in all the following experiments where the structure of the mutant was resolved, the rearrangements did not disturb the conformation of

the catalytic triad. The kinetic results indicated that contrary to the simple specificity model, further residues must be involved as specificity determinants, and sequence/structure comparisons led to further experiments exploring potential targets to interchange.

The next steps toward trypsin with chymotrypsin-like specificity were mutants D189S+YG217-219SGG and D189S+YG217-219SGG+Q192M (Gráf *et al.*, 1990). These variants exhibited similar trypsin-like activity to the single mutant D189S, and showed a trend of gaining a modest chymotrypsin-like activity in response to the successive alterations, but the activity was still four orders of magnitude lower than that of wild type chymotrypsin (**Figure 12**). This study indicated that this trend might continue by increasing the number of involved S1 site residues.

	138	172	190	200	210	220	230
TR2 RAT	I	Y	FLEGGKDSCQGDSSGGPVVC--NGE--LQGIVSWGYG-CALPDN-PGVYTK				
D189S	I	Y	.....S.....				
D189S+	I	Y	.....S.....SGG.....				
D189S++	I	Y	.....S..M.....SGG.....				
L1+L2	T	Y	-ASG-GS..M.....SGT.-STST.....				
L1+L2+	T	W	-ASG-GS..M.....SGT.-STST.....				
CTRB RAT	T	W	-ASG-VSSCMGDSGGPLVCQKDGVTLAGIVSWGSGVC-STST-PAVYSR				
CTR BOV	T	W	-ASG-VSSCMGDSGGPLVCQKDGVTLAGIVSWGSSTC-STST-PGVYSR				
CTRB RAT	T	W	-ASG-VSSCMGDSGGPLVCQKDGVTLAGIVSWGSGVC-STST-PAVYSR				
S189D CT	T	W	.....D.....				
L1+L2 CT	I	Y	FLE.GKD..Q.....YG-.ALPDN.....				
TR2 RAT	I	Y	FLEGGKDSCQGDSSGGPVVC--NGE--LQGIVSWGYG-CALPDN-PGVYTK				
TR2 RAT	I	Y	FLEGGKDSCQGDSSGGPVVC--NGE--LQGIVSWGYG-CALPDN-PGVYTK				
L1+L2 TR	I	Y	FLEGGKDSCQGDSSGGPVVC--NGE--LQGIVSWGYG-CALPDN-PGVYTK				
ELA1 RAT	I	Y	-GDGVRSGCQGDSSGGLHCLVNGQYSVHGVTSEFVSSMGCNVSKKPTVFTR				
CTRB RAT	T	W	-ASG-VSSCMGDSGGPLVCQKDGVTLAGIVSWGSGVC-STST-PAVYSR				
216+226	T	W	.....V.....T....				
L1+L2 CT	T	W	.GNGVR.....V.SMG.NV.KK.T....				
ELA1 RAT	I	Y	-GDGVRSGCQGDSSGGLHCLVNGQYSVHGVTSEFVSSMGCNVSKKPTVFTR				

**Fig. 11 Sequence alignments of the S1 region of wild type and mutant enzymes**

For each type of interchange experiments (see the text for further details), the template sequences of the mutagenesis are shown first, followed by the mutant sequences and finally by the sequences of the target enzymes. Dots denote residues similar to the template. The order of the presented sequences follows that of the experiments described in the text.

The next study in line demonstrated, however, that even the interchange of all the amino acids of the S1 site including loops L1 and L2 and residue 138 failed to transfer specificity (Hedstrom *et al.*, 1992). As a breakthrough, the mutant enzyme was equivalent to chymotrypsin in its catalytic rate, but its substrate binding was impaired, and showed a catalytic efficiency 2-3 orders of magnitude lower when compared to wild type chymotrypsin. Further interchange of residues over the borders of the S1 site came into focus.

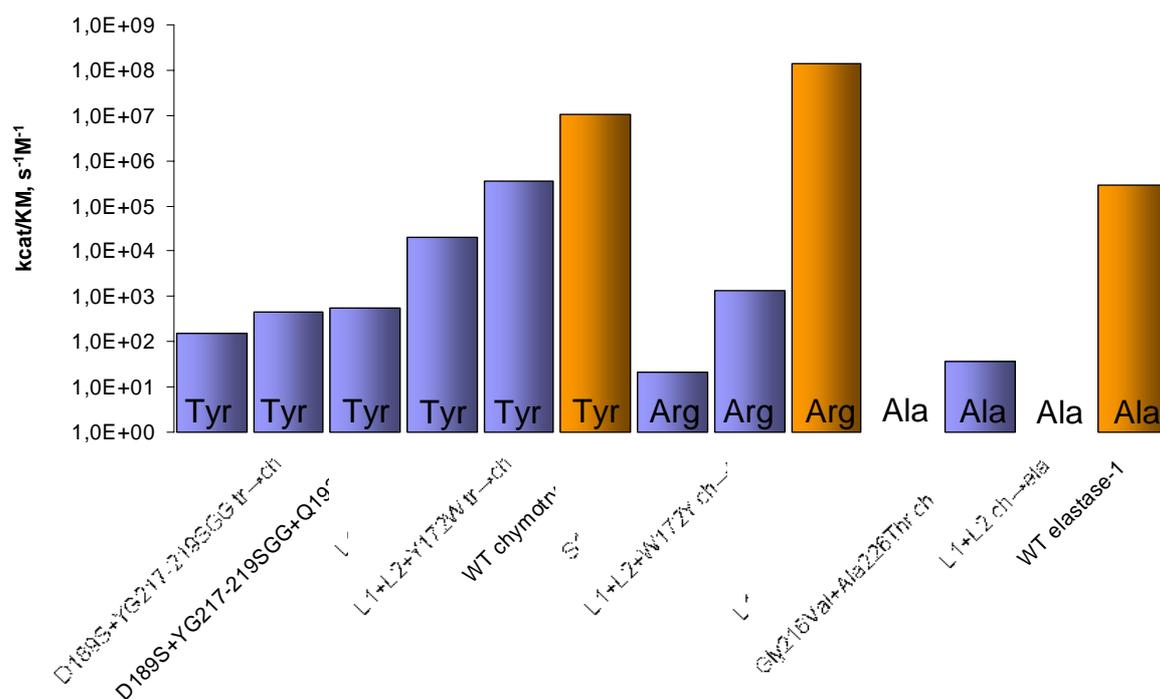
Finally, addition of the mutation Y172W in an adjacent surface loop (Hedstrom *et al.*, 1994) has brought the chymotrypsin-like activity of the trypsin mutant to the range of wild type chymotrypsin (roughly 10%), improving substrate binding by 50-fold. Crystallographic analysis (Perona *et al.*, 1995) showed that the mutation Y172W improved the stability of residues in the adjacent loop L2 and also in loop L1. These results suggested a complex model for substrate specificity that is basically different from the original simple one, since numerous residues that do not directly contact the substrate are considered as specificity determinants, and the differential specificities of trypsin and chymotrypsin are explained by the control of extended structural units (Gráf, 1995, Perona *et al.*, 1995).

### 1.9.2 Attempts to convert chymotrypsin to trypsin

To further test the above complex model, the reverse substitutions were introduced into chymotrypsin (Venekei *et al.*, 1996b). First, the S189D mutation was tested as a control for the simple specificity model. The single mutant showed a greatly reduced activity, while its specificity remained basically chymotrypsin-like. The trypsin-like activity of the mutant was six orders of magnitude below that of wild type trypsin, although the activity of trypsin on its preferred Arg substrate is one order of magnitude higher than that of chymotrypsin on its preferred Tyr substrate (**Figure 12**). A subsequent structural analysis (Szabó *et al.*, 2003) demonstrated that the S1 site was even more distorted, when compared to the D189S trypsin mutant, with the 191-220 cystine inside the pocket and Asp189 turned outward to the solvent. Furthermore, the Ile16-Asp194 salt bridge was missing, Asp194 was in a zymogen-like conformation and the oxyanion hole was not formed.

The further substitutions in the S1 region included different combinations mirroring the trypsin to chymotrypsin conversions, and they resulted at best in a slightly trypsin-like specificity coupled to an activity of five orders of magnitude below the wild type level. This was far below the results of the trypsin to chymotrypsin conversion, and suggested that

reaching the trypsin-like specificity in chymotrypsin requires even further mutations of sites around the pocket that, in addition, must be basically involved in determination of specificity. Such a possible target site might be the autolysis loop. A recent *in silico* study analyzed the correlated motions of surface residues in wild type trypsin and in the successful trypsin to chymotrypsin conversion mutant (Ma *et al.*, 2005). The authors found that motions in wild type trypsin, loops L1 and L2 show correlation with motions in the autolysis loop. In the mutant, however, this correlation was not present. This might suggest that the chymotrypsin-like loops in the trypsin mutant are not involved in interactions with the autolysis loop. Furthermore, the cleaved autolysis loop in chymotrypsin does not have a structure as defined as that of the uncleaved autolysis loop of trypsin. In the reverse conversion experiment, however, the trypsin-like loops introduced to chymotrypsin might miss their wild type interactions with the intact autolysis loop of trypsin.



**Fig. 12 Comparison of specific activities of wild type and mutant enzymes**

$k_{cat}/K_M$  values (given in  $s^{-1}M^{-1}$  and shown on a logarithmic scale) were measured on succ-Ala-Ala-Pro-Xaa-AMC substrates where Xaa denotes the actual P1 residue shown on the base of the value bars. The specific activities of wild type enzymes (orange) are shown on their preferred substrate, while specific activities of the mutant enzymes (blue) are shown on the most preferred substrate according to the targeted enzyme specificity. The order of the presented activities follows that of the experiments described in the text.

### **1.9.3 Attempts to convert trypsin to elastase**

As another test of the complex model, a trial very similar to the successful trypsin to chymotrypsin conversion was carried out (Hung and Hedstrom, 1998). Substitutions at the differing sites were introduced into trypsin in order to convert it to elastase-1 as well, but surprisingly the mutants had no measurable amidase activity, only assays on ester substrates showed a slight elastase-like specificity. Also this result suggests that a unique, extended structure determines each specificity of pancreatic serine proteinases, with key stabilizing components outside the S1 region, in some cases. The complex model became even more complex.

### **1.9.4 Attempts to convert chymotrypsin to elastase**

Looking over the previous attempts shows that each of them involved the removal or the introduction of the negative charge of Asp189 that have severe consequences, as shown by the kinetic results and the structures of the single D189S trypsin and S189D chymotrypsin mutants: when removed from trypsin, an important interaction partner that forms the center of the H-bonding network is lost, and when introduced into chymotrypsin the scaffold of the S1 site is not prepared for the dissipation of the charge. In this next study in line (Venekei and Jelinek, unpublished results, 2000) we supposed that the severe nature of this intervention might be responsible for the high number of mutations needed in the only successful trypsin to chymotrypsin conversion, and also for the fail of the other attempts. The chymotrypsin to elastase conversion seemed to overcome this problem, and mutants according to both the simple and the complex specificity model were prepared to explore the basis of S1 site specificity when the problem of exchanging the charge of Asp189 is not present.

The double mutant Gly216Val+Ala226Thr according to the simple model showed an aspecific profile with a slight Ala over Tyr preference, but its activity was four orders of magnitude lower than that of elastase-1 on the Ala substrate. The mutant according to the complex model lost 5-6 orders of magnitude of activity and its specificity profile remained chymotrypsin-like with no measurable activity on the Ala substrate. In this specific case of conversion, the approach of the simple model yielded somewhat better results.

## 2. Aims of the study

The template enzyme used for mutagenesis in the chymotrypsin to trypsin conversion studies was chymotrypsin-B. The A and B isoforms of chymotrypsin show very high sequence homology, the only important difference in the S1 region is at site 226: isoform A contains a Gly at this site while isoform B contains an Ala. Examination of the effect of this difference on the specificity profile of the two isoforms (Hudáky *et al.*, 1999) proved that the impaired activity of chymotrypsin B on the largest Trp substrate residue is caused by the slight steric hindrance of the extra methyl group at site 226, while the smaller P1 residues Phe and Tyr are not involved. As the two isoforms show no other important differences, Ala at site 226 was not taken into account in the design of the chymotrypsin to trypsin experiments.

However, sequence comparison showed that Gly at site 226 is conserved among trypsins, which may indicate an important role in the determination of a trypsin-like substrate specificity. Indeed, introducing a single methyl group at site 226 of trypsin reduces the activity on trypsin substrates by 3-4 orders of magnitude (Craik *et al.*, 1985). The crystal structure of this mutant revealed that the G226A mutation causes the misalignment of both Arg and Lys substrates at the S1 site, and showed the lack of the direct electrostatic interaction between the P1 Arg and Asp189 of trypsin (Wilke *et al.*, 1991). The large extent in which this mutation lowered the tryptic activity raised doubts about the results obtained in the chymotrypsin to trypsin conversion experiments.

To investigate whether position 226 is important for the chymotrypsin to trypsin specificity conversion as well, we introduced the A226G replacement in mutants that were prepared according to both the simple and the complex model of specificity: into the single S189D chymotrypsin-B mutant (S189D+A226G), and into one with further substitutions in the S1 region (S1+A226G) (**Figure 13**).

As the previous studies indicated, determination of the crystal structures of the mutants can give a much deeper understanding of the effects of mutations than kinetic results alone. The encouraging kinetic results of the S189D+A226G mutant raised important questions about the possible rearrangements at the S1 site, therefore we determined the crystal structure of the mutant, and also investigated the pH dependence of the activity and the internal tryptophan fluorescence.

	190	200	210	220	230
CTRB RAT	-ASG-VSSCMGDSGGPLVCQKDGVTLAGIVSWGSGVC-	STSTP	AVYSR		
S189D	.....D.....				.....
S189D+A226G	.....D.....				.....G.....
S1	FLE.GKD..Q.....			Y.-.ALP.....	.....
S1+A226G	FLE.GKD..Q.....			Y.-.ALP.....	.....G.....
TRY2 RAT	FLE.GKD..Q.....	V.--N.E--.Q.....		Y.-.ALPDN.	G..TK

**Fig. 13 Sequence alignment of the S1 region of wild type enzymes and the mutants**  
The template sequence of wild type chymotrypsin-B is shown first, followed by the mutant sequences and the sequence of trypsin. Dots denote residues similar to the template; site 226 is highlighted in red.

When looking over to the long history of specificity conversion experiments, the results show a rather confusing picture. Conversions according to the simple model failed to transfer specificity, and apart from the only successful trypsin to chymotrypsin conversion, the exchange of the whole S1 site also resulted in a demolished activity. The reason for this seems to be that interactions between the S1 region and other parts of the enzyme might be of high importance, as the S1 site is basically embedded in the whole structure. This suggestion is in accord with the high gain of activity upon the Y172W exchange in the trypsin to chymotrypsin conversion. Another theory suggest that specificity conversions could be done by the exchange of only a few carefully selected key residues of the S1 site, without altering its general scaffold including the interactions with distal parts of the parent enzyme. In fact, the complete exchange of the S1 site might include the introduction of residues and/or structural motifs that are not in accord with the scaffold that surrounds the S1 site of the enzyme. The aims of this study were i) to further explore the structural background of S1 site specificity, with focus on the stabilization of the specificity determinant Asp189 charge in trypsin, and ii) to further improve the model that describes specificity to a point where it can pass the test of the conversion experiments. This better understanding of the structural basis of specificity aims to contribute to improvements in the fields of protein engineering and pharmacology.

## 3. Materials and Methods

### 3.1 Materials

For the activation of the S1+A226G chymotrypsin mutant, we used the highly purified enterokinase of Biozyme (EK-3), free of pancreatic protease contamination. Bovine chymotrypsinogen, bovine trypsin, SBTI-sepharose, MUGB, Succinyl-Ala-Ala-Pro-Phe-AMC, AMC and 7-methylumbelliferon were from Sigma Chemical Co. The substrate succinyl-Ala-Ala-Pro-Lys-AMC is not commercially available, it was prepared at our department as described (Gráf *et al.*, 1988). The oligopeptide substrate library was prepared as described (Antal *et al.*, 2001) at the peptide synthesis facility of our department.

### 3.2 Construction of mutants

Chymotrypsin-B mutants S189D+A226G and S1+A226G were constructed according to Kunkel (Kunkel, 1985), from a previous S189D and a multiple substituted S1 region mutant construction, respectively (**Figure 13**). M13mp18 mutant phage stocks were used to prepare the uracil-containing single stranded DNA templates in the BO 265 *E. coli* strain lacking the enzyme deoxyuridine transferase. The uracil content of the templates was controlled by phage-titration with JM 101 (no plaques formed) and BO 265 (covered with plaques) strain plate cultures. Purity of the templates was controlled by agarose electrophoresis and by measuring the 260nm/280nm absorbance ratios for detecting protein contamination. Sense direction oligos containing the desired mutation were phosphorylated with the enzyme T4 polynucleotide kinase, annealed to the antisense single-stranded templates, and the mutant strands were completed with the enzymes T4 polymerase and T4 ligase. Hybrid DNAs were transformed into JM 101 cells where the original antisense template strand containing uracil was degraded and the mutant strand was amplified. The new mutant sequences were confirmed by DNA sequencing.

Similarly to the previous chymotrypsin to trypsin conversion mutants (Venekei *et al.*, 1996a,b) the new mutants were also expressed as a propeptide chimera. Following the activation cleavage of chymotrypsinogen by trypsin, the chymotrypsin propeptide remains anchored to the active enzyme via the 1-122 cystine. Trypsinogen can be activated by enterokinase and its propeptide is released then to the solvent. In these chymotrypsin mutants

the propeptide was replaced by the trypsin propeptide and site Cys122 was mutated to Ser, so that they could be activated with enterokinase. This excluded the contamination of the enzyme preparations with trypsin otherwise necessary for activation, and an accurate determination of even low tryptic activities of the mutant enzymes became possible. The kinetic parameters of the chimera do not differ from those of wild-type chymotrypsin under physiological conditions (Venekei *et al.*, 1996a). On the other hand, the wild-type propeptide increased the stability of the protein as studied under denaturing conditions (non-physiological pH, temperature and denaturing agents) (Kardos *et al.*, 1999). One could expect that the introduced mutations might have a destabilizing effect similar to the above denaturing conditions, and the disulfide-linked chymotrypsin propeptide might have a stabilizing effect in the mutants. To examine this possibility, the chymotrypsin propeptide-containing forms of the mutants were also expressed and characterized. They did not, however, show any significant difference in their catalytic activities when compared to their chimeric counterparts. The crystal structure of the form with the wild type propeptide was determined to compare the possible structural differences of the propeptide binding region to that of the single mutant S189D chimera structure. The crystallized mutant contained an extra MSTQA amino acid sequence at the N-terminus, before the Cys1 of the propeptide.

### **3.3 Expression and purification of the mutants**

The wild-type and the mutant chymotrypsinogen sequences were subcloned into the pET-17b expression vector with the restriction endonucleases HindIII and BamHI. These plasmids were transformed into the BL21 (DE3) pLysS *E. coli* strain chemically competent cells by a 1min heat shock at 42°C, and the expression was conducted according to a protocol based on the plasmid manufacturer's instructions (pET System Manual, 1997): transformed cells were incubated in 1.0 ml LB medium for 1h at 37°C, then transferred into 50 ml LB containing 20 µg ampicillin, and incubated overnight at 37°C. 5ml of this culture was transferred into 500ml LB containing 200 µg ampicillin and was incubated until the OD<sub>600nm</sub> of the culture reached 0.6, then cells were induced with 0.5mM IPTG and another 200 µg ampicillin was added to ensure selective growth. After 3h incubation at 37°C, cells were collected in 1/10 volume of TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0) and were frozen at -20°C. Similarly to wild type chymotrypsin, the mutant proteins formed inclusion bodies in the cells. Cells were thawed and sonicated, and the inclusion body fraction was collected by

centrifugation and washed three times with TE buffer. For renaturation of the expressed protein, the inclusion body fraction was solubilized with 6M GuHCl, 0.1M TRIS-HCl (pH 8.0), 100mM DTT. The solution contained ~10 mg/ml protein. The solubilized protein was diluted 200-fold into the refolding buffer containing 1M GuHCl, 5 mM cysteine, 1 mM cystine, 50 mM Tris-HCl (pH 8.0), 5 mM EDTA. The renaturation process was conducted at 4°C, overnight. The renatured protein solutions were then dialyzed against 2mM HCl, 10mM CaCl<sub>2</sub>, and ultracentrifuged (45000 rpm, 30min, 4°C).

Zymogens were activated by highly pure enterokinase at a 100:1 (w/w) zymogen/enterokinase ratio, and then the active forms were purified by affinity chromatography on an SBTI-Sepharose column. Fortunately, both mutants showed enough affinity toward SBTI to use this single purification step. The purity of the preparations was analyzed by SDS-PAGE. The enzyme concentration was determined by Bradford assay for the low activity mutants, and by active site titration with MUGB and MUTMAC (Jameson *et al.*, 1973) for S189D+A226G mutant and wild-type chymotrypsin, respectively. For crystallization, the S189D+A226G mutant was further purified on a Pharmacia HiPrep Sephacryl-100 column in presence of 0.2 mg/ml benzamidine.

### **3.4 Enzyme assays**

#### **3.4.1 Introduction**

In the present study we used the steady state kinetics approach for the investigation of specificity. The term steady state refers to the situation in which the value of a particular quantity is constant, because of its similar formation and depletion rates. When an enzyme is mixed with a large excess of substrate, an initial period called pre-steady state of the reaction is started. During this period the concentration of enzyme-bound intermediates reach a constant, steady state level that is maintained over the following steady state period of the reaction. The steady state here is an approximation, but with large excess of substrate and a relatively short measurement time, the depletion of the substrate can be neglected.

The theoretical basis of steady state kinetics is a reaction that occurs according to the Michaelis-Menten mechanism described by the following scheme:



It is assumed that the concentration of the enzyme is negligible to that of the substrate, and the initial rate  $v_0$  of the depletion of the substrate is measured. The first step is the formation of the enzyme-substrate complex with the dissociation constant  $K_S$ . This step is assumed to be rapid and reversible, with no chemical changes taking place. The next step is that of the chemical processes, assumed to be irreversible and occurring with the first-order rate constant of  $k_{cat}$ . The dissociation of the enzyme-product complex is also assumed to be fast. It was found experimentally in many cases that the measured rate  $v_0$  is proportional to the concentration of the enzyme,  $[E]_0$ , but  $v_0$  follows saturation kinetics when the substrate concentration  $[S]$  is raised. At saturating  $[S]$ ,  $v_0$  reaches its maximum  $V_{max}$ . These kinetic phenomena were interpreted by the Michaelis-Menten mechanism, resulting in the Michaelis-Menten equation (*Eq. 2*):

$$V_0 = ([E]_0[S] k_{cat}) / (K_M + [S]) \quad \text{Eq. 2}$$

where

$$k_{cat}[E]_0 = V_{max} \quad \text{Eq. 3}$$

and

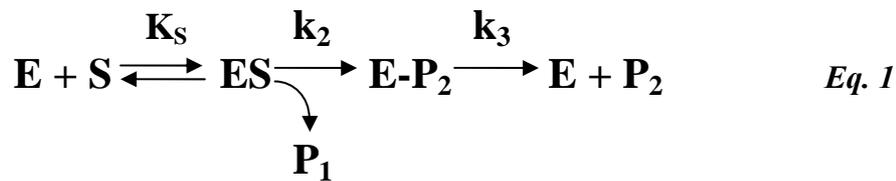
$$K_M = (k_{-1} + k_{cat}) / k_1 \quad \text{Eq. 4}$$

The concentration of substrate at which  $v_0 = V_{max}/2$  is termed the Michaelis constant  $K_M$ , and when  $k_{-1} \gg k_{cat}$  then  $K_M = K_S$ . When the assumptions of the Michaelis-Menten conditions are not met, the meanings of the  $k_{cat}$  and  $K_M$  parameters will change.

Generally,  $k_{cat}$  will be the function of all the first-order rate constants of the reaction and could only be assigned to any particular step if simplifying features occur, e.g. if one of the steps is much slower than the others. In this case,  $k_{cat}$  will refer to the first order rate constant of the rate limiting step of the reaction. Similarly,  $K_M$  will not be the true dissociation constant of the enzyme-substrate complex (called also the Michaelis-complex), but an apparent overall dissociation constant of all the enzyme-bound species of the reaction.

$k_{cat}$  and  $K_M$  values presented in this study were calculated using the Michaelis-Menten equation. However, the mechanism of the measured hydrolysis reaction of pancreatic serine proteinases is substantially different from the Michaelis-Menten mechanism, as shown in

equation 1 (Zerner *et al.*, 1964), where  $k_2$  represents the acylation rate and  $k_3$  represents the deacylation rate:



For a proper interpretation of the results it must be noted that in the measured reactions the acylation is the rate limiting step, therefore  $k_{\text{cat}} = k_2$  and  $K_M = K_S$  (Perona *et al.*, 1995).<sup>2</sup>

In this study, apart from the measurements on the competitive oligopeptide library, specificity of wild type and mutant enzymes is described by the generally used comparison of  $k_{\text{cat}}/K_M$  values measured on different substrates.  $k_{\text{cat}}/K_M$  is often referred as the specificity constant, for the reason that it determines the specificity for competing substrates, not detailed here. In the specific case of pancreatic serine proteinases measured on amide substrates,  $k_{\text{cat}}/K_M = k_2/K_S$ . This means that the specificity constant is determined by ground state binding and by the acylation rate or, in other words, specificity is determined by the specific substrate binding process and the proper positioning of the scissile bond. An extensive characterization of numerous mutant trypsins demonstrated that the key kinetic determinant of substrate specificity is the rate of the acylation reaction (Hedstrom *et al.*, 1994b).

### 3.4.2 Description of the experiments

Amide hydrolysis was measured on Succinyl-Ala-Ala-Pro-Phe-AMC and Succinyl-Ala-Ala-Pro-Lys-AMC substrates in a 50mM Tris-HCl pH 8.0, 10mM CaCl<sub>2</sub>, 0.1 M NaCl reaction buffer at 37°C using a Spex Fluoromax spectrofluorimeter. The excitation and emission wavelengths were 380nm and 460nm, respectively. The data were analyzed with the Enzfitter software.

Autoactivation experiments were conducted with incubation at room temperature in the reaction buffer above, and were monitored by SDS-PAGE and activity measurements.

Chymotrypsinogen activation was carried out at 37°C in the reaction buffer above. 1 μM chymotrypsinogen was incubated with 16 μM S189D, 14 μM S189D+A226G and 5nM

<sup>2</sup> In the case of ester substrates, where deacylation is the rate-determining step,  $K_M = K_S[k_3/(k_2+k_3)]$  and  $k_{\text{cat}} = k_2k_3/(k_2+k_3)$ .

bovine trypsin for 80 minutes. Chymotrypsin activity was measured every 10 minutes on Succinyl-Ala-Ala-Pro-Phe-AMC substrate and was expressed as percentage of the total activity.

For the measurement of benzamidine inhibition of the S189D and S189D+A226G mutants, progress curves were recorded with enzyme concentrations of 80nM to 160nM and with substrate concentrations of 50μM to 100μM. Constants were determined by analyzing the progress curves using the DynaFit software (Kuzmic, 1996) according to the Michaelis-Menten equation.

pH dependency of amide hydrolysis was measured on 400 μM Succinyl-Ala-Ala-Pro-Phe-AMC and 400 μM Succinyl-Ala-Ala-Pro-Lys-AMC substrates, in 40 mM buffers of Na-Acetate (pH4.0-pH5.0), MES (pH5.5-pH6.5), MOPS (pH7.0-7.5), Tris-HCl (pH8.0-pH8.5) CHES (pH9.0-pH9.5) and CABS(pH10.0-10.5). The reaction buffers contained 10 mM CaCl<sub>2</sub> and 0.1 M NaCl, measurements were carried out at 37°C using a Spex Fluoromax spectrofluorimeter. The differences between the activity values of three measurements were within 10%. Unweighted curve fittings (**Figure 20**) were performed using the Origin software (OriginLab), using the following equations (*Eq. 5- Eq. 7*):

for chymotrypsin on Phe substrate and the mutant on Lys substrate:

$$y = V_m / (1 + 10^{-x} / 10^{-pK_a} + 10^{-pK_b} / 10^{-x}) \quad \text{Eq. 5}$$

for chymotrypsin on Lys substrate:

$$y = V_{m_1} / (1 + 10^{-x} / 10^{-pK_{a1}}) + V_{m_2} / (1 + 10^{-x} / 10^{-pK_{a2}}) \quad \text{Eq. 6}$$

for the mutant on Phe substrate:

$$y = V_{m_1} / (1 + 10^{-x} / 10^{-pK_a} + 10^{-pK_b} / 10^{-x}) + V_{m_2} / (1 + 10^{-pK_{b2}} / 10^{-x}) \quad \text{Eq. 7}$$

*Eq. 5* describes a bell-shaped pH dependence (Segel, 1993), where  $x = \text{pH}$  and  $V_m$  is an amplitude parameter. We derived *Eq. 6* and *Eq. 7* from equation *Eq. 5* to fit a half bell curve with two titration steps, and to fit a bell shaped curve with two titration steps on the right side, respectively.

pH dependence of the internal tryptophan fluorescence of the double mutant and wild type chymotrypsin was also measured under similar conditions, using the pH 10.5 and the pH 8.0 reaction buffers. The enzyme concentration was 0.1mg/ml, the excitation and emission wavelengths were 297nm and 460nm, respectively. Fluorescence intensities of i) the mutant, ii) the mutant complexed with 200 $\mu$ M benzamidine and iii) chymotrypsin at the two different pH values were compared. The differences between the changes of fluorescence intensities of three measurements were within 0.2 %.

### **3.5 Specificity profiling**

Besides defining specificity by the usual comparison of individual  $k_{cat}/K_M$  measurements, we performed specificity profiling on a competing oligopeptide substrate library (Antal *et al.*, 2001). This approach has the advantage of defining specificity under conditions where the enzyme can discriminate between several potential substrates present in the same reaction mixture (Cornish-Bowden, 1995).

The oligopeptide library has seven members with a sequence HAAPXSADIQIDI, where X represents the different P1 residues, Lys, Arg, Tyr, Leu, Phe and Trp. Since most serine proteinases are unable to cleave bonds following a P1 proline, X=Pro served as internal standard. This sequence was designed to meet several demands: it can be accepted both by trypsin and chymotrypsin, it focuses the cleavage at the P1 site, all the initial components and cleavage products can be easily identified by RP-HPLC, and its solubility allows to reach the needed concentration under the required conditions for the enzymes. These substrates have the advantage of forming similar interactions with the enzymes as the natural peptide substrates, while those having an artificial leaving group might form interactions different from the normal peptide substrate interactions. These seven individual peptide substrates compete for the proteinase during the enzymatic reaction.

Enzyme concentrations were optimized to ensure comparable enzyme reaction rates: 0.6  $\mu$ M for S189D and S189D+A226G mutants, 0.075  $\mu$ M for chymotrypsin-B and 0.0012  $\mu$ M for bovine trypsin. The concentration of each substrate mixture component was 40  $\mu$ M. The digestions were terminated at different reaction times (0, 1, 4, 16, 32 and 64 min) by the addition of 20  $\mu$ l of 5 M acetic acid to the 100  $\mu$ l aliquots. The reaction was monitored by RP-HPLC separation of the components. The calculated peak areas, normalized with the 0 min value of the X=Pro internal standard were used to obtain the percentage specificity data.

## **3.6 Crystallization, structure solution and refinement**

### **3.6.1 Introduction**

Crystallographic models depict the average structure of somewhere between  $10^{13}$  and  $10^{15}$  molecules throughout the crystal, showing an average and static image of the dynamic molecules. The image is obtained from the intensities and directions of X-ray beams diffracted by protein crystals. Visible light with wavelengths of 400-700 nm cannot produce an image of individual atoms that are only about 0.15 nm (=1.5Å) apart - the electromagnetic radiation of this wavelength falls into the X-ray range. A single protein molecule alone would be a very weak scatterer of X-rays. This problem is solved by analyzing diffraction patterns from crystals containing many ordered protein molecules. Diffraction patterns are collected from all orientations of the crystal with respect to the X-ray beam.

Another problem is that X-rays cannot be focused by lenses; therefore a focused image of the protein cannot be obtained. So the next step consists of a computer simulation of the effect of objective lenses that reconstruct the image, by a process called Fourier transform. The Fourier transform describes the mathematical relationship between an object and its diffraction pattern, and basically this process consists of obtaining the mathematical function whose graph is the desired image, called the electron-density map. To obtain this function, amplitude, frequency and phase of each reflection is required. However, diffraction patterns show only intensities that give the amplitudes and positions that refer to the frequencies, but they contain no information about the phases. The phase of each reflection has to be obtained from new diffraction patterns of experiments with heavy atom complexes of the proteins, or from known homologous crystal structures.

In the final step called map fitting or model building, the map is interpreted by building a molecular model that is on one hand consistent with the obtained map, and, on the other hand is chemically, stereochemically and conformationally reasonable. This iterative process involves manual adjustments and automatic corrections of bond lengths and angles, and is guided to convergence by monitoring the criteria above along with the residual index (R-factor) of the model. The R-factor shows the difference between the measured structure-factor amplitudes and those calculated from the current model. A more revealing criterion in judging the quality of the model is the R<sub>free</sub> factor that uses a randomly chosen test-set of

intensities that were not part of the refinement process for the cross-validation of the calculated (i.e. the model) and the observed data. When all criteria are met and the R factors do not decline any more, the best possible model is obtained from the diffraction data.

One of the most important properties of a crystallographic model is the resolution, given in Ångströms. Contrary to microscopy, resolution refers here to the amount of data used for determination of the structure: a crystal that diffract measurable reflections out to a distance of say  $1/(3\text{Å})$  from the origin will yield a model with  $3\text{Å}$  resolution. However, the precision of the atom positions presented by the model depends also critically on the quality of the data, as reflected by the R-factor. Tools like the Luzzati plot (Luzzati, 1952) can help to estimate the upper limit of rms errors in the atomic coordinates of the model, based on the calculated R-factors.

### 3.6.2 Description of the experiments

S189D+A226G chymotrypsin containing the wild type propeptide was crystallized at  $20^{\circ}\text{C}$  by the hanging drop, vapor diffusion method, mixed 1:1 with a 0.1M HEPES, pH 7.0 precipitant solution containing 30% PEG6000 and equilibrated against 0.5 ml of precipitant solution. The final protein concentration in the drop was 7.5 mg/ml, and the solution contained benzamidine in a twofold molar excess. Crystals were grown in 6 days. The crystals were flash cooled in liquid nitrogen after briefly soaking in cryosolution containing the crystallization solution and 20% glycerol. Crystallographic data were collected from a single crystal at the ESRF synchrotron source on beamline ID 14 EH2 ( $\lambda = 0.933\text{ Å}$ ) at 100 K.

The X-ray diffraction displayed high mosaicity (estimated to be  $1.8^{\circ}$ ). Macroscopic crystals of proteins are not perfect arrays of the unit cells; they are rather mosaics of many submicroscopic arrays in rough alignment with each other. The crystals of the mutant suffered of high mosaicity that resulted in greater mosaic spread of the X-ray reflections. The quality of the diffraction is the most important criterion in determining the quality of the crystallographic model; therefore weaknesses of the model described later are the consequences of the mosaicity of the crystals yielding lower quality data.

Crystallographic intensities were integrated and scaled to a resolution of  $2.2\text{ Å}$  using Mosflm and Scala of the CCP4 package (CCP 4, 1994). The crystal structure belonged to space group  $P2_1$ . Completeness of the data was 98.5 % at  $2.2\text{ Å}$  resolution. The structure was solved by obtaining the phases by molecular replacement using the program Amore (Navaza,

2001) of the Collaborative Computing Project 4 (CCP 4, 1994). The phasing model was derived from the X-ray structure of S189D chymotrypsin (PDB entry: 1kdq) (Szabó *et al.*, 2003). A single well defined solution appeared with correlation coefficient of 0.55 and R-factor of 38.3 %, providing an adequate initial estimate of phases.

After the initial rigid body refinement of the crude model, it became obvious that regions of 186-190 and 218-221 were significantly different compared to the S189D single mutant structure. Therefore these loops were systematically rebuilt from scratch using iterative cycles of phase improvements in RESOLVE (Terwilliger, 2000), manual rebuilding using the program Coot (Emsley and Cowtan, 2004) and restrained TLS refinement with REFMAC (Murshudov, 1997). For modeling anisotropic disorder in the crystal structure two TLS groups were defined: the alpha-chain domain (residue ranges 19-146, 151-157 and the covalently bound propeptide) and beta-chain domain (residue range 158-243). Introduction of more TLS groups did not significantly improve the Rfree value. The switch to TLS refinement was justified by a drop of 5.2 % in Rfree value indicating that anisotropic disorder is an important characteristic of the diffraction data, and a further slight improvement in the Rfree value was achieved by including some 40 more solvent molecules in the structural model. Final Rfree values, however, remained still unusually high.

The final protein model contained 230 residues of rat chymotrypsin including the bound propeptide. The stereochemistry of the structure was assessed with Whatcheck (Hooft *et al.*, 1996) and PROCHECK (Laskowski *et al.*, 1993). Data collection and refinement statistics are summarized in **Table II**. The crystallographic data and the refined model have been deposited at the Protein Data Bank ([www.pdb.org](http://www.pdb.org)) with accession code 2jet.

Images that present protein structures throughout this study were generated using the software DeepView 3.7 (Guex and Peitsch, 1997) and PyMol (Delano, 2002).

### **3.7 Structure analysis**

Superposition of X-ray structures was improved with iterative sieve fitting (Lesk, 1991) in LSQMAN (Madsen and Kleywegt, 2002) until about 50% of the matched residues were under the final distance cutoff (typically 0.5-0.7 Å). Structural comparisons were performed against bovine trypsin (1tpo) (Marquart *et al.*, 1983), S189D mutant of rat chymotrypsin (1kdq) (Szabó *et al.*, 2003) and bovine chymotrypsin (6gch) (Brady *et al.*, 1990). The precision of the crystal structure was evaluated by the Luzzati plot (Luzzati, 1952)

and the DPI method (Cruickshank, 1999) using the program SFCHECK (CCP 4, 1994). Using the conservative Luzzati estimate of 0.48 Å for the coordinate error, the structural changes described for the residue ranges 185-191 and 219-223 are significant.

## 4. Results

### 4.1 Kinetic results

To study the significance of Gly226 in specificity determination A226G mutants were constructed according to both the simple and the complex specificity model, and the kinetic properties of the mutants were determined on trypsin and chymotrypsin substrates. The A226G substitution was introduced into the S189D chymotrypsin-B single mutant with a slight chymotrypsin-like specificity profile, resulting in the S189D+A226G double mutant. The other attempt based on the complex model used a multiple-substituted chymotrypsin-B to trypsin mutant as template which has a slight trypsin-like specificity profile; the substitution resulted in the S1+A226G mutant (**Figure 13**). To compare effects of the A226G mutation on the S1 site specificity, kinetic constants were calculated from hydrolysis rates measured on Succinyl-Ala-Ala-Pro-Lys-AMC and Succinyl-Ala-Ala-Pro-Phe-AMC fluorometric amide substrates.

The A226G replacement exerted opposite effects on the mutants' specificities (**Table I**). The S1 mutant gained two orders of magnitude in activity, but its specificity profile changed back to that of a chymotrypsin-like protease, with two orders of magnitude higher activity on the chymotrypsin than on the trypsin substrate. On the other hand, the basically chymotrypsin-like profile of the S189D mutant became trypsin-like by the single A226G substitution. The catalytic activity on the Lys-containing substrate was increased by two orders of magnitude and at the same time it was decreased by one order of magnitude on the chymotrypsin substrate. The level of tryptic activity of the double mutant surpasses that of the best previous chymotrypsin to trypsin mutant by 50-fold. The trypsin-like activity of the mutant allowed the use of active site titration for the measurement of enzyme concentration, which was not possible with previous chymotrypsin to trypsin mutants. It is interesting to note that the increase in the specificity constants of both mutants S189D+A226G and S1+A226G resulted mainly from the elevation of the catalytic rate constants, similarly to previous observations of trypsin mutants (Hedstrom *et al.*, 1994b).

To further characterize its specificity, the S189D+A226G chymotrypsin mutant was tested on a competing oligopeptide substrate library developed for specificity profile analysis of serine proteases (see Materials and Methods). In this experiment where substrates with different P1 side chains are present in the same solution, the mutant showed a profile highly

similar to that of trypsin, with a considerable Arg over Lys preference, which is a typical feature of trypsin. The strong preference for the Arg substrate suggested that the P1 Arg and the Asp189 side-chains can form even a direct charge interaction. Affinity toward the chymotrypsin substrates Tyr, Phe and Trp were equally low, the elastase substrate with P1=Leu was not cleaved (**Figure 14**). Comparison of the profile of the single S189D mutant to that of wild type chymotrypsin-B shows that in contrast with its impaired activity and the serious distortions of its S1 site, the mutant discriminates between P1 side chains in a very similar way. The A226G substitution clearly turned this chymotrypsin-like specificity profile of the parent mutant S189D to that of a trypsin-like protease.

**Table I**  
Kinetic constants determined on Succinyl-Ala-Ala-Pro-Xaa-AMC polypeptide amide substrates

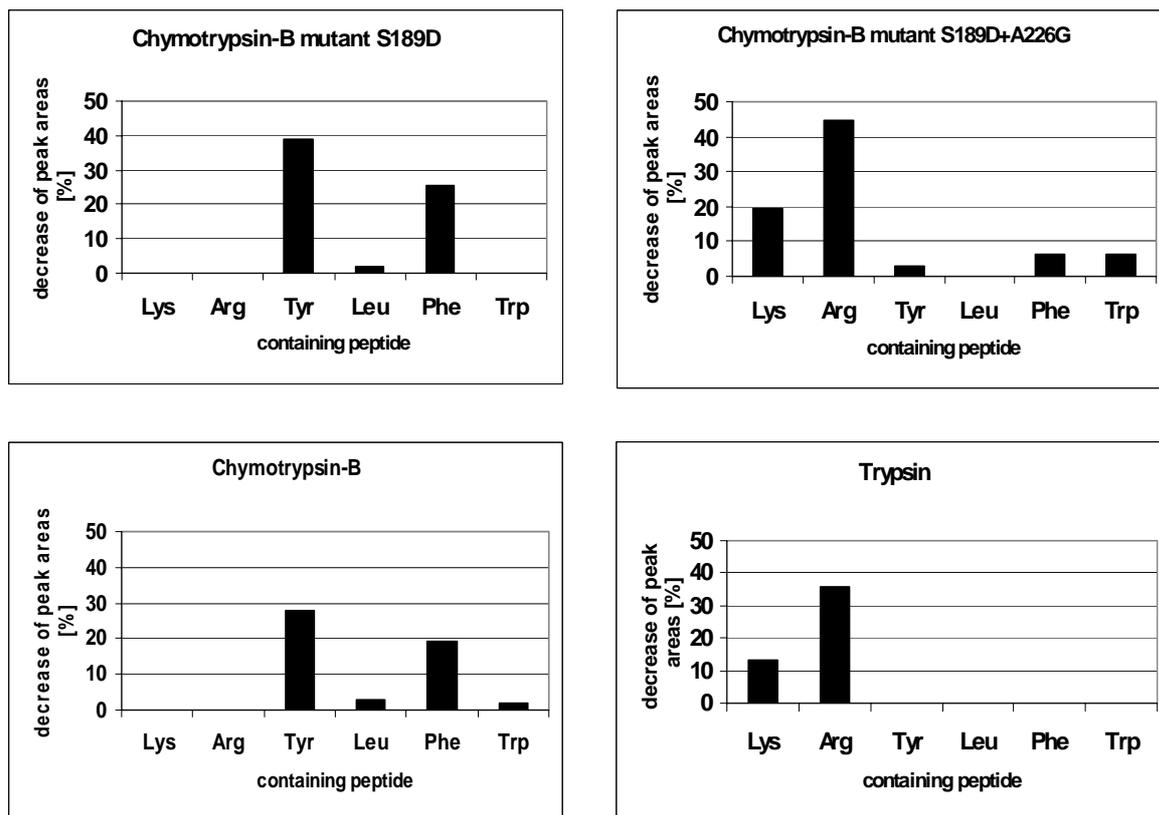
Enzyme	Xaa=Lys			specificity**
	$k_{cat}, s^{-1} \times 10^3$ (mean±SEM)	$K_M, mM$ (mean±SEM)	$k_{cat}/K_M$ $s^{-1}M^{-1}$	
wild-type chymotrypsin*	18±4	0.35±0.07	51	
S <sub>1</sub> *	11±5	0.13±0.03	84	
S <sub>1</sub> +A226G	31±4	1.5±0.8	21	
S189D*	5.6±1	0.17±0.08	33	
S189D+A226G	4100±200	0.99±0.04	4100	
wild-type trypsin*	70000±3000	0.0032±0.0001	2.2x10 <sup>7</sup>	

Enzyme	Xaa=Phe			specificity**
	$k_{cat}, s^{-1} \times 10^3$ (mean±SEM)	$K_M, mM$ (mean±SEM)	$k_{cat}/K_M$ $s^{-1}M^{-1}$	
wild-type chymotrypsin*	86000±2000	0.023±0.002	3.7x10 <sup>6</sup>	4.9
S <sub>1</sub> *	0.43±0.2	0.38±0.06	1.1	- 1.9
S <sub>1</sub> +A226G	180±40	0.095±0.003	1900	2.0
S189D*	100±30	0.041±0.009	2400	1.9
S189D+A226G	320±70	0.39±0.05	820	-0.7
wild-type trypsin*	100±20	0.13±0.05	770	-4.4

\* Data from Venekei *et al.* 1996b.  
\*\*specificity =  $\log(k_{cat}/K_M^{Phe} / k_{cat}/K_M^{Lys})$

Another important consequence of the A226G replacement was that the S189D+A226G chymotrypsinogen underwent autoactivation, which is a typical property of trypsinogen. The same mutation, however, did not turn the zymogen of the S1 mutant to an

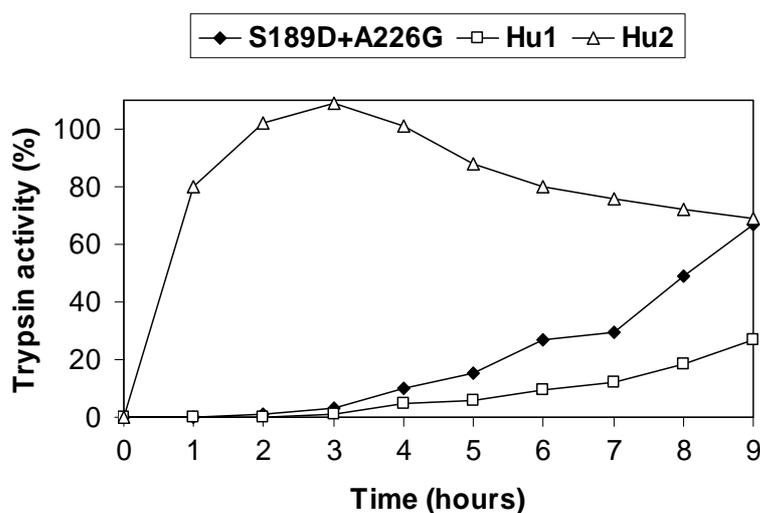
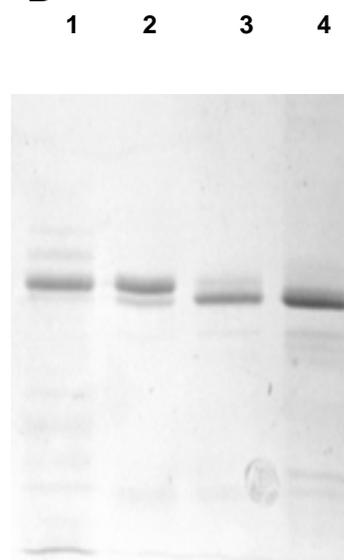


**Fig. 14** Graphical presentation of enzyme specificity profiles as determined on a competing oligopeptide substrate library (from Jelinek et al., 2004)

The sequence of the oligopeptide is HAAPXSADIQIDI, where X represents the different P1 residues. Decreases of peak areas after 16-min digestion are presented. Chymotrypsin-B and trypsin data were taken from Antal et al., 2001. Note that the enzyme concentrations in these experiments were different, because they were optimized to ensure comparable enzyme reaction rates (see Methods).

autoactivating one, and all the other previous chymotrypsin to trypsin mutants lacked this feature. This suggests that there is a lower limit in the structural integrity level of the trypsinogen-like conformation necessary for autoactivation, that was only surpassed by the zymogen of the double mutant. Similarly to trypsinogen, the S189D+A226G zymogen fully activated itself during an overnight incubation. The rate of autoactivation of the double mutant is comparable to those of wild-type trypsinogens (**Figure 15**). The large difference between the autoactivation rate of the human trypsin isoforms that exhibit otherwise similar catalytic activity suggests that apart from specific activity, other structural features also determine the ability of autoactivation in trypsin.

The S189D+A226G chymotrypsin mutant also gained the ability of activating chymotrypsinogen (**Figure 16**), that is activated exclusively by trypsin *in vivo*. Contrary to the autoactivation rates, chymotrypsinogen activation rates of trypsin and the mutant enzyme

**A****B**

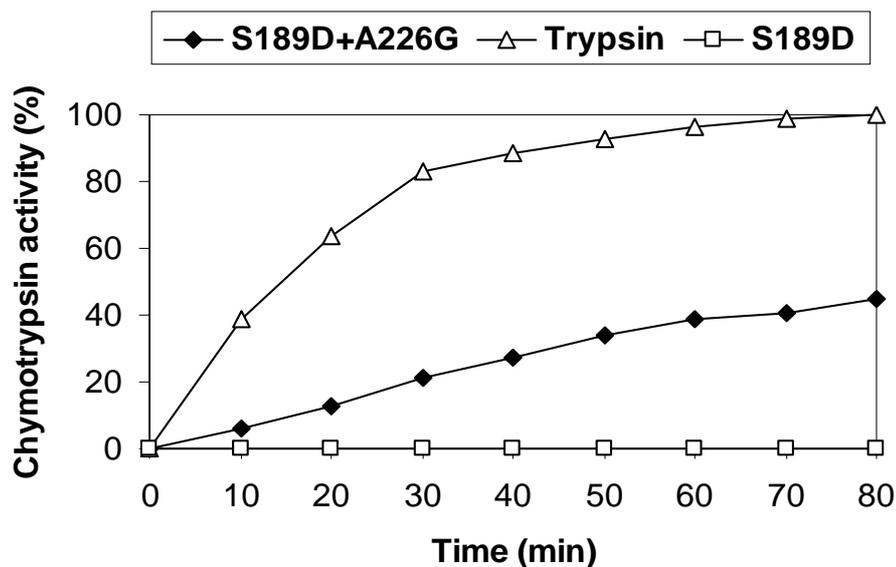
**Fig. 15 Autoactivation of the mutant chymotrypsinogen S189D+A226G** (from Jelinek et al., 2004)

**A:** Autoactivation of the mutant S189D+A226G ( $8\mu\text{M}$  final zymogen concentration) ( $\blacklozenge$ ) was compared to that of wild type human recombinant trypsinogen 1 ( $8\mu\text{M}$ ) ( $\square$ ) and wild type human trypsinogen 2 ( $1.8\mu\text{M}$ ) ( $\Delta$ ). Zymogens were incubated at  $37^\circ\text{C}$ , in the presence of  $50\text{ mM}$  Tris-HCl (pH 8.0),  $0.1\text{ M}$  NaCl and  $10\text{ mM}$   $\text{CaCl}_2$ . Aliquots were withdrawn from reaction mixtures at indicated times and trypsin activity was determined with the substrate Succinyl-Ala-Ala-Pro-Lys-AMC. Activity was expressed as percentage of the potential total activity, as determined on similar zymogen samples activated with enterokinase.

**B:** SDS-PAGE of the autoactivation of mutant S189D+A226G, reducing conditions, 15 % polyacrylamide gel. Lane 1, 0 min; lane 2, 5 hours; lane 3, 15 hours of incubation; lane 4, activated with enterokinase. Note that the mutant does not autolyse, probably because its chymotryptic activity is not large enough to cleave peptide bonds at the autolytic chymotrypsin cleavage sites.

were only comparable when the concentration of the mutant in the reaction mixture was at least one thousand-fold higher than that of trypsin.

Pancreatic serine proteinases can be also substrates of themselves; in a process called autolysis specific bonds on the surface of the molecules are cleaved successively by the others resulting finally in the loss of activity. The rate of autolysis depends on many conditions, especially on the concentration of the enzyme. In most chymotrypsin types, the first cleavage generally occurs in the autolysis loop of the molecule, resulting in a cleaved but fully active form that can be seen in the crystal structures. The autolysis loop of trypsins, as it contains no trypsin-specific site, is always intact. Autolysis of the S189D+A226G mutant under the conditions of these experiments was also not observed. This may be due to the lack of trypsin



**Fig. 16 Activation of chymotrypsinogen** (from Jelinek *et al.*, 2004)

1  $\mu\text{M}$  bovine chymotrypsinogen was incubated at 37 °C with the mutant enzymes S189D (16  $\mu\text{M}$ ) ( $\square$ ), S189D+A226G (14  $\mu\text{M}$ ) ( $\blacklozenge$ ) and wild-type trypsin (5 nM) ( $\Delta$ ) for 80 minutes. Aliquots were assayed every ten minutes for chymotrypsin activity that is represented as percentage of the total activity.

sensitive peptide bonds at the critical autolytic cleavage sites of chymotrypsin (Bódi *et al.*, 2001), on one hand, and to the low chymotryptic activity of the mutant on the other.

Benzamidine, a small reversible trypsin inhibitor that, similarly to the P1 Arg residue, forms a direct charge interaction with the Asp189 side-chain of trypsin (Marquart *et al.*, 1983) inhibited the S189D+A226G mutant with a  $K_i$  value similar to that of trypsin inhibition by benzamidine:  $K_i^{\text{S189D+A226G}} = 6.9 \times 10^{-6}$  M,  $K_i^{\text{trypsin}} = 6.3 \times 10^{-6}$  M. The wild-type level inhibition by benzamidine, however, indicates only a shift of the mutant S1 site towards a functionally more competent S1 structure. The result of the low  $K_i$  value similar to the one obtained with wild-type trypsin taken together with the poor  $K_m$  values of the double mutant are compatible with an S1 region of the mutant which can form a stable complex with benzamidine but not with the substrate. The reason for this is that contrary to the substrates, benzamidine of rigid structure and small size can adopt an orientation which is suitable for interactions with the mutant S1 site. In this respect the S189D+A226G chymotrypsin may be similar to the G226A rat trypsin mutant. The latter, relative to the wild-type enzyme, has a 30-fold higher  $K_m$  on oligopeptide substrates while only a 5-fold higher  $K_i$  on benzamidine (Craik *et al.*, 1985; Wilke *et al.*, 1991). X-ray crystallography showed that the mutant bound benzamidine in an orientation very different from that in the wild-type enzyme (Wilke *et al.*, 1991).

## 4.2 The S189D+A226G mutant structure

### 4.2.1 Introduction

The facts that i) the S189D+A226G chymotrypsinogen mutant undergoes autoactivation, ii) the active form has a trypsin-like specificity profile with a relatively high specific activity and that iii) the mutant activates chymotrypsinogen assumed a certain degree of structural integrity in its substrate binding region. The crystal structure of the S189D single mutant (Szabó *et al.*, 2003) shows that the S1 region contains severe deformations, mostly in the loop segments 185-195 and 217-224 including the 191-220 cystine. The Asp189 side chain at the bottom of the pocket is turned out to the solvent, presumably because contrary to trypsin, the S1 region of chymotrypsin is not suitable for the stabilization of such a partially buried charge. These structural deformations explained the poor activity and the lack of trypsin-like specificity of the S189D mutant, but did not provide any useful information concerning the possible structural effects of the A226G substitution. From the increased trypsin-like features of the S189D+A226G mutant, however, we concluded that the negative charge of Asp189 might be in a more available position for the positive charge of the P1 substrate side chain. To test this assumption, we determined the three-dimensional structure of the S189D+A226G mutant, which provided further insights into the structural basis of this unexpected specificity conversion.

### 4.2.2 The overall structure

The structure of the S189D+A226G rat chymotrypsin-B double mutant was determined at 2.2Å resolution. Data collection and refinement statistics are shown in **Table II**. Crystallization was conducted in presence of the small-molecule inhibitor benzamidine, which is not visible, however, in the structure. The activation domain including the site of substitutions is well defined, whereas residues 11-18 at the activation cleavage site and amino acids 147-150 in the cleaved autolysis loop were not visible. While chymotrypsins have a high tendency of autolytic cleavage at Tyr146, which does not inactivate the enzyme, the double mutant was stable against autolysis in solution, probably because of its trypsin-like specificity profile. However, under the conditions of crystallization (at very high protein concentration and pH 7.0, where specificity becomes more chymotryptic (see below)), autolysis was observed in the structure of the double mutant as well.

<b>Table II: Crystallographic data and refinement statistics</b>			
Resolution (Å) *	43.44 – 2.20 (2.32 – 2.20)		
Cell parameters	Space group $P2_1$ a = 34.7 Å, b = 64.4 Å, c = 44.2 Å, $\beta = 102.1^\circ$ , $\alpha, \gamma = 90^\circ$		
Number of observed reflections *	29867 (3843)		
Number of unique reflections *	9608 (1357)		
Completeness (%)*	98.5 (95.8)		
Mosaicity (°)	1.8		
Oscillation range per image (°)	1		
I/ $\sigma$ *	10.2 (2.3)		
R <sub>merge</sub> (%) <sup>†*</sup>	9.0 (44.5)		
R <sub>work</sub> (%) <sup>‡</sup>	27.6		
R <sub>free</sub> (%) <sup>‡</sup>	33.4		
R.m.s. bond length (Å)	0.005		
R.m.s. bond angles (°)	0.8		
No. of modeled atoms			
Protein	1679		
Solvent	67		
Ramachandran plot (by Procheck)			
Favoured (%)	84.5		
Additionally allowed (%)	14.4		
Generously allowed (%)	0.5		
Disallowed (%)	0.5		
Estimates of coordinate error (Å)			
Luzzati (Luzzati, 1952)	0.48		
DPI (Cruickshank, 1999)	0.32		
Average B-factors (Å <sup>2</sup> )	Protein **	Waters	Overall
	55.3	56.8	55.4

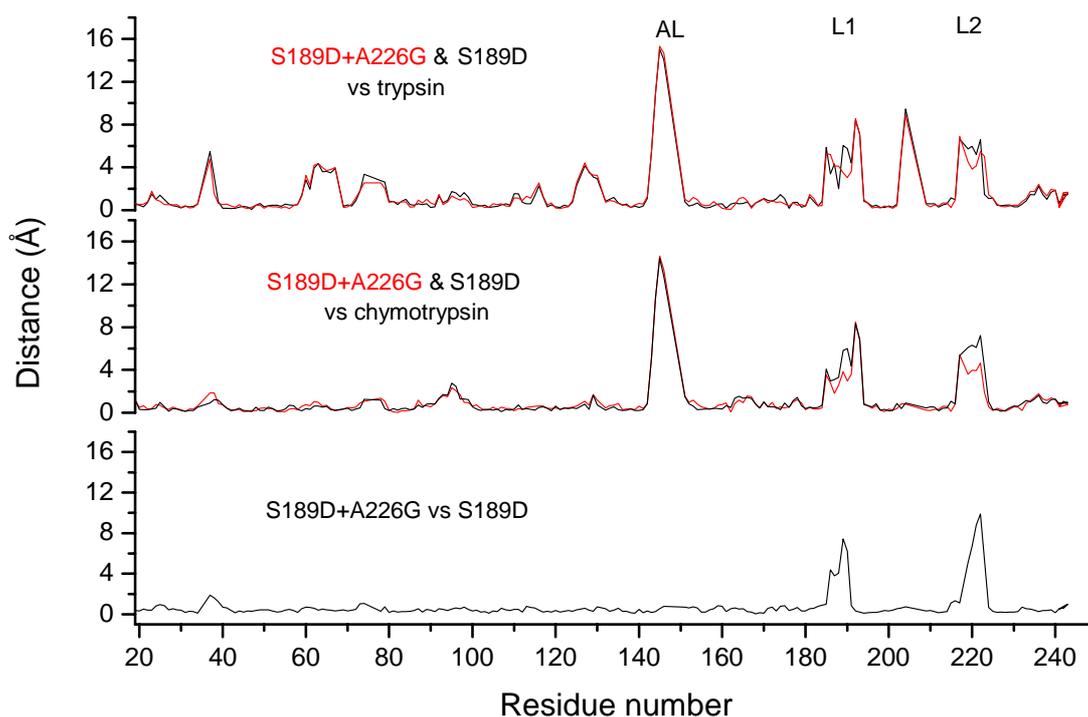
\* Values in parentheses indicate statistics for the highest resolution shell.

\*\* Isotropic residual B-factors without TLS contribution.

†  $R_{\text{merge}} = \frac{\sum |I_o - \langle I \rangle|}{\sum I_o} \times 100\%$ , where  $I_o$  is the observed intensity of a reflection and  $\langle I \rangle$  is the average intensity obtained from multiple observations of symmetry related reflections.

‡  $R \text{ factor} = \frac{\sum ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum |F_{\text{obs}}|} \times 100\%$ .

The removal of the methyl group at site 226 generated substantial rearrangements at the S1 region, compared to the S189D single mutant, while there were minor differences in other regions. When compared to wild type trypsin and chymotrypsin, substantial differences were found in the loops of the activation domain, loops L1 and L2, and especially in the autolysis loop. The equivalent C $\alpha$  distances measured after superposition of the structures are presented in **Figure 17**.



**Fig. 17** *C $\alpha$  distance comparison plots* (from Jelinek et al., 2007)

*C $\alpha$  distances of corresponding positions were compared using superimposed structures: the single and double mutant enzymes against wild type trypsin and chymotrypsin, and also S189D+A226G vs. S189D. Highlighted areas refer to the autolysis loop (AL) and loops L1 and L2.*

### 4.2.3 The S1 specificity site

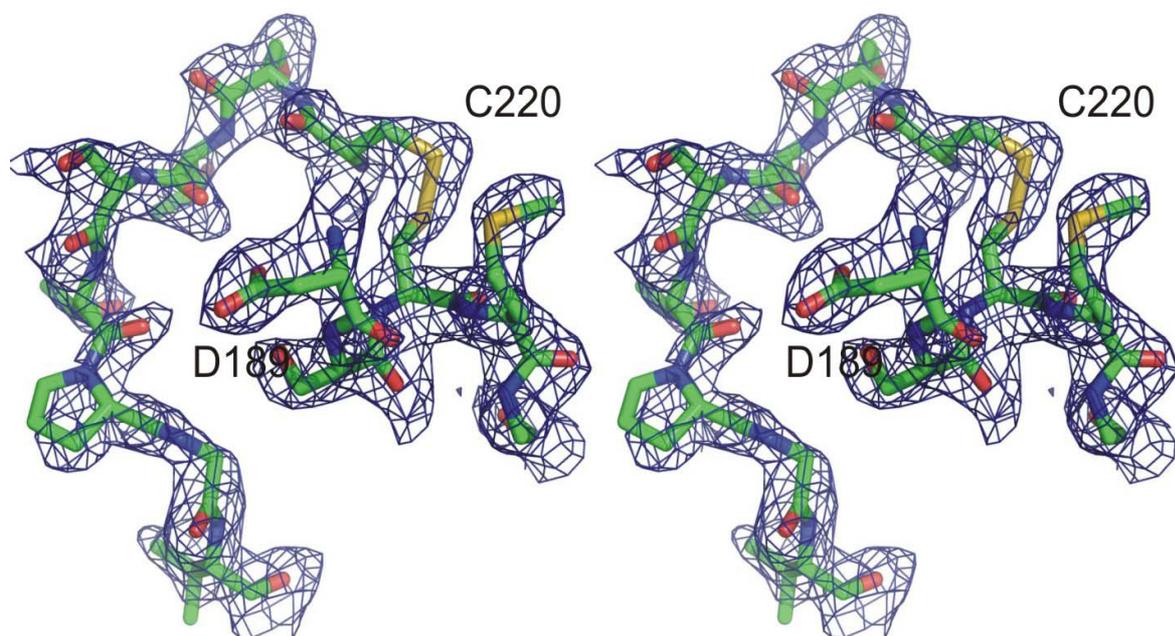
Comparison of the S189D+A226G double mutant structure to that of the S189D single mutant shows that rearrangements induced by the removal of the methyl group from position 226 in the S1 site were, in some respects, directed towards a wild type trypsin conformation (**Figure 18A and B**). In S189D chymotrypsin, the Ala226 side chain forms a van der Waals contact with the 191-220 cystine. This is the only interaction formed by the methyl group of Ala226, suggesting an important role in the rearrangements caused by the A226G mutation. The interaction appears to stabilize the cystine in a conformation which is substantially different from that in active enzymes. As a result, the cystine and the 216-218 segment of loop L2 bulge into the S1 pocket occluding it seriously (Szabó *et al.*, 2003). In the S189D+A226G mutant, where the Ala226 methyl- 191-220 cystine interaction is missing, the cystine moved out from the pocket, close to the wild type position.

The 191-220 cystine is an important determinant of the S1 site structure (Dementiev *et al.*, 2006), connecting loops L1 and L2 (and thereby the two sides of the S1 pocket), and has an identical conformation in trypsin and chymotrypsin. Analysis of the 191-220 cystine chi angles in the wild type and mutant enzymes according to Katz and Kossiakoff (1986) showed that the conformation of the cystine is right handed, except for the S189D+A226G mutant, where it is left handed. Neither wild type nor mutant 191-220 cystines fit well into any of the subclasses defined by Katz and Kossiakoff, and the mutants showed high variance in their chi values when compared to the wild type. Upon activation, the 191-220 cystine in chymotrypsinogen undergoes a conformational change from left handed to right handed, therefore the left handed 191-220 cystine in the S189D+A226G mutant is a chymotrypsinogen-like property.

The conformation of loops L1 and L2 of the double mutant was also reshaped, so that loop L1 became able to stabilize the negative charge of Asp189, which therefore turned towards the interior of the S1 site from its surface position in S189D chymotrypsin (**Figure 18B**). However, the new H-bonding partners that dissipate the half-buried charge of the Asp189 side chain, the hydroxyl group of Ser186 and the Ala185 main chain nitrogen atom, were still different from those in wild type trypsin. This contribute to the remaining distortions in loop L1 and, consequently, in loop L2. Thus, similarly to the S189D single mutant, part of loop L2 still bulged into the S1 pocket so that residue 217 was in the way of a P1 residue.

The structure of the S189D+A226G mutant also retained the chymotrypsinogen-like features at the S1 site observed in the S189D chymotrypsin structure, such as the oxyanion hole that was not formed and also the side chain of Asp194 that formed a salt bridge with His40 instead of the N-terminus.

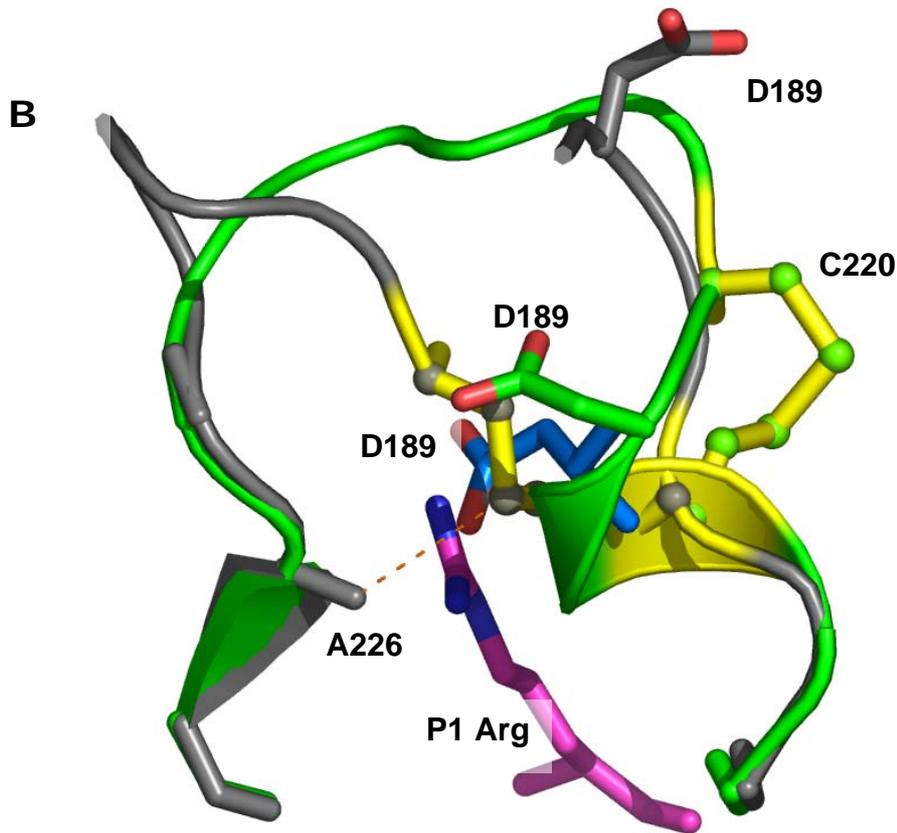
**A**



**Fig. 18 S1 specificity site structures of mutants S189D (gray) and S189D+A226G (green)**  
(from Jelinek et al., 2007)

**A:** Stereo view of the final 2Fo-Fc electron density maps of S189D+A226G, from 189 to 193, and from 220 to 227 regions contoured at  $1\sigma$  level (blue).

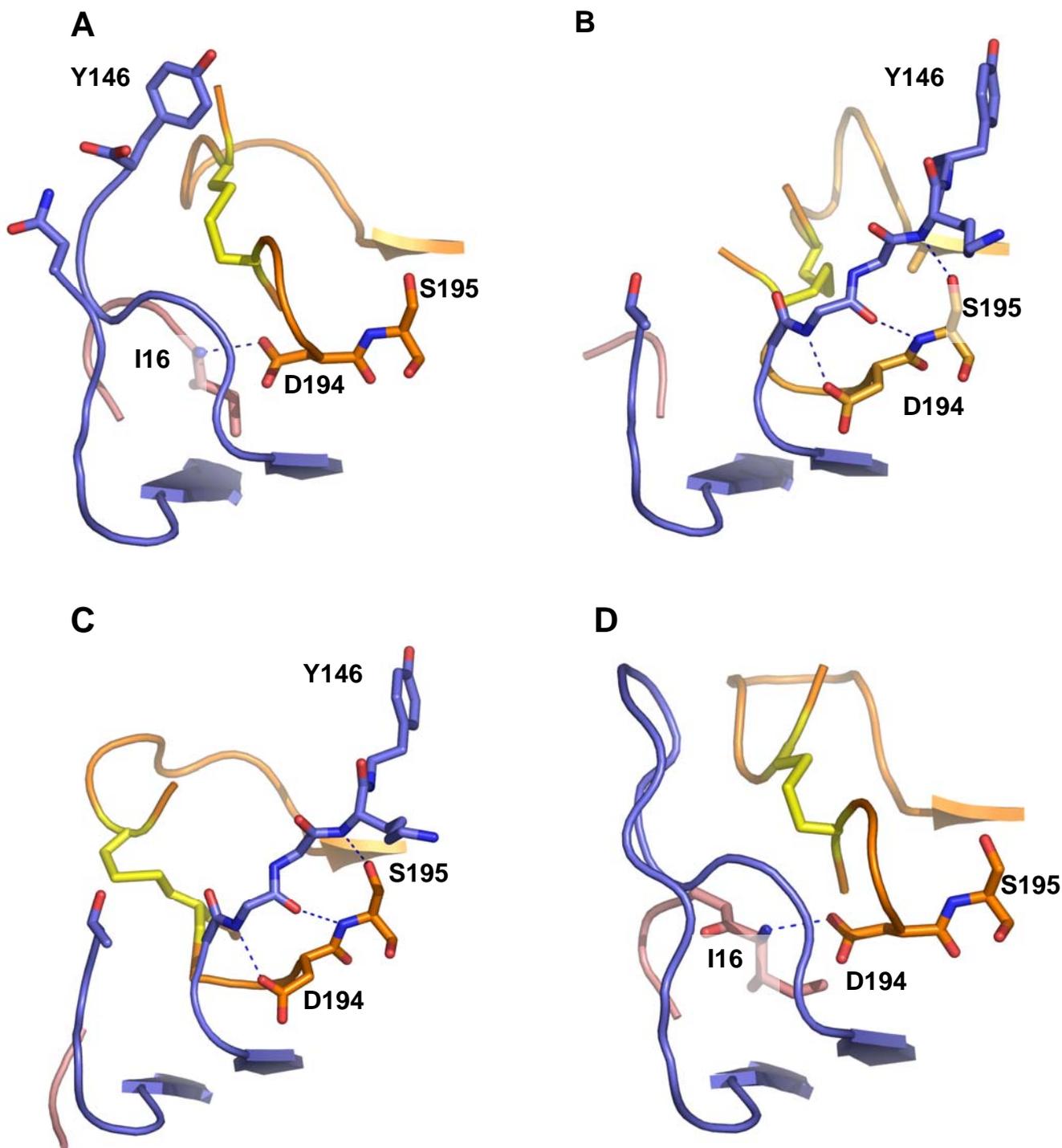
**B, on the next page:** The following parts of the S1 sites are shown and compared: the backbone from 189 to 193, and from 220 to 227; the 191-220 cysteine (gold in S189D, yellow in S189D+A226G); residues 189 and 226. Asp189 (blue) and P1Lys (magenta) positions in wild type trypsin are also shown. The Ala226 side chain of S189D is in Van der Waals interaction (distance is  $3.9\text{\AA}$ , blue dotted line) with the 191-220 cysteine connecting the two loops.



Crystal contacts may also influence the X-ray structure of proteins. The S189D+A226G and S189D mutant chymotrypsins crystallized in different space groups, hence the crystal environment is different in the two crystal structures. Since the overall fold is almost identical in the two structures (except for the 185-191 and 219-223 regions) we can assume that the general structure is mostly unaffected by the crystal contacts. In the S189D+A226G mutant residue range 219-223 are stabilized by two hydrogen bonds to a neighboring protein molecule (Leu123<sub>N</sub> - Ser221<sub>O</sub> and Ile47<sub>O</sub> - Ser221<sub>O<sub>γ</sub></sub>), while the 185-191 region is free of crystal contacts. In the S189D mutant structure, both regions are involved in crystal contacts through the Phe114<sub>N</sub> - Ser223<sub>O<sub>γ</sub></sub>, Glu49<sub>O<sub>ε1</sub></sub> - Ser223<sub>N</sub>, Glu49<sub>O<sub>ε2</sub></sub> - Thr222<sub>N</sub> and Glu49<sub>O<sub>ε2</sub></sub> - Gly187<sub>N</sub> hydrogen bonds. In these chymotrypsin mutants these flexible loops adopt various conformations which might be explained on one hand by the structural differences genuinely attributed to the A226G mutation. On the other hand, there is little free energy difference between the alternative conformations and, unlike the rest of the structure; they can be easily influenced by the environment. In the crystalline form this means that weak crystal contacts are sufficient to alter the conformation whereas in solution changes in pH, ionic strength etc. may play a similar role.

#### 4.2.4 The activation domain

The diagram of C $\alpha$  distances between the mutant and wild type chymotrypsins showed the highest differences in the region of the activation domain, especially in the autolysis loop (**Figure 17**) which is cleaved in both mutants and the wild type crystal structures. In wild type enzymes, this loop makes contacts with loops L1 and L2 at the 191-220 cysteine, and also with the N-terminal residues 16-17 (**Figure 19**). Two H-bonds are formed between Lys143<sub>O</sub> - Ile16<sub>N</sub> and Lys143<sub>N</sub> - Met192<sub>O</sub>. All these contacts seem to be crucial to the stabilization of the active conformation of the wild type enzymes. In both the S189D and the S189D+A226G mutants, however, the autolysis loop moved towards the substrate binding site, and occupied the P2-P2' substrate positions. Contacts with the 191-220 cysteine and the N-terminus were lost, and the new conformation of the autolysis loop was stabilized by five H-bonds to amino acids 194 and 195, as well as by van der Waals contacts with residues 40, 41, 57, 58, 217, 218. This position of the autolysis loop sterically blocked the formation of the oxyanion hole, and brought the amide nitrogen of Lys145 in H-bond distance from the active hydroxyl group of Ser195. These structural features of the activation domain showed that both the single and the double mutants crystallized in a conformation which is completely incompatible with canonical binding which is presumably a prerequisite for efficient activity (Fodor *et al.*, 2006; Radisky *et al.*, 2006; Fuhrmann *et al.*, 2006; Liu *et al.*, 2006). The identical zymogen-like features of the two structures indicate that the equilibrium of the active and inactive forms was disturbed in both cases by the mutations.



**Fig. 19: Activation domain structures, A: chymotrypsin, B: S189D, C: S189D+A226G, D: trypsin (from Jelinek et al., 2007)**

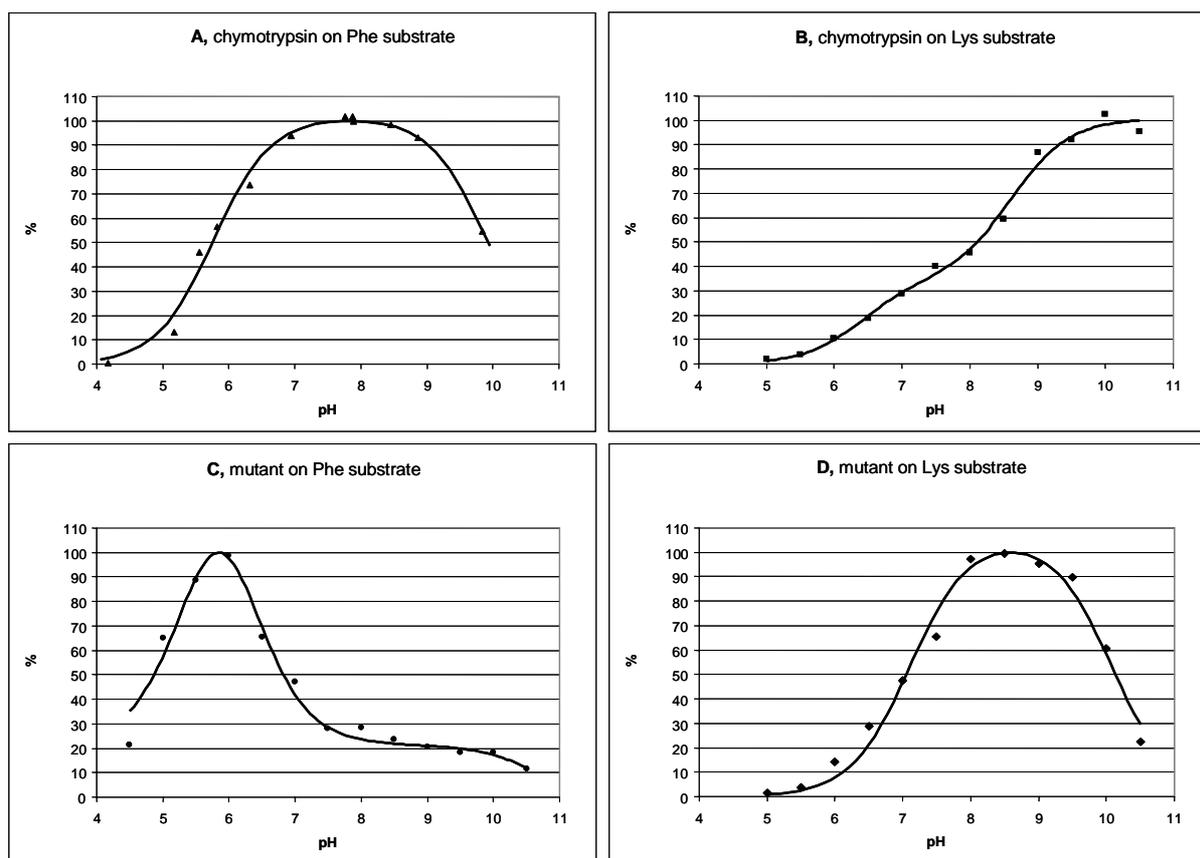
The following parts of the activation domains are shown and compared: loop L1 from 191 to 195 and sites 220 to 227 (orange); the 191-220 cystine (yellow); the autolysis loop (blue) and the N-terminus (salmon). Dotted blue lines represent H bonds.

### 4.3 pH dependence experiments

While several changes in the structure of the double mutant might be favorable for the activity (e.g. the rearrangement of Asp189 and the 191-220 cystine), the serious distortions in the S1 region structure of both the single and the double mutants are difficult to reconcile with enzyme activity measured in solution (Venekei *et al.*, 1996b; Jelinek *et al.*, 2004). It is known that the S1 regions of wild type trypsin and chymotrypsin are sensitive to changes in the environment. For example, the effect of changes in the pH on enzyme activity is due primarily to alterations in this region: to the loss of the salt bridge between the Asp194 and the N-terminal amino group ( $pK_b$ ) which disrupts the oxyanion hole, and to the protonation of His57 ( $pK_a$ ). Previous studies (Fersht, 1972; Verheyden *et al.*, 2004) have found that there is a pH dependent equilibrium of the active and inactive states of wild type trypsin and chymotrypsin, depending on the transition between the active and the zymogen-like structure of the S1 region. Thus, altered parameters of the pH profile of the mutants can also be useful indicators of perturbations in the solution structure of the S1 region due to amino acid substitutions and perhaps can be informative about the congruence of the measured activities and the zymogen-like crystal structures. Therefore we compared the influence of pH on the activity of the S189D+A226G double mutant and that of wild type chymotrypsin using Succinyl-Ala-Ala-Pro-Lys-AMC (succAAPK-AMC) and Succinyl-Ala-Ala-Pro-Phe-AMC (succAAPF-AMC) tryptic and chymotryptic fluorogenic substrates.

On its preferred succAAPF-AMC substrate, wild type chymotrypsin showed wide activity range around a pH optimum of 7.9 (**Figure 20 A, E**) with  $pK_a=5.8$ ,  $pK_b=9.9$ , and a slightly visible  $pK_{a2}$  of around 6.9. At the same time, on the tryptic substrate succAAPK-AMC its pH optimum shifted upward to above 9.5, and in the acidic side two pH steps are evident ( $pK_{a1}=6.4$  and  $pK_{a2}=8.6$ ) (**Figure 20 B**).

In the case of the S189D+A226G double mutant on the succAAPF-AMC substrate (**Figure 20 C**), the pH profile had a narrower range of activity, and showed a significant acidic shift in the parameters relative to the wild type enzyme with two pH steps on the basic side (apparent values of pH optimum,  $pK_a$ ,  $pK_{b1}$  and  $pK_{b2}$  were 5.9, 4.9, 7.0 and 8.7, respectively). The activity on tryptic substrate also showed narrower range of pH optimum relative to wild type chymotrypsin but, with a maximum at pH 8.5 (**Figure 20 D**). The higher maximum was due solely to the basic shift in  $pK_a$  from 5.8 to 7.1. The  $pK_b$  (10.1) was the same as that of the wild type enzyme on the Phe substrate.



E						
Enzyme	Substrate	pK <sub>a</sub>	pK <sub>a2</sub>	pH <sub>opt</sub>	pK <sub>b</sub>	pK <sub>b2</sub>
WT chymotr.	Phe ▲	5.8		7.9	9.9	
WT chymotr.	Lys ■	6.4	8.6	9.5<	-	
S189D+A226G	Phe ●	5.5		5.9	6.2	10.6
S189D+A226G	Lys ◆	7.1		8.5	10.1	

**Fig. 20: pH dependency experiments** (from Jelinek et al., 2007)

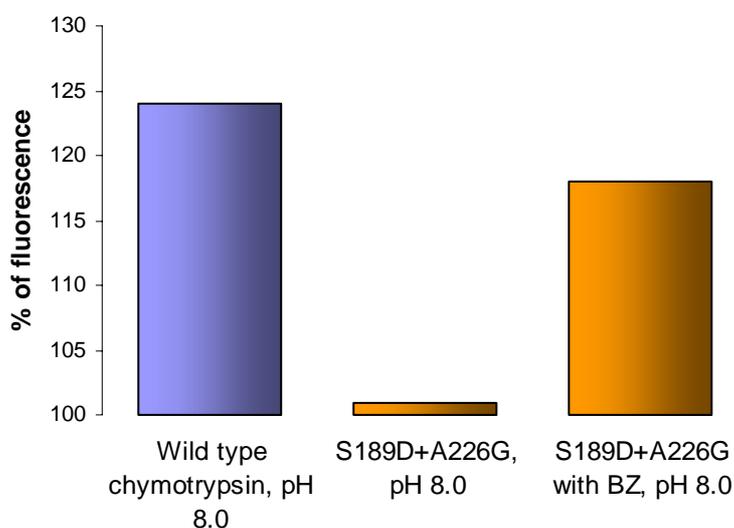
**A:** pH dependence of the activity of the S189D+A226G mutant and wild type chymotrypsin was measured with 400  $\mu$ M succAAPKAMC and succAAPFAMC fluorogenic substrates, in 40 mM buffers of Na-Acetate (pH4.0-pH5.0), MES (pH5.5-pH6.5), MOPS (pH7.0-7.5), Tris-HCl (pH8.0-pH8.5) CHES (pH9.0-pH9.5) and CABS(pH10.0-10.5). The reaction buffers contained 10 mM CaCl<sub>2</sub> and 0.1 M NaCl, measurements were carried out at 37°C, enzyme concentration was 0.02  $\mu$ M. Activity is represented as the percentage of the calculated maximum (=100%) activity.

**A:** chymotrypsin on Phe substrate (from Venekei et al., unpublished); **B:** chymotrypsin on the Lys substrate; **C:** mutant on Phe substrate; **D:** mutant on Lys substrate.

**E:** Calculated pK and pH<sub>opt</sub> values for the titration curves. Note that in the case of the mutant on Phe substrate, the calculated pK<sub>a</sub> and pK<sub>b</sub> values do not match the apparent pK values, see the discussion for an explanation.

Taken together, the pH profiles and the maximum activity values on the P1 Lys and Phe substrates show that, unlike wild type chymotrypsin, the mutant has either a trypsin- or chymotrypsin-like specificity in a pH dependent manner, such that the trypsin-like specificity at pH 8.0 becomes chymotrypsin-like at pH 6.0.

The transition between the active and inactive forms described above can be followed by detecting changes in the internal tryptophan fluorescence of the enzyme (Verheyden *et al.*, 2004). To investigate this conformational transition of the mutant, we compared the fluorescence intensities of the mutant and wild type chymotrypsin at two distinct pH points. Measurements at pH 10.5 and pH 8.0 showed that chymotrypsin exhibits a 24% gain in fluorescence upon the transition to the active form. This change is in the range as that obtained in our later pH-jump experiments with trypsin (Tóth *et al.*, 2006). The double mutant showed only 1% increase in fluorescence intensity, suggesting strongly that even at the optimal pH, most of the double mutant molecules are in their inactive form, supposedly the one seen in the crystal structure.



**Fig.21 Measurements of the pH dependence of internal tryptophan fluorescence**

*The percentage gains of fluorescence intensities upon the pH 10.5 to pH 8.0 transition were compared. Intensities measured at pH 10.5 were taken as 100%.*

In the presence of benzamidine, the increase in fluorescence of the mutant raised to 18%, suggesting that the positive charge of benzamidine can induce the formation of the active conformation. This is in good agreement with the results of the pH dependence of activities. As the amount of the increase of fluorescence upon the transition of all the mutant molecules

to the active form, and also the possible effect of the bound benzamidine on the internal fluorescence of the mutant are unknown, further experiments are needed to measure quantitatively the effect of benzamidine binding.

## 5. Discussion

In the first part of my doctoral work (Jelinek *et al.*, 2004) we presented the kinetic results of mutants demonstrating that site 226 has an important role in converting the specificity of chymotrypsin to that of trypsin. In the case of the S1+A226G mutant prepared according to the complex model, removal of the single methyl group resulted in a considerable rise in chymotrypsin-like activity, moreover it annulled the results of the S1 mutant in conferring trypsin-like specificity to chymotrypsin. These changes are hard to interpret in the absence of a structure of the mutant. According to the specificity profiles it can be supposed that similarly to the S189D mutant with a chymotrypsin-like profile, Asp189 is not available for an electrostatic interaction with a positively charged P1 residue. As glycine residues can increase flexibility, the gain of activity on the large chymotrypsin substrates might be explained on one hand with a higher flexibility of the distorted S1 site, and, on the other hand, with the lack of the charge at the base of the pocket.

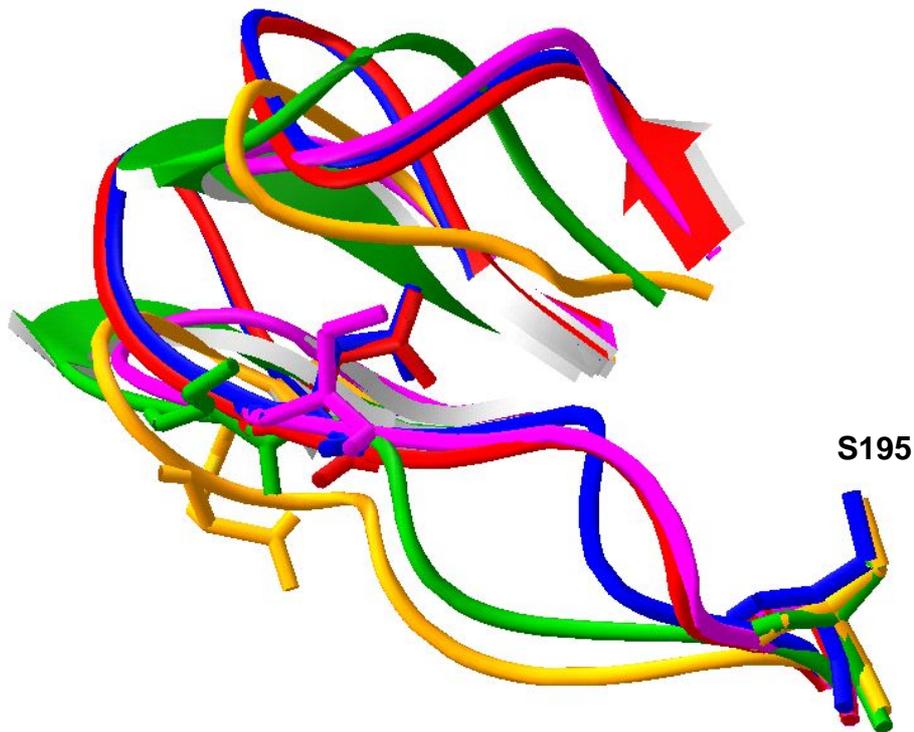
Similarly to the chymotrypsin to elastase conversion experiment, the mutant according to the simple model showed the better results. In fact, the A226G substitution in this case can be regarded as a conversion of the single S189D chymotrypsin-B mutant to a single S189D chymotrypsin-A mutant. The chymotrypsin-B mutants were not appropriate counterparts of the trypsin to chymotrypsin conversion mutants, since site 226 is a conserved Gly in trypsins. The fact that the G226A mutation in trypsin reduces the activity on trypsin substrates by 3-4 orders of magnitude (Craik *et al.*, 1985) also documents the importance of Gly replacing Ala at site 226. Therefore the S189D+A226G double mutant can be considered as a single S189D chymotrypsin-A mutant that is the true counterpart of the D189S trypsin single mutant. Our unexpected finding was that in contrast with the chymotrypsin to trypsin conversion experiments, the single mutation S189D converts the specificity profile of chymotrypsin to that of a trypsin-like protease with enough catalytic potential even for autoactivation, whereas the exchange of the whole S1 site failed to confer trypsin-like specificity. This seems to support the view (also suggested by Hung and Hedstrom, 1998) that there may be different strategies and routes to convert trypsin to a chymotrypsin-like protease and *vice versa* and that the key residues involved in substrate discrimination and therefore to mutate in order to interchange the specificities of these proteases may not be identical in the two structures.

In the second part of my doctoral work (Jelinek *et al.*, 2007), we determined the structure of the trypsin-like S189D+A226G mutant. The structural alterations in the S189D mutant upon the A226G substitution can be attributed to the loss of the interaction between the methyl group of Ala226 and the 191-220 cystine. The rearrangements that underlie the increased activity of the double mutant include the shift of the cystine and parts of loop L1 and L2 to positions which are closer to those in wild type trypsin, and the stabilization of the Asp189 residue in an internal conformation, closer to the S1 pocket (**Figure 17**). At the same time, serious distortions remain in the structure of two S1 site loops and in the nearby elements of the activation domain, including certain zymogen-like features.

A comparison of the main chain conformations at the sequence 189-195 of trypsin-like proteinases (**Figure 10**) suggests that the specific position of Asp 189 is a direct consequence of the conformation of this part of loop L1 that is very similar in all trypsin-like enzyme structures, and that the rest of loop L1 can be subject to the evolutionary divergence observed in these structures. In trypsinogen, however, where this region has a different conformation as the oxyanion hole is not formed, Asp189 is in an identical position with that of the active enzyme (**Figure 22**). This suggests that rather the conformation of the whole loop L1 defines the position of Asp189, together with the stabilizing interactions of loop L2, that are also partly provided by different sites in the zymogen and the active enzyme. Thus, the comparison of the S1 site of trypsin and trypsinogen shows two different possible setups for the stabilization of the Asp189 charge in the similar position. In contrast to trypsinogen, the position of site 189 in chymotrypsinogen is not identical with that of the active enzyme. The sequence 189-195 undergoes a larger conformational change upon activation that results in a 2.5Å approach of site 189 towards the active serine. In the double mutant structure, Asp189 is in a chymotrypsinogen-like position but, contrary to the single mutant, the charge points toward the inside of the pocket. A dissipation of the Asp189 charge identical to that of trypsin is not possible, since the S1 site chymotrypsin cannot provide all the necessary interaction partners. As an induced fit mechanism that results in the active form, a 2.5Å approach similar to the activation process might occur in the double mutant when the dissipation of the Asp189 charge is provided by a positively charged P1 residue. The next goal of the chymotrypsin to trypsin conversion will be to provide proper charge dissipation for the Asp189 charge in the absence of the substrate, as well.

A somewhat similar induced fit binding process was observed in the serine proteinase hepatocyte growth factor activator (Shia *et al.*, 2005). The apo structure of this wild type enzyme shows that the access of the P1 sidechain to Asp189 is blocked by residues 215, 216

and 217 of loop L2, which was considered to be an enzymatically non competent arrangement. However, the enzyme is fully active in solution, and when complexed with the first Kunitz domain of its physiological inhibitor, the structure shows a pocket with identical conformation to other members of the S1 family.



**Fig. 22 Comparison of the substrate binding pocket of the double mutant with that of trypsin, chymotrypsin and their zymogen forms**

*The crystallographic structures of the following enzymes and zymogens were superimposed:*

- 1. Bovine trypsin, red (PDB code: 1EJM)*
- 2. Bovine chymotrypsin, pink (PDB code: 1CBW)*
- 3. Bovine trypsinogen, blue (PDB code: 1TGC)*
- 4. Bovine chymotrypsinogen, green (PDB code: 1CHG)*
- 5. The double mutant S189D+A226G, yellow (PDB code: 2JET)*

*Parts 183-195 and 216-226 are shown as ribbons, residues Asp189 and Ser195 are shown as sticks.*

The changes in the conformation of the single mutant upon the A226G exchange can have a direct effect on the binding of the P1 residue of the substrate, explaining the significant shift in specificity towards trypsin. However, the crystal structure shows a number of structural features which do not allow canonical substrate binding and catalysis, and therefore

seem to be inconsistent with the observed three orders of magnitude increase in catalytic efficiency. An explanation might be that the binding of the P1 residue can have a direct effect on the conformation of the mutant, as suggested by the effect of benzamidine binding on the change in the pH-dependent tryptophan fluorescence of the mutant. The S1 region, as part of the activation domain is known to be sensitive to external effects which can easily be more pronounced in mutant enzymes (Szabó *et al.*, 1999). In solution, the active and the zymogen-like inactive forms are in equilibrium, and even at the optimal pH, only 90% of wild type chymotrypsin is in its active form (Fersht, 1972). Similarly to the changes in the pH, mutations at the S1 site might also have an effect on this equilibrium by disturbing the structure of the activation domain, as suggested by both the fluorescence measurements and the zymogen-like structures of the single and double mutants. None of the determined structures of trypsin to chymotrypsin conversion mutants (Perona *et al.*, 1995; Szabó *et al.*, 1999) showed zymogen-like features, contrary to the two structures of the chymotrypsin to trypsin conversion mutants. This suggests that the equilibrium of the activation domain of chymotrypsin might be more sensitive to mutations of the S1 site, when compared to trypsin.

Based on detailed analysis of the pH dependence of refolding and the knowledge of the structure, the pK values can provide some information on the structure of the S1 region in solution, such as the protonation state of the catalytic His57, and the existence and strength of the salt bridges of Asp194  $\gamma$ -carboxyl with either His40 or the Ile16 N-terminal amine (the assigned pK values of wild type chymotrypsin are:  $pK_{a1}=5.8$ ,  $pK_{a2}=7.0$  and  $pK_b=9.9$ , respectively (Fersht and Requena, 1971; Fersht, 1972; Verheyden *et al.*, 2004)). Of the latter two, the Asp194  $\gamma$  carboxyl – Ile16 N-terminal amine interaction is characteristic to the active enzyme, while the Asp194  $\gamma$  carboxyl – His40 interaction is a zymogen like state of Asp194.

In the case of wild type chymotrypsin, two of the titration steps,  $pK_{a1}=5.8$  and  $pK_b=9.9$ , are clearly visible (**Figure 20 A**), when the titration is followed by activity measurement on the chymotryptic substrate. The  $pK_{a2}$  of His40 is also slightly visible. When the tryptic substrate, succ-AAPK-AMC is used  $pK_{a1}=6.4$  is probably due to the titration of the Asp194-His40 interaction (Fersht, 1972), and the  $pK_a=5.8$  of the titration of His57 is not visible due to the very low activities below pH 6.0 (**Figure 20 B**). The  $pK_{a2}$  at pH 8.6 in this titration curve results from the deprotonation of the P1 Lys residue (Gráf *et al.*, 1988), which makes it better accepted by the hydrophobic S1 site of chymotrypsin. Up to pH=10.5, this gain of activity seems to compensate the loss of it due to the titration of the Asp194 - Ile16 N-terminal salt bridge, thus the titration step of wild type chymotrypsin at pH~10 is not seen.

In contrast, deprotonation of the P1 Lys residue is not favorable for the S189D+A226G double mutant, therefore a  $pK_b=10.1$  appears (**Figure 20 D**), which corresponds to the  $pK_b=9.9$  of the wild type enzyme measured on the Phe substrate above. Thus, at variance with the structural data, this titration indicates the existence of the Asp194  $\gamma$ -carboxyl – Ile16 N-terminal amine salt bridge, and the  $pK_b$  value similar to that of chymotrypsin also suggests that the interaction might be as strong as in the wild type enzyme.

The S189D+A226G mutant on the succAAPF-AMC chymotryptic substrate shows a narrow titration curve with unique  $pK$  values (**Figure 20 C**). It is tempting to interpret  $pK_{b1}=6.2$  as the titration of the Asp189 side chain, which – when losing its charge – becomes much more favorable for the activity on the substrate with hydrophobic P1 residue. As the constraint of stabilizing the negative charge disappears, the pocket of the mutant might become very similar to that of wild type chymotrypsin, therefore only a few percent of the molecules with an uncharged Asp189 could produce a relatively high chymotrypsin-like activity. When lowering the pH, this effect could then give rise to a high gain of activity but it is limited by the loss of activity due to the parallel titration of His57 ( $pK_a=5.5$ ). It seems reasonable to suppose that due to their closeness, the two titrations influence each other. Thus the apparent  $pK_a$  and  $pK_{b1}$  of the curve are combinations of the two titration effects and, therefore, do not mean a 50% titration of the residues. Data points between pH 7.5 and pH 10.0 might suggest an additional titration with a  $pK$  of  $\sim 8.8$ , which is hard to interpret but if it is a downshifted  $pK$  of the titration of the Asp194  $\gamma$  carboxyl - Ile16 N-terminal amine salt bridge, then it indicates a weaker interaction than in the wild type enzyme, or when the P1 substrate residue is Lys (see above). This interpretation implies residual activity following the titration of the salt bridge that was in turn also observed – at a lower extent, 12% - in wild type chymotrypsin, too (Fersht, 1972).

This latter comparison, together with the fluorescence experiments, shows a stabilization of the salt bridge and the S1 region in the double mutant by the positive charge of a P1 residue. A similar effect was observed when trypsinogen was complexed with BPTI: the S1 region rearranged from the zymogen to an active conformation, which also included the formation of the salt bridge when the Ile-Val dipeptide was also added, since the Ile16 N-terminus is not available in trypsinogen (Bode *et al.*, 1978). In the case of the S189D+A226G double mutant, an attraction between the P1 Lys and the slightly displaced Asp189 might induce similar rearrangements.

Taken together our data offer some structural explanation for the effect of the A226G substitution on the activity and specificity of the S189D chymotrypsin mutant. Whereas the

crystal structure shows that, of the constituents of the S1 region, only the Asp189 and the 190-220 cystine are in better positions for activity but others are in very unfavorable ones, the pH profile of activity and the pH dependence of the internal fluorescence indicate alternative structures in solution that seem to be induced by the substrate or inhibitor binding and are dependent on the P1 residue of the substrate.

Thus the A226G substitution, on one hand, induces in itself rearrangements of the inherently flexible S1 region of the S189D chymotrypsin mutant towards a trypsin-like conformation (seen in the crystal structure). On the other hand, in solution it appears to permit substrate- or inhibitor-induced conformational states with enhanced trypsin-like activity, while such changes may be restricted in the S189D single mutant.

As the kinetic results demonstrated, when within the chymotrypsin scaffold, the S1 site of chymotrypsin is more appropriate to stabilize the negative charge of Asp189 than the S1 site exchanged to that of trypsin. We also conclude that Asp189 is indeed a key specificity determinant. However, the stabilization of the charge requires a unique interaction system that is not provided in the S1 site of chymotrypsin. To confer trypsin-like specificity to chymotrypsin, one route could be the formation of these stabilizing interactions by both the exchange of loops L1 and L2 to that of trypsin and the involvement of more extended structures of the chymotrypsin scaffold. Alternatively, one or a few more key exchanges like the A226G mutation might result in an S1 site that can, on one hand, provide the appropriate interaction partners for Asp189 and, on the other hand, would fit well into the parent chymotrypsin scaffold. Careful design and selection is required here, since our results strongly suggest that apart from causing possible distortions, mutations at the S1 site can also affect the equilibrium between the active and the zymogen-like inactive forms of pancreatic serine proteinases.

## 6. Summary

My doctoral work focused on the structural background of the S1 site substrate specificity in pancreatic serine proteinases. The S1 region of these enzymes include the substrate binding pocket, a cavity composed of residues 185-195, 213-223 and 226-228. The substrate binding pockets evolved to accommodate different substrate residues. Trypsin, with the negative charge of Asp189 at the base of the pocket prefers Arg and Lys, while the less polar pocket of chymotrypsin with Ser189 has the highest affinity for Tyr, Phe and Trp substrate residues. The shallow pocket of elastase is most suitable for Ala, Leu and Val because of the occlusion at the entrance of the pocket by Val216 and Thr226.

In a previous successful attempt to convert trypsin to a chymotrypsin-like protease, fifteen residues of the substrate binding region of trypsin were replaced with the corresponding ones in chymotrypsin. This suggests a complex mechanism of substrate recognition instead of a relatively simple one proposed in early crystallographic studies that involve only sites 189, 216 and 226. However, the further trypsin to elastase, chymotrypsin to elastase and chymotrypsin to trypsin conversion experiments carried out according to the complex model resulted in nonspecific proteases with low catalytic activity. Chymotrypsin used in the latter studies was of B-type, containing an Ala residue at position 226. Trypsins, however, contain a conserved Gly at this site. The substantially decreased trypsin-like activity of a G226A trypsin mutant also suggested a specific role for this site in substrate specificity, and raised doubts about the results obtained in the chymotrypsin to trypsin conversion experiments.

We investigated the role of site 226 by introducing the A226G substitution into chymotrypsin to trypsin conversion mutants which were constructed according to both the simple and the complex model of specificity determination: i) A226G into the S189D chymotrypsin-B single mutant with a slight chymotrypsin-like specificity profile and ii) A226G into a multiple-substituted chymotrypsin-B to trypsin mutant named S1 which has a slight trypsin-like specificity profile. The kinetic properties of the mutants were determined on trypsin and chymotrypsin substrates. The results show that the A226G substitution in the S1 mutant increased the chymotrypsin-like activity while the trypsin-like activity did not change, namely it annulled the results of the S1 mutant in conferring trypsin-like specificity to chymotrypsin. In contrast, this substitution in the S189D chymotrypsin mutant resulted in a 100-fold increase in trypsin-like activity and a trypsin-like specificity profile. Its specificity

profile was further tested on a competing oligopeptide substrate library where the mutant showed a profile highly similar to that of trypsin, with a considerable Arg over Lys preference. Additionally, the S189D+A226G mutant is the first trypsin-like chymotrypsin that undergoes autoactivation and is capable to activate chymotrypsinogen, which are exclusive properties of trypsin among pancreatic serine proteases. Taken together, the S189D+A226G mutant is the best “trypsin” so far in the line of trials to convert chymotrypsin to trypsin, aiming a more complete understanding of the structural basis of substrate specificity in pancreatic serine proteases.

From the increased trypsin-like features of the S189D+A226G mutant we concluded that the negative charge of Asp189 might be in an available position for the positive charge of the P1 substrate side chain. To test this assumption, we determined the three-dimensional structure of the S189D+A226G mutant, at 2.2 Å resolution. Crystallization was conducted in presence of the small inhibitor benzamidine, which is not visible, however, in the structure. The removal of the methyl group at position 226 caused significant rearrangements at the S1 site, relative to S189D chymotrypsin. This is mainly due to the elimination of the interaction between A226 and the 191-220 cystine present in S189D chymotrypsin, which resulted in the movement of the cystine out of the substrate binding pocket to a position that is much closer to that of wild type enzymes. This rearrangement allows a new conformation of the Asp189 residue which is now oriented towards the interior of the specificity site and occupies a position closer to that in trypsin. At the same time, serious distortions remain in the structure of two S1 site loops and in the nearby elements of the activation domain, including certain zymogen-like features. Changes in the internal tryptophan fluorescence upon decreasing the pH from 10.5 to 8.0 suggest that the S189D mutation disturbed the equilibrium between the active and the inactive zymogen-like forms, and the zymogen-like form seen in the crystal structure is also abundant in the pH-optimum solution. Binding of benzamidine can induce the formation of the active form of the mutant at pH 8.0, and the pH profile of activities suggest that the conformation of the S1 region of the mutant is influenced also by the nature of the P1 residue of the substrate. Thus the A226G substitution increased tryptic activity by approximating the structure to that of the wild type trypsin and by allowing other, substrate or inhibitor induced changes which may be restricted in the S189D single mutant.

## 7. Összefoglalás

Doktoranduszi kutatómunkám során azt vizsgáltam, hogy mely szerkezeti elemek felelősek a pankreatikus szerin proteázok S1 szubsztrátkötőhely-specifitásának kialakításáért. A pankreatikus szerin proteázok szubsztrátkötő zsebe az evolúció során különböző oldalláncok befogadásához adaptálódott. A tripszin a pozitív töltésű Arg és Lys, míg a kimotripszin a Tyr, Phe és Trp szubsztrát oldalláncok mellett hasít, az elasztáz pedig a kisebb, apoláros oldalláncokat (Ala, Val, Leu) részesíti előnyben. Az enzimek szubsztrát preferenciája és a szubsztrátkötő helyek szerkezetének összehasonlítása alapján a hetvenes években kidolgoztak egy szubsztrát specifitási modellt, melyből azóta tankönyvi alappélda lett. A modellben három aminosav oldallánc (189, 216, 226) határozza meg az enzim specifitását: a tripszin kötőzsebének alján található Asp189 negatív elektrosztatikus potenciálja képes a szubsztrát pozitív töltésű oldalláncának stabilizálására, az elasztáz kötőhelyét a zseb bejáratánál található Val216 és Thr226 oldallánc szűkíti le (tripszinben mindkettő Gly). A kimotripszin a 189-es pozícióban szerint tartalmaz, kevésbé poláros kötőzsebe a nagyobb, apoláros oldalláncoknak kedvez. A kimotripszin-A esetében a 216 és 226-os aminosav Gly, míg a kimotripszin-B a 226-os helyen egy alanint tartalmaz, mely enyhén befolyásolja a specifitási profilt.

A specifitás szerkezeti hátterét tovább vizsgáló helyspecifikus mutagenézis kísérletekben az enzimek specifitását próbálták egymásba alakítani oly módon, hogy a kiindulási enzimben különböző kombinációkban cseréltek ki aminosavakat a cél-enzim szekvenciájának megfelelő aminosavakra. A tripszin-kimotripszin átalakítás során az egyedüli Asp189Ser csere nem hozott eredményt, csak 14 további aminosav kicserélésével sikerült a tripszin specifitását kimotripszin szerűvé alakítani. Az eredmények alapján kidolgozott komplex szubsztrát-specifitási modellben a kötőhelytől távolabb eső oldalláncok is szerepet kapnak szubsztrát diszkriminációban. Az új modell alapján kidolgozott – fordított – kimotripszin-tripszin, kimotripszin-elasztáz, valamint a tripszin-elasztáz átalakítási kísérletek azonban nem hoztak eredményt, a mutánsok aktivitása és specifitás-változása messze elmaradt a várakozásoktól. Az eddigi eredmények alapján úgy tűnik, hogy az egyes szubsztrát specifitási profilok kialakításához a kötőhelyet és környékét alkotó kiterjedt szerkezeti elemek specifikus kombinációja szükséges.

A kimotripszin-tripszin irányú átalakítási kísérletek kimotripszin-B enzimből indultak ki, mely a 226-os helyen alanint tartalmaz. Tripszinben azonban itt egy konzervált Gly

található, a Gly226Ala mutáció pedig 3-4 nagyságrenddel csökkenti a tripszin aktivitását, mert befolyásolja a szubsztrát oldallánc kötődését és leárnyékolja a zseb alján található negatív töltést. Ezek alapján felmerült, hogy a 226-os pozíció befolyással lehet a kimotripszin-tripszin átalakítás eredményére is, ezért megvizsgáltuk az A226G mutáció hatását az egyszerű és az összetett modell alapján készült kimotripszin-B mutánsokban.

Az A226G mutációt két kimotripszin-B mutánsba vittük be, egyikük az egyszerű modell alapján készült alacsony aktivitású, kimotripszinszerű specificitású S189D mutáns, a másik a komplex modellnek megfelelően számos aminosav cserét tartalmazó S1 mutáns, alacsony aktivitással és tripszinszerű specificitási profillal. Az S1 mutáns esetében a Lys és Phe szubsztrátokon végzett fluorimetriás kinetikai mérések azt mutatták, hogy az A226G csere hatására a mutáns kimotripszin-szerű aktivitása nőtt, tripszinszerű aktivitása pedig nem változott. Az S189D mutáns esetében azonban az A226G csere két nagyságrenddel megnövelte a mutáns tripszinszerű aktivitását. Az S189D+A226G mutáns tripszinszerű specificitási profilját egy kompetitív oligopeptid szubsztrát könyvtáron vizsgáltuk tovább, és a mutáns profiljában megjelent a tripszinre jellemző Arg preferencia is. Az S189D+A226G mutáns a tripszinhez hasonlóan autoaktiválódott és képes volt a kimotripszinogén aktiválására.

A tripszinszerű tulajdonságok szerkezeti hátterének vizsgálatához meghatároztuk az S189D+A226G mutáns kristályszerkezetét, 2,2Å felbontással. A kristályosítást benzamidin inhibitor hozzáadása mellett végeztük, a kristályosuló forma azonban a benzamidint nem tartalmazta. A kimotripszinszerű S189D mutáns szerkezetével összehasonlítva jól látszik, hogy az A226G csere jelentős változásokat okoz az S1 szubsztrátkötő hely szerkezetében: a 226-os alanin és a 191-220 diszulfid közötti kölcsönhatás megszűnt, így a diszulfid a kettős mutánsban a zseb belsejéből kifordulva a vad típushoz hasonló pozícióba került. Egy másik fontos különbség a tripszinszerű specificitásért felelős Asp189 pozíciója: a kettős mutánsban az oldallánc a tripszinhez hasonlóan a szubsztrátkötő zseb belseje felé néz. Az Asp189 új konformációját stabilizáló kölcsönhatások azonban jelentősen eltorzítják a zsebet alkotó két hurok szerkezetét. Feltehetően ez okozza, hogy – az S189D mutánsához hasonlóan – az Asp194 oldallánc zimogénszerű konformációban van és az oxianion üreg sem alakul ki. Az autolízis-hurok konformációja is hasonló: a szubsztrátkötőhely jelentős részét elfoglalva H-kötést alakít ki a 195-ös aktív szerinnel.

A kristályban megfigyelt zimogénszerű szerkezetből az enzimaktivitás teljes hiánya következik, ezért feltételezzük, hogy oldatban más, aktív szerkezet(ek) is kialakulhat(nak). Ezt a mutáns triptofán fluoreszcencia változásának pH-függés vizsgálata is megerősíti. Az

eredmények azt mutatják, hogy a mutáció jelentősen eltolta a vad típusra jellemző, inaktív (zimogén-szerű) és aktív formák között fennálló egyensúlyt, és a mutáns optimális pH-n is nagyjából inaktív formában van jelen. Benzamidin inhibitor jelenlétében viszont kialakul az aktív forma, és az aktivitás pH-függés vizsgálatai is azt mutatják, hogy a szubsztrát P1 oldallánca befolyással lehet a régió stabilitására. Az A226G mutáció hatására megnövekedett tripszinszerű aktivitás magyarázata tehát egyrészt az, hogy kialakulhat egy a vad típusú tripszinhez közelebb álló szerkezetet, másrészt lehetővé válnak olyan szubsztrát kiváltotta szerkezetváltozások is, melyek az egyszeres S189D mutánsban korlátozottak.

## 8. Acknowledgements

First, I would like to thank to my Supervisor, Prof. László Gráf who, as Head of Department, made me possible to prepare my thesis at the Department of Biochemistry, conducted the research with thought-provoking consultations and provided constant support to my work.

I thank Dr. István Venekei for teaching me the art of biochemistry and meanwhile a lot more about science and life, and also for the possibility to work in his laboratory.

I would like to acknowledge Dr. Gergely Katona, Dr. Krisztián Fodor and Dr. József Antal, my further co-workers in this project for the fruitful cooperation and the discussions of which I learned a lot.

I also acknowledge Dr. András Málnási-Csizmadia and Dr. Mihály Kovács, my present directorate, for their support in preparing my thesis.

Finally, I would like to say a big “thank you” to all my former and present colleagues at the Department of Biochemistry, for all their technical and personal help during these beautiful years of science.

Balázs

24<sup>th</sup> July 2007.

## References

- Antal, J., Pál, G., Asbóth, B., Buzás, Z., Patthy, A. and Gráf, L. (2001) Specificity assay of serine proteinases by reverse-phase high-performance liquid chromatography analysis of competing oligopeptide substrate library. *Anal. Biochem.*, **288**, 156-167.
- Barrett, A.J. and Rawlings, N.D. (1995) Families and clans of serine peptidases. *Arch Biochem Biophys* **318**, 247-250.
- Bode, W., Schwager, P. and Huber, R. (1978) The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding. The refined crystal structures of the bovine trypsinogen-pancreatic trypsin inhibitor complex and of its ternary complex with Ile-Val at 1.9 Å resolution. *J. Mol. Biol.*, **118**, 99-112.
- Bódi, Á., Kaslik, G., Venekei, I. and Gráf, L. (2001) Structural determinants of the half-life and cleavage site preference in the autolytic inactivation of chymotrypsin. *Eur. J. Biochem.*, **268**, 6238-6246.
- Brady, K., Wei, A.Z., Ringe, D. and Abeles, R.H. (1990) Structure of chymotrypsin-trifluoromethyl ketone inhibitor complexes: comparison of slowly and rapidly equilibrating inhibitors. *Biochemistry* **29** 7600-7607
- Brünger, A.T., Huber, R. and Karplus M. (1987) Trypsinogen-trypsin transition: a molecular dynamics study of induced conformational change in the activation domain. *Biochemistry*. **26**(16):5153-62.
- CCP 4 (1994) *Acta Cryst. D* **50**, 760-763
- Cornish-Bowden, A. (1995) *Fundamentals of Enzyme Kinetics*, Portland Press, London
- Craik, C.S., Largman, C., Fletcher, T., Rocznik, S., Barr, P.J., Fletterick, R.J. and Rutter, W.J. (1985) Redesigning trypsin: alteration of substrate specificity. *Science*, **228**, 291-297.
- Cruikshank, D. W. J. (1999) Remarks about protein structure precision. erratum *Acta Cryst. D* **55**, 583-601
- DeLano W. L., (2002) The PyMOL Molecular Graphics System. San Carlos, CA, USA, DeLano Scientific. <http://www.pymol.org> Ref Type: Computer Program
- Del Mar, E.G., Largman, C., Brodrick, J.W., Fassett, M. and Geokas, M.C (1980) Substrate specificity of human pancreatic elastase 2. *Biochemistry* **19** 468-472
- Dementiev, A., Dobó, J. and Gettins, P.G.W. (2006) Active site distortion is sufficient for proteinase inhibition by serpins: structure of the covalent complex of alpha1-proteinase inhibitor with porcine pancreatic elastase. *J. Biol. Chem.*, **281**(6), 3452-3457
- Dixon, M.M., Brennan, R.G. and Matthews, B.W. (1991) Structure of gamma-chymotrypsin in the range pH 2.0 to pH 10.5 suggests that gamma-chymotrypsin is a covalent acyl-enzyme adduct at low pH. *Int. J. Biol. Macromol.*, **13**(2), 89-96.

Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Cryst.* D60, 2126-32.

Evnin, L.B., Vásquez, J.R. and Craik, C.S. (1990) Substrate specificity of trypsin investigated by using a genetic selection. : *Proc Natl Acad Sci U S A.* 87(17):6659-63.

Fersht, A.R. and Requena, Y. (1971) Mechanism of the  $\alpha$ -chymotrypsin-catalyzed hydrolysis of amides. pH dependence of  $k_{cat}$  and  $K_m$ . Kinetic detection of a tetrahedral intermediate. *J. Am. Chem. Soc.*, 93, 7079-7087.

Fersht, A.R. (1972) Conformational equilibria in  $\alpha$ - and  $\beta$ -chymotrypsin. The energetics and importance of the salt bridge. *J. Mol. Biol.*, 64(2), 497-509.

Fersht, A.R. and Renard, M. (1974) pH dependence of chymotrypsin catalysis. Appendix: substrate binding to dimeric  $\alpha$ -chymotrypsin studied by x-ray diffraction and the equilibrium method; *Biochemistry* **13** 1416-1426

Fodor, K., Harmat, V., Neutze, R., Szilágyi, L., Gráf, L. and Katona, G. (2006) Enzyme:substrate hydrogen bond shortening during the acylation phase of serine protease catalysis. *Biochemistry* 21;45(7):2114-2121.

Fuhrmann, C.N., Daugherty, M.D. and Agard, D.A. (2006) Subangstrom crystallography reveals that short ionic hydrogen bonds, and not a His-Asp low-barrier hydrogen bond, stabilize the transition state in serine protease catalysis. *J Am Chem Soc.* 128(28):9086-9102.

Hooft, R.W., Vriend, G., Sander, C. and Abola, E.E. (1996) Errors in protein structures. *Nature* **381**, 272.

Gráf, L., Jancsó, A., Szilágyi, L., Hegyi, Gy., Pintér, K., Náray-Szabó, G., Hepp, J., Medzihradský, K. and Rutter, W.J. (1988) Electrostatic complementarity within the substrate-binding pocket of trypsin. *Proc. Natl. Acad. Sci. USA*, **85**, 4961-4965.

Gráf, L., Boldogh, I., Szilágyi, L. and Rutter, W.J. (1990) Redesigning the substrate specificity of trypsin: can trypsin be converted to a chymotrypsin-like protease? In: Protein Structure-Function (Zaidi, Z.H., Abbasi, A. and Smith, D.L., eds) TWEL Publishers, Karachi, New York and London, pp. 49-55.

Gráf, L. (1995) In Zwilling, R. (ed), *Natural Sciences and Human Thought*. Springer-Verlag, Berlin, Heidelberg, pp. 139-148.

Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling; *Electrophoresis* 18 2714-2723  
<http://www.expasy.org/spdvb>

Hedstrom, L., Szilágyi, L. and Rutter, W.J. (1992) Converting trypsin to chymotrypsin: the role of surface loops. *Science*, **255**, 1249-1253.

Hedstrom, L., Perona, J.J. and Rutter, W.J. (1994) Converting trypsin to chymotrypsin: residue 172 is a substrate specificity determinant. *Biochemistry*, **33**, 8757-8763.

- Hedstrom,L., Farr Jones,S., Kettner,C.A. and Rutter,W.J. (1994) Converting trypsin to chymotrypsin: ground-state binding does not determine substrate specificity. *Biochemistry*, **33**, 8764-8769.
- Heremans L and Heremans K 1989 Raman spectroscopic study of the changes in secondary structure of chymotrypsin: effect of pH and pressure on the salt bridge; *Biochim. Biophys. Acta* **999** 192-197
- Huber, R. and Bode, W. (1978) Structural basis of the activation and action of trypsin; *Acc. Chem. Res.* **11** 114-122.
- Hudáky,P., Kaslik,Gy., Venekei,I. and Gráf,L. (1999) The differential specificity of chymotrypsin A and B is determined by amino acid 226. *Eur. J. Biochem.*, **259**, 528-533.
- Hung,S. and Hedstrom,L. (1998) Converting trypsin to elastase: substitution of the S1 site and adjacent loops reconstitutes esterase specificity but not amidase activity. *Protein Eng.*, **11**, 669-673.
- Jameson,G.W., Adams, D.V., Kyle, W.S. and Elmore, D.T. (1973) Determination of the operational molarity of solutions of bovine alpha-chymotrypsin, trypsin, thrombin and factor Xa by spectrofluorimetric titration. *Biochem. J.*, **131**, 107-117.
- Jelinek,B., Antal,J.,Venekei,I.,Gráf,L. (2004) Ala226 to Gly and Ser189 to Asp mutations convert rat chymotrypsin B to a trypsin-like protease. *Protein Eng. Des. Sel.*, **17**, 127-131.
- Jelinek, B., Katona, G., Fodor, K., Venekei, I. and Gráf, L. (2007) The crystal structure of a trypsin-like mutant chymotrypsin: The role of position 226 in the activity and specificity of S189D chymotrypsin. *The Protein Journal*, online first status, DOI: 10.1007/s10930-007-9110-3; [http://www.springerlink.com/content/111872/?k=au%3a\(jelinek\)](http://www.springerlink.com/content/111872/?k=au%3a(jelinek))
- Jencks, W.P. (1969) *Catalysis in Chemistry and Enzymology*,Dover, New York
- Kahne, D. and Still, W.C. (1988) Hydrolysis of a peptide bond in neutral water. *J. Amer. Chem. Soc.* **110**. 7529-7534.
- Kardos,J., Bódi,A., Závodszy,P., Venekei,I. and Gráf,L. (1999) Disulfide-linked propeptides stabilize the structure of zymogen and mature pancreatic serine proteases. *Biochemistry*, **38**, 12248-12257.
- Katz, B. and Kossiakoff, A. (1986) The crystallographically determined structures of atypical strained disulfides engineered into subtilisin. *J Biol Chem.* **261**(33):15480-15485.
- Kuhne, W. (1877) *Verh. Naturhist.-Med. Ver. Heidelberg*, **1**, 194-198.
- Kunkel,T.A. (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA*, **82**, 488-492.
- Kuzmic, P. (1996) Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.*, **237**, 260-273.

Laskowski,R.A., MAcArtur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures *J. Appl. Crystallog.* **26**, 283–291.

Lesk, A. M. (1991) Protein Architecture: A Practical Guide, IRL Press, Oxford

Liu,B., Schofield,C.J. and Wilmouth, RC. (2006) Structural analyses on intermediates in serine protease catalysis. *J Biol Chem.* **281(33)**:24024-24035.

Luzzati, P.V., (1952) Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Cryst.*, **5**, 802-810.

Ma, W., Tang, C. and Lai, L. (2005) Specificity of trypsin and chymotrypsin: loop-motion-controlled dynamic correlation as a determinant. *Biophys J.* **89(2)**:1183-93. Epub 2005 May 27.

Madsen,D. and Kleywegt,G.J. (2002) *J. Appl. Crystallog.* **35**, 137–139.

Marquart,M., Walter,J., Deisenhofer,J., Bode,W. and Huber,R. (1983) The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr. Sect. B.* **39**, 480-490.

Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst.* **D53**, 240-255.

Navaza,J (2001) Implementation of molecular replacement in AMoRe. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 1367-72

Northrop, J.H., Kunitz, M. and Herriott, R. (1947) Crystalline Enzymes, Columbia University Press, New York.

Novagen pET System Manual TB055, 7<sup>th</sup> Ed. 4/97. Novagen, Madison, WI.

Perona, J.J., Hedstrom, L., Wagner, R.L., Rutter,, W.J., Craik,, C.S. and Fletterick, R..J. (1994) Exogenous acetate reconstitutes the enzymatic activity of trypsin Asp189Ser. *Biochemistry.* **33(11)**:3252-9.

Perona,J.J., Hedstrom,L., Rutter,W.J. and Fletterick,R.J. (1995) Structural origins of substrate discrimination in trypsin and chymotrypsin. *Biochemistry*, **34**, 1489-1499.

Perona, J.J. and Craik, C.S. (1997) Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. *J Biol Chem.* **272(48)**:29987-90.

Radisky, E.S., Lee, J.M., Lu, C.J. and Koshland, D.E. Jr. (2006) Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. *Proc Natl Acad Sci U S A.* **103(18)**:6835-6840.

Schechter,I. and Berger,A. (1967) On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.*, **27**, 157-162.

Segel, I. H. (1993) Enzyme Kinetics: Behavior and analysis of rapid equilibrium and steady-state enzyme systems, Wiley Classics Library Edition

Shia, S., Stamos, J., Kirchhofer, D., Fan, B., Wu, J., Corpuz, R.T., Santell, L., Lazarus, R.A. and Eigenbrot, C. (2005) Conformational lability in serine protease active sites: structures of hepatocyte growth factor activator (HGFA) alone and with the inhibitory domain from HGFA inhibitor-1B. *J Mol Biol.* **346(5)**:1335-49.

Shotton, D.M. and Watson, H.C. (1970) Three-dimensional structure of tosyl-elastase *Nature* **225**, 811-816

Steitz, T.A., Henderson, R. and Blow, D.M. (1969) Structure of crystalline alpha-chymotrypsin. 3. Crystallographic studies of substrates and inhibitors bound to the active site of alpha-chymotrypsin. *J. Mol. Biol.*, **46**, 337-348.

Stoesz, J. D. and Lumry, R. W. (1978) Refolding transition of alpha-chymotrypsin: pH and salt dependence; *Biochemistry* **17** 3693-3699

Szabó, E., Böcskei, Zs., Náray-Szabó, G. and Gráf, L. (1999) The three-dimensional structure of Asp189Ser trypsin provides evidence for an inherent structural plasticity of the protease. *Eur.J.Biochem.* **263(1)**, 20-26.

Szabó, E., Venekei, I., Böcskei, Zs., Náray-Szabó, G. and Gráf, L. (2003) Three dimensional structures of S189D chymotrypsin and D189S trypsin mutants: the effect of polarity at site 189 on a protease-specific stabilization of the substrate-binding site. *J.Mol.Biol.* **331(5)**, 1121-1130.

Terwilliger, T.C. (2000) Maximum likelihood density modification *Acta Cryst.* **D56**, 965-972.

Thompson, R.C. and Blout, E.R. (1973) Dependence of the kinetic parameters for elastase-catalyzed amid hydrolysis on the length of peptide substrates *Biochemistry* **12/1**, 57-65

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22(22)**:4673-80.

Tóth, J., Simon, Z., Medveczky, P., Gombos, L., Jelinek, B., Szilágyi, L., Gráf, L. and Málnási-Csizmadia, A. (2006) Site directed mutagenesis at position 193 of human trypsin 4 alters the rate of conformational change during activation: role of local internal viscosity in protein dynamics. *Proteins.* **67(4)**:1119-27.

Venekei, I., Gráf, L. and Rutter, W.J. (1996) Expression of rat chymotrypsinogen in yeast: a study on the structural and functional significance of the chymotrypsinogen propeptide. *FEBS Lett.*, **379**, 139-142.

Venekei, I., Szilágyi, L., Gráf, L. and Rutter, W.J. (1996) Attempts to convert chymotrypsin to trypsin. *FEBS Lett.*, **379**, 143-147.

Verheyden,G., Matrai,J., Volckaert,G., Engelborghs,Y. (2004) A fluorescence stopped-flow kinetic study of the conformational activation of alpha-chymotrypsin and several mutants. *Prot. Sci.*, 13, 2533-2540.

Wilke,M.E., Higaki,J.N., Craik,C.S. and Fletterick,R.J. (1991) Crystallographic analysis of trypsin-G226A. A specificity pocket mutant of rat trypsin with altered binding and catalysis. *J. Mol. Biol.*, **219**, 525-532.

Zerner,B. & Bender,M.L. (1964) The kinetic consequences of the acyl-enzyme mechanism for the reactions of specific substrates with chymotrypsin. *J Am Chem Soc* **86**, 3669-3674.