Max-Planck-Institut für Biochemie
Abteilung Strukturforschung

Structural and Biochemical Studies on Tricorn Protease: A Molecular Cage for Protein Degradation

Jeong-Sun Kim

Table of Contents

5 STRUCTURE DESCRIPTION

6 MUTAGENESES AND INHIBITOR SYNTHESES

7 RESULTS

1 Acknowledgements

The work for the thesis was carried out from July of 1999 to September of 2002 in the Abteilung für Strukturforschung at the Max-Planck-Institut für Biochemie, Martinsried, Germany under the supervision of Prof. Dr. Robert Huber, FRS.

During my stay in Martinsried, a number of people helped me accomplishing the final goal that this thesis presents. Work companions and friends alike, people that taught me and helped me, and must be here properly acknowledged. They will be so in a more or less random order.

I would like to thank greatly Prof. Dr. Robert Huber for giving me the chance to work here, his full financial support for my small family here during the thesis work, and his priceless advices when I met unexpected problems.

Specials thanks go for a secretary, Renate Rüller for taking care of all affairs of our family during my whole thesis work and retired Gina Beckman, too.

I acknowledge the kindness and thoughtfulness of Prof. Franz P. Schmidtchen and Prof. Wolfgang Hiller at the Technical University of München.

I would like to thank Dr. Hans Brandstetter and Dr. Michael Groll for sharing the tricorn protease project.

A special acknowledgement is due to Prof. Dr. Luis Moroder, Dr. Raymond Behrendt, Hans-Jürgen Musiol and Markus Kaiser for the valuable tricorn inhibitor syntheses and for helping the analyses of degradation products.

I appreciate Dr. Hyun Kyu Song's family, Dr. Hyun-Ju Cha's family, David Nixon & Kiley Lee, and Taewon's family for helping our family to settle down and taking consistent cares here.

I acknowledge Dr. Tim Clausen for teaching the Program O, Dr. Stefan Steinbacher for teaching the cyclic averaging program, Dr. Peter Sonderman for the invaluable frequent discussions, Dr. Ravishankar Ramachandran for helping activity assays of

2 Summary

Protein degradation by proteases, the last step in the life of a protein, is essential for the maintenance of required intracellular components. The initial terminators for this special event are ATP-dependent proteases. Among them, proteasomes are responsible for most of cytosolic protein recycles by taking part in the removal of incorrectly folded or damaged proteins (Etlinger and Goldberg, 1977), the activation of transcription factors (Palombella *et al.*, 1994), the degradation of proteins involved in cell cycle control (Glotzer *et al.*, 1991), and the generation of peptides destined for presentation by class I molecules of the major histocompatibility complex (Rock *et al.*, 1994). Predominant proteasomal products, peptides of 7 to 9 amino acid long, need further processing for the recovery of free amino acids pool, since they are, if any, of no use. Then, free amino acids can be used for the protein synthesis and the energy production. Recently, a new proteolytic system was identified in the model organism *Thermoplasma acidophilum*, which performs this processing. The hexameric core protein of this modular system is named as the tricorn protease (Tamura *et al.*, 1996a).

Tricorn is a hexameric protease of 720 kDa and can assemble into a giant icosahedral capsid, which might serve as the organizing center of a multiproteolytic complex, as determined by electron microscopy (Tamura *et al.*, 1996a; Walz *et al.*, 1997). 2.0 Å resolution crystal structure reveals that the 120 kDa subunit has a mosaic structure, two open Velcro β-propeller structures of 6 and 7 blades (Met39-Asp310 and Ala326-Lys675), a helical bundle (Val679-Ser745), a PDZ domain (Arg761-Asp855) and an αβ sandwich structure (Arg856-Asn1061). The five domains combine to form one of 6 sub-units, which further assemble to form 3-2 symmetric core protein. The active site residues, Ser965 and His746 which are located in the C-terminal half, were visualized by crystallographic structural studies of ligand complexes, TLCK and TPCK, and confirmed by constructing the single site mutants, S965A and H746N, both of which are amidolytically inactive. The hexameric ring is introduced by a number of hydrogen bonds between the two 6-bladed β-propeller domains, the two 7-bladed β-propeller domains, and the two C-terminal domains on the dimeric

interfaces. On the other hands, dimerizations are derived from the tight interactions between the 7-bladed β-propeller structure and the PDZ domain, together with two arms (Leu520-Val535 and Asn930-Asn949).

Obstruction of the pore within the 7-bladed β-propeller domain by alkylation or oxidation of an engineered double cysteine mutant (R414C and A643C) exhibited strongly decreased enzymatic activities. In line herewith, the crystal structure of the tricorn protease in complex with a tridecapeptide inhibitor modifying the catalytic Ser965 revealed part of the peptide trapped inside the channel of the 7-bladed β-propeller domain. A cysteine mutation in the lumen of the 6-bladed β-propeller domain (L184C) enhanced catalytic activity, which was restored to normal values after its alkylation. A charge reversal mutant at the putative anchor site of substrate C-terminus (R131E-R132E) drastically reduced the proteolytic activity. The complex crystal structure of an inhibitor with a diketo group at the cleavage site exhibited that the carbonyl oxygen of peptide bond between the P2' and the substrate C-terminal is anchored to these two arginines (Arg131-Arg132). Biochemical and structural results suggest the 7-bladed β-propeller domain to serve as a substrate selective filter and guide to the sequestered active site, while the 6-bladed β-propeller domain is the product channel to the protease surface from the active site. Several inhibitor complex structures explain the unique substrate anchoring to Arg131-Arg132 as well as the substrate binding to the active site by forming a local anti-parallel β-sheet between the inserted substrate (P2-P2') and the two short strands of the protease matrix (Gly993-Thr995 and Gly916-Gly918 at the unprimed and the primed site, respectively).

3 Introduction

3.1 Proteases

Proteolytic enzymes are found in all cells and tissues, where they degrade other proteins by splitting the peptide bond between two amino residues or aid in the digestion of food. The recommended designation for this class of enzymes is a peptidase, a synonym of the more widely used term protease. They are thus peptide bond hydrolases, which are classified according to their catalytic types and evolutionary relationships (Barret & Rawlings, 1991; Rawlings & Barret 1993). Five main types or classes of proteases are presently known, differing in their catalytic mechanisms. These classes are termed according to the most prominent functional group in their active sites.

3.1.1 Serine proteases

The serine protease class comprises the mammalian enzymes of the chymotrypsin family (including trypsin, elastase and the coagulation cascade protease, as well as chymase and tryptase), and the bacterial proteases belonging to the subtilisin family. Even though two enzyme families share a common catalytic triad and a common catalytic mechanism, they show quite different three-dimensional structures. The different substrate specificities are determined by the shape of the various enzyme subsites and/or the amino acid residues delineating them. The catalytic mechanism can be divided into two main steps. The first step involves the formation of a covalent bond between the carbonyl carbon of the P1 residue and the hydroxyl group of the catalytic serine residue (acylation). Production of acyl-enzyme intermediate proceeds through a negatively charged transition intermediate, where carbonyl carbon of P1 residue assumes a tetrahedral geometry, in contrast to the planar triangular geometry in the peptide group. During this step the peptide bond is cleaved, one smaller peptide product remains attached to the enzyme, and the other peptide product diffuses away. In the second catalytic step, a water molecule, which is activated by the imidazole ring of catalytic histidine, hydrolyzes the acyl-enzyme intermediate, releases the substrate with a free C-terminus, and restores the hydroxyl group of the catalytic

serine (deacylation). This step also proceeds through the formation of a negatively charged tetrahedral transition state intermediate. The histidine and the aspartate residues of catalytic triad form the charge relay system, providing a binding site for the proton in the transition state, which is being shuttled between the reactive serine and the leaving or entering group.

3.1.2 Threonine Proteases

Threonine proteases are the newest class of endopeptidases, which only one family has been identified so far, the multicatalytic proteolytic complex (proteasome). In the yeast 20S proteasome, a water molecule acts as a general base by abstracting a proton from the hydroxyl group of the active Thr1, which drives the nucleophilic attack of Thr Oγ to the carbonyl carbon of P1 residue (acylation). When a proton is transferred to the amino group of Thr1 from the water molecule, the water molecule hydrolyzes the acyl-enzyme intermediate, releases the N-terminal peptide product, and restores the hydroxyl group of the Thr1 (deacylation). Asp17 and Lys33 compete the catalytic triad of the proteasome, forming a hydrogen bond network that enhances the nucleophilic character of Thr1 N and Oγ as well as of the catalytic water molecule (Groll *et al.*, 1997).

3.1.3 Aspartic Acid Proteases

This group of proteases contains two aspartates as catalytic residues, which are originated from two domains in monomeric enzymes (pepsin, chymosin and *etc.*) or from two subunits in the oligomer-forming enzymes (human immunodeficiency virus type 1(HIV-1) protease). The nucleophilic attack is performed by a water molecule, which is activated by one carboxylate group of those two aspartates. A second simultaneous proton transfer is taken place from the other carboxylate group to the carbonyl oxygen of P1 residue. At the acidic optimal pH of these enzymes, only one of the aspartate carboxylates is ionized. Two further proton transfers to the tetrahedral substrate intermediates lead to the cleavage of the peptide bond, in a general acid-base catalytic mechanism.

3.1.4 Cysteine Proteases

Two major types of proteases belong to this class of proteases, papain and caspases (cysteine dependent aspartyl proteases). They have a common characteristic catalytic mechanism, even though they have different foldings. Like serine proteases, the catalytic mechanism of cysteine endopeptidases involves the formation of a covalent tetrahedral intermediate. The Sγ of a cysteine residue, which is activated by a histidine, is the attacking nucleophile, thus playing the same role as the serine Oγ in serine proteases. The imidazole group of a histidine residue stabilizes the attacking thiolate ion, which adds to the P1 carbonyl carbon resulting in a tetrahedral intermediate covalently bound to the enzyme. The rupture of peptide bond concomitantly releases the C-terminal peptide product, resulting in an acyl-enzyme complex. Hydrolysis of the S-C bond by a water molecule, which is activated by the imidazole ring, releases the N-terminal product and regenerates the thiol group of the enzyme.

3.1.5 Metalloproteases

They possess metal ions (normally a zinc) at the catalytic site, although the coordinating groups might be variable. Like aspartic acid proteases, metalloproteases cleave their substrates without formation of a covalent intermediate. A water molecule, which is bound to the metal ion and hydrogen-bonded to a neighboring glutamic acid, is activated by the latter and attacks the P1 carbonyl carbon. A tetrahedral intermediate forms, coordinated to the zinc via both the attacking water and the carbonyl oxygen of P1. Upon transfer of a proton from the glutamic acid to the leaving nitrogen, this intermediate decomposes, the C-terminal peptide leaves and the N-terminal peptide is replaced by a water molecule. The glutamic acid seems to act as a proton shuttle, in analogy to one of the catalytic groups in the aspartic acid proteases and to the histidine in the serine and cysteine proteases.

3.2 Protein Degradation in the Cytosol

Protein synthesis and degradation in cells are balanced to maintain homeostasis of the required cellular components. Initial intracellular protein degradation in the cytosol is carried out mostly by sequence-unspecific proteases with sieving mechanisms for substrate selection. Commonly, they form molecular cages by oligomerization to exclude tertiary folded substrates from their central cavities harboring the active sites. Proteasomes are representatives of this class of proteases. They form large multi-subunit complexes reaching molecular weights of 2.0 MDa. They take part in the degradation of most cytosolic proteins (Rock *et al.*, 1994) and at least some proteins of the endoplasmic reticulum (Sommer *et al.*, 1993, 1997; Hiller *et al.*, 1996). Their central entry ports are not wide enough to allow folded substrates to reach the active sites (Löwe *et al.*, 1995; Groll *et al.*, 1997; Bochtler *et al.*, 2000; Figure 1). Accessory proteins form the regulatory particle of the proteasome and are attached to the core particle to regulate the substrate access to the active sites in an ATP-dependent way (Groll *et al.*, 2001).
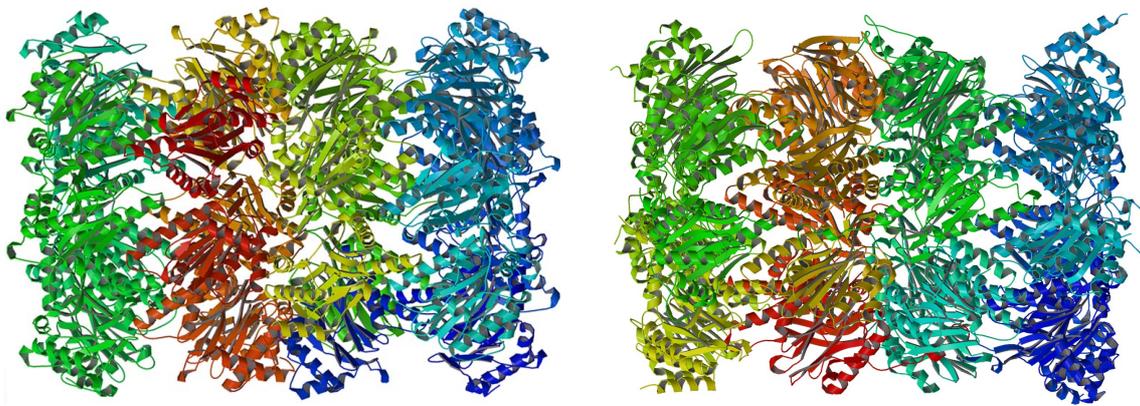
Figure 1. Ribbon representations of proteasomes from *Thermoplasma acidophilum* (left, 1pma) and yeast, *Saccharomyces cerevisiae* (right, 1ryp) (Löwe *et al.*, 1995; Groll *et al.*, 1997). Their central cavities are not wide enough to allow tertiary folded proteins to be located inside, but they take part in various tasks in controlling the cell life.

3.3 Degradations of Proteasomal Products by the Tricorn Protease in *Thermoplasma acidophilum*.

The proteasome removes misfolded or unneeded proteins in eukaryotes and archaea. Peptides, with a preferred length of between 7 to 9 amino acids, are the major products of the proteasome. However, peptides of these lengths are of no use, if any. Therefore, these peptides must be further degraded into free amino acids, which can be used for the protein synthesis and the energy production. In the archaeon *Thermoplasma acidophilum*, the degradation of the proteasomal products is performed by the tricorn protease. The further and final degradations to free amino acids are accomplished by the tricorn protease interacting factors, a proline iminopeptidase termed F1 and two metalloproteases termed F2 and F3 (Tamura *et al.*, 1996a, b, 1998).

The tricorn protease consists of six identical subunits of 120 kDa and its complex is a trimer of dimers forming a toroid traversed by a channel along its threefold axis. The hexameric toroid structure of the tricorn protease is approximately 18 nm in its diameter and 10 nm in
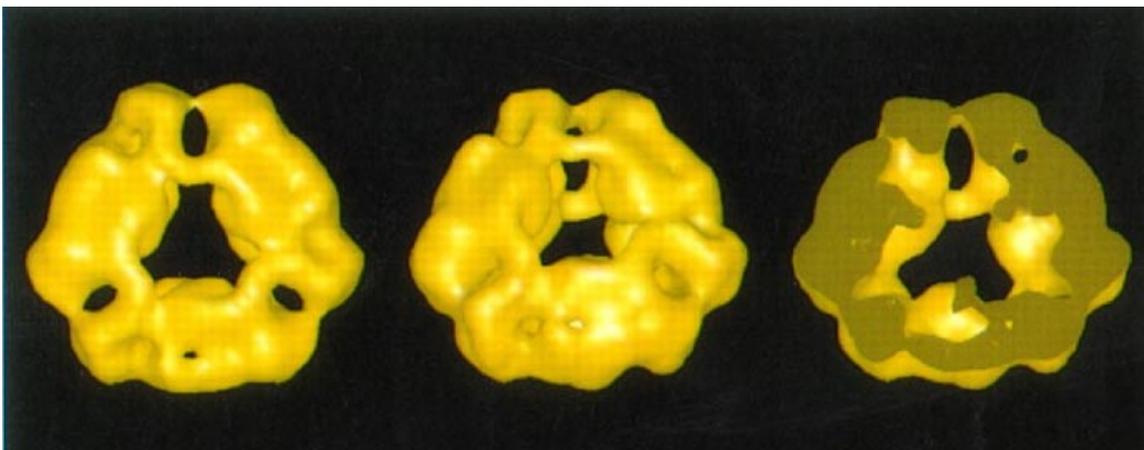
Figure 2. Electron microscopic characterization of the tricorn protease from *Thermoplasma acidophilum* (Tamura *et al.*, 1996a). Three-dimensional reconstruction of a tricorn protease hexamer with 32-point group symmetry imposed. This is a shaded isosurface representation viewed along the threefold axis of the molecule. Three twofold axes are in the plane perpendicular to the threefold axis (left). The view was obtained by tilting the model around an axis that is in the plane

of the twofold axes and perpendicular to one of them (center). The same view but with the upper half of the molecule removed is shown at right.

the height (Tamura *et al.*, 1996a; Figure 2). The quaternary structure of the tricorn protease is quite similar to that of the Gal6 or bleomycin hydrolase from *Saccharomyces cerevisiae* (Zheng *et al.*, 1995; Figure 3), which are also hexamers with a central channel. However, the sequence of the tricorn protease gene is not related to Gal6/bleomycin hydrolase, nor is any other tricorn homologues in the databases. The only detectable homology to other proteins occurs in a stretch of ~160 residues at the C-terminal (Thr878-Ala1036) of the tricorn subunit (Figure 4). This segment is related to several C-terminal processing proteases of bacterials and eukaryotic origin and may form a domain that participates in substrate binding.
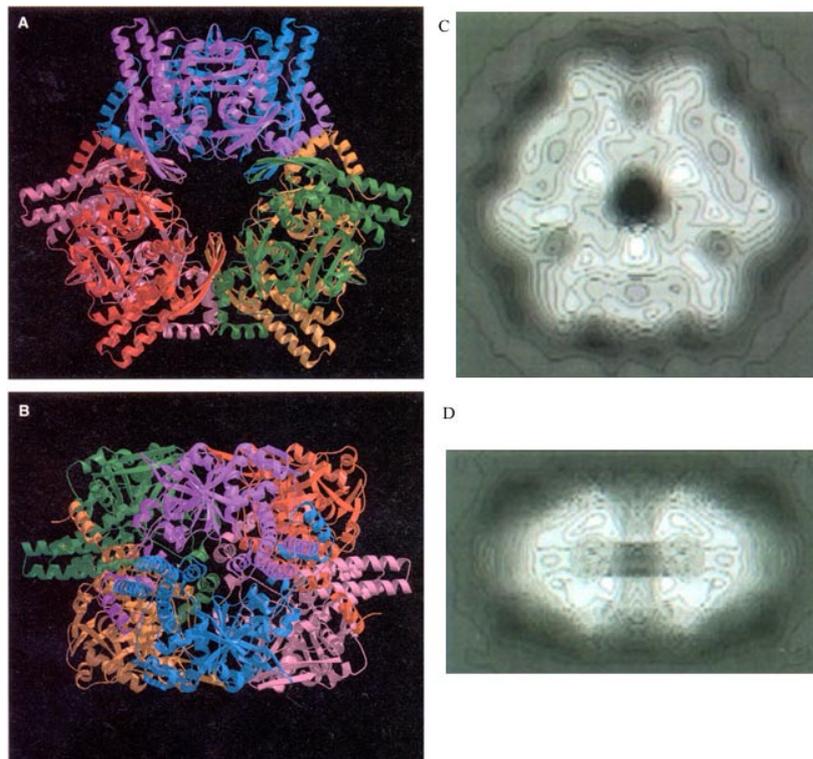
Figure 3. Comparison of a hexameric overall shape of the tricorn protease (Tamura *et al.*, 1996a) with the Gal6/belomycin hydrolase from *Saccharomyces cerevisiae* (Joshua-Tor *et al.*, 1995). There are striking resemblances in overall appearances of the hexameric forms between the Gal6/bleomycin hydrolase and the tricorn protease. (A) A top view of the Gal6 along the threefold axis. (B) A side view of Gal6 perpendicular to the top view. (C) A top view of the tricorn protease from *Thermoplasma*

acidophilum along the three-fold axis. (D) A side view of the tricorn protease perpendicular to the top view.
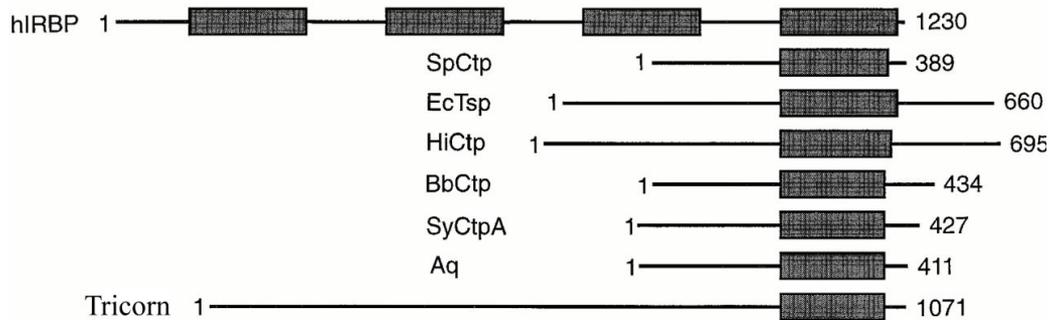
Figure 4. Schematic alignment of the tricorn protease (Tricorn) from *Thermoplasma acidophilum*, human interphotoreceptor retinol-binding protein (hIRBP), and C-terminal specific protease homologs from eubacteria (Ec, *E. coli*; Hi, *Haemophilus influenzae*; Bb, *Bartonella bacilliformis*; Sy, *Synechocystis sp.* PCC 6803; and Aq, *Agmenellum quadruplicatum*) and a eukaryote (Sp, *Spinach*). Shaded boxes indicate the regions of homology with the IRBP domain (Tamura *et al.*, 1996a).

Unlike the 20S proteasome from *Thermoplasma acidophilum*, which cleaves chymotrypsin-like substrates, the tricorn protease preferentially cleaved trypsin-like substrates except one chymotrypsin-like substrate which was not cleaved either, when it carried a negative charge at its N-terminal end (Tamura *et al.*, 1996a; Table 1). The intrinsic proteolytic activities of the tricorn protease were enhanced and new activities were generated, upon interacting with low molecular weight factors (F1, F2, and F3). While the proteasomes require the ATP-dependent protein machines for the substrate selection, there was no evidence for an ATP requirement for the enzymatic activity of the tricorn protease (Tamura *et al.*, 1996a, 1998).

20 copies of hexamer can assemble further to form an icosahedral capsid structure with a molecular mess of 14.6 MDa, which might serve as the organizing center of a multi-proteolytic complex (Walz *et al.*, 1997, 1999; Figure 5). This capsid appears to enclose a cavity approximately 37 nm in diameter, large enough to accommodate a ribosome and was observed only *in vivo*.

Table 1. Hydrolysis of synthetic substrates of the tricorn protease and the proteasome (Tamura *et al.*, 1996a).

	Tricorn protease (nmol hr^{-1} mg^{-1})		Proteasome (nmol hr^{-1} mg^{-1})
	Native	Recombinant	
Trypsin-like activity			
Z-Gly-Gly-Arg-AMC	138	209	0
Boc-Leu-Arg-Arg-AMC	347	389	0
Z-Ala-Arg-Arg-AMC	334	308	0
Bz-Val-Gly-Arg-AMC	997	1257	0
Chymotrypsin-like activity			
Suc-Leu-Tyr-AMC	0	0	0
Suc-Ala-Ala-Phe-AMC	0	0	400
H-Ala-Ala-Phe-AMC	5125	6460	978
Suc-Leu-Leu-Val-Tyr-AMC	0	0	1432
Z-Gly-Gly-Leu-AMC	0	0	1750

Figure 5. Isosurface representation of the capsid portion of the tricorn protease at 13 Å resolution (Walz *et al.*, 1999). Views of the capsid down the fivefold (left), threefold (center), and twofold axes (right). The scale bar corresponds to 50 nm.

3.4 Protein Crystallography

A detailed discussion of protein crystallography is beyond the scope of this work and can be found in relevant textbooks (Blundell & Johnson, 1994; Drenth, 1994; Glusker *et al.*, 1994; McRee, 1993) and in *Meth. Enzymol.* **276**, 1997.

Crystallization trials were carried out using standard materials and solutions, home-made and sparse-matrix screens purchased from Hampton Research (USA).

4 Materials and Methods

4.1 DNA Work

Unless mentioned specifically, standard techniques and materials were used.

4.1.1 Design of Primers

The nucleotide sequences of the primers for amplification were based on the reported tricorn protease encoding gene sequences from the archaeon *Thermoplasma acidophilum* (Tamura *et al.*, 1996a) and contained the *NdeI* and the *XhoI* restriction enzyme sites at each end like followings:

Primer 1: 5'-CCACCACCACCACATATGCCATCCCTTATGTCCTTTGGGTC-3'
Primer 2: 5'-CCACCACCACCACTCGAGTTAAGAGGGTCTCTGAGGCAAC-3'

One more extra stop-codon was added immediately after the ending codon site, even though the gene sequences contain one at the translation end.

4.1.2 Polymerase Chain Reaction

Polymerase Chain Reaction (Innis *et al.*, 1988; Saiki *et al.*, 1988) was carried out with *Pfu* polymerase (Stratagene, Heidelberg, Germany) on a stratagene Robocycler Gradient 96 (Stratagene, Heidelberg, Germany) obeying the following steps:

$10 \times$ Reaction buffer: 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 100 mM Tris·HCl (pH 8.8), 20 mM $MgSO_4$, 1 % (w/v) Triton$^\circledR$ X-100, 1 mg/ml nuclease-free bovine serum albumin (BSA)

- initial denaturation step: 10 min, 95 °C   1cycle;
- elongation step: denaturation 1min, 95 °C

                    annealing 1min, 52 °C

elongation 4.5 min, 68 °C 30 cycles;

• termination step: 10 min, 68 °C   1cycle.

Reaction mixtures for the PCR reaction were prepared like followings:

2 µl Template which corresponds to 20-100 ng genomic DNA

3 µl Primer 1 which corresponds to 30 pM

3 µl Primer 2 which corresponds to 30 pM

2 µl dNTP mix which corresponds to 0.5 mM each nucleotide

5 µl $10 \times pfu$ DNA polymerase buffer

1 µl pfu DNA polymerase that corresponds to 2.5 units

34 µl distilled water to a final reaction volume of 50 µl

and 40 µl of mineral oil was added on the top of the reaction mixtures.

4.1.3 Vector for Cloning

pRSet6c (Schoepfer, 1993) was chosen for the cloning and the protein expression.

4.1.4 Restriction Enzyme Treatment

Amplified tricorn protease encoding DNA and the vector, pRSet6c were double-digested by the *NdeI* and the *XhoI* restriction enzymes on the following condition:

NEB buffer 2: 10 mM Tris·HCl, 10 mM $MgCl_2$, 50 mM NaCl, 1 mM DTT, pH 7.9 at 25 °C

7 µl insert gene or vector

1 µl *NdeI* (20 units, New England Biolab)

1 µl *XhoI* (20 units, New England Biolab)

1 µl $10 \times$ NEB buffer (2, New England Biolab)

10 µl total reaction volume

Sample sets were incubated at 37 ˚C overnight and the double-digested products were purified by running and eluting the 0.8 % agarose gel.

4.1.5 Ligation

The double digested DNA and the vector were mixed with different ratios on the following condition:

T4 DNA ligase buffer: 50 mM Tris·HCl, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 25 ug/ml BSA, pH 7.5

        x μl insert gene

        (8-x) μl vector

        1 μl 10 × T4 DNA ligase buffer (New England Biolab)

        1 μl T4 DNA ligase (20 units/μl, New England Biolab)

        10 μl total reaction volume

Ligation reactions were accomplished by incubating the above mixtures at 4 ˚C, overnight.

4.1.6 Transformation into *E.coli*

Ligase-treated mixtures of the vector and insert gene were then transformed into *E.coli* DH5α cell line by heat shock method, which was prepared by following the below steps.

E.coli DH5α

(ITC Biotechnology GmbH, Heidelberg), (Hanahan, 1983)

Genotype: *deo*R, *end*A1, *gyr*A96, *hsd*R17($r_k^-m_k^+$), *rec*A1, *rel*A1, *sup*E44, *thi*-1, D(*lac*ZYA-*arg*F), f80*lac*ZDM15F1⁻.

Transformation Buffer I & II were prepared and the components are described on the below. *E.coli* cells were harvested at the log phase (optical density (OD) $_{600\ nm}$ = 0.5) and cooled for 30min on ice. They were resuspended into 30 ml of *Transformation*

Buffer I and incubated for 30 min on ice. Then, bacteria was pelleted by centrifugation at 4 °C and resuspended again by 10 ml of *Transformation Buffer II*. Cells were aliquot by 100 μl units after another incubation for 10 min on ice (Hanahan, 1983).

Transformation Buffer I: 50 mM $MnCl_2$

30 mM KAcOH

100 mM RbCl

10 mM $CaCl_2$

15 % (w/v) Glycerol

pH titration with 0.2 M AcOH to 5.8

Transformation Buffer II: 10 mM RbCl

75 mM $CaCl_2$

10 mM MOPS

15 % (w/v) Glycerol

pH titration with NaOH to 6.8

Luria-Bertani (LB) Meduim:

10 g bacto-tryptone

5 g bacto-yeast extract

10 g NaCl

pH titration to 7.0 with NaOH

Adjustment of volume to 1 liter with distilled water.

Sterilization by autoclaving for 20 minutes at 15 lb/sq. at 121 ˚C.

For preparation of LB-agarose plate, 15 grams of bacto-agar from Sigma was melted into 1 l LB medium (1.5 % (w/v)) and autoclaved. Required antibiotics were added.

5 μl of ligation mixture was added into 100 μl competent cells prepared at above steps and incubated for 30 min on ice. Transformation was done by heat shock method for 1 min 45 sec at 42 ˚C and the heat-shocked cells were cooled for 10 min on ice. Cells were incubated for 1 hr at 37 ˚C after adding 900 μl of LB media, spread on the LB-

agarose plate with 0.1 mg/ml ampicillin (final concentration), and incubated overnight at 37 ˚C.

4.1.7 Purification of Plasmid

A single colony was picked, inoculated into the LB medium with 0.1 mg/ml ampicillin (final concentration) and cultured overnight at 37 ˚C. Cell debris were collected by centrifugation and plasmids were extracted by following the instruction of the Qiagen miniprep kit:

Pelleted bacterial cells were resuspended into 250 μl of Buffer P1, lysed with adding 250 μl of Buffer P2 and mixed gently by inverting. Lysed proteins were precipitated by adding of 350 μl of Buffer N3. The precipitation was removed by centrifugation and DNA was absorbed to the QIAprep column by decanting or pipetting and centrifugation for 1 min at 13,000 rpm. Column was washed two times by centrifugation after adding 500 μl of buffer PB and 750 μl of buffer PE, successively. Absorbed DNA was eluted by adding 30 μl of distilled water to the center of QIAprep column and collected by centrifugation for 2 min at 13,000 rpm.

Buffer P1:   50 mM Tris·HCl, pH 8.0

10 mM EDTA

100 μg/ml RNAse A, cooled at 4 °C

Buffer P2:   200 mM NaOH

1 % (w/v) SDS

Buffer N3:   3 M KAcOH, pH 5.5

Buffer PB: 750 mM NaCl

50 mM MOPS, pH 7.0

15 % (v/v) ethanol

0.15 % (v/v) Triton X-100

Buffer PE: 1.25 M NaCl

50 mM MOPS, pH 8.5

15 % (v/v) ethanol

4.1.8 Sequencing the Inserted Gene

Inserted DNA into the expression vector was sequenced on the basis of PCR method (Innis *et al.*, 1988; Saiki *et al.*, 1988). This construct, pRSet6c-tricorn from *Thermoplasma acidophilum* with the *NdeI* and the *XhoI* restriction sites for insertion, was used for all required protein works and DNA works for the mutageneses.

4.2 Protein Biochemistry

Unless mentioned specifically, standard techniques and materials were used. Protein concentration was determined considering the molar extinction coefficients (1.05 in case of the tricorn protease). Protein solutions were concentrated using Amicon Ultrafiltration kit and Centricons (Amicon, USA) and Ultrafree centricon devices (Millipore, USA):

• Size exclusion chromatography was performed on a FPLC system (Pharmacia) with a Superose 6 HR (1.0/30 cm) column (void volume 2.4ml). Column calibration was performed with globular protein size standards (Fluka and BioRad);

• SDS-PAGE (8-12 % (w/v)) was performed according to Laemmli (Laemmli, 1970) on a Biorad system (BioRad). Gels were stained with Coomassie brilliant blue G-250 or silver (Heukeshoven, 1985);

• UV/Visible and fluorescence were performed using Lambda 17 UV/Vis (Perkin Elmer) and an LS 50B Luminescence (Perkin Elmer) spectrometers, respectively.

4.2.1 Expression and Purification of the Tricorn Protease

E.coli BL21(DE3)

(Novagen, Inc., Madison, Wisconsin, USA), (Studier & Moffatt, 1986)

Genotype: F⁻, *omp*T, *hsd*S$_B$(r$_B$⁻m$_B$⁻), *gal*, *dcm*, (DE3).

E.coli B834(DE3)

(Novagen, Inc., Madison, Wisconsin, USA), (Studier & Moffatt, 1986)

Genotype: F⁻, *omp*T, *hsd*S$_B$(r$_B$⁻m$_B$⁻), *gal*, *dcm*, *met* (DE3)

The wild-type tricorn construct (pRSet6c-Tricorn with the *Nde*I and the *Xho*I restriction enzyme sites for insertion) was transformed into the B834(DE3) or BL21(DE3) *E.coli* strain and protein expression was induced with 1 mM isopropyl β-D-thiogalactoside (IPTG, final concentration) at a cell density OD$_{600nm}$ 0.8 at 30 °C. Induced cell dump was harvested after 5 hours by pelleting at 6,000 rpm for 30 min at 4 °C. Harvested cells were resuspended into buffer A (20 mM Tris·HCl, pH 7.5 and 2 mM β-mercaptoethanol), then frozen overnight at –20 °C. Frozen cells were thawed in warm water for 30 min, and disrupted by sonification. Cell debris was removed by centrifugation at 25,000 rpm for 30
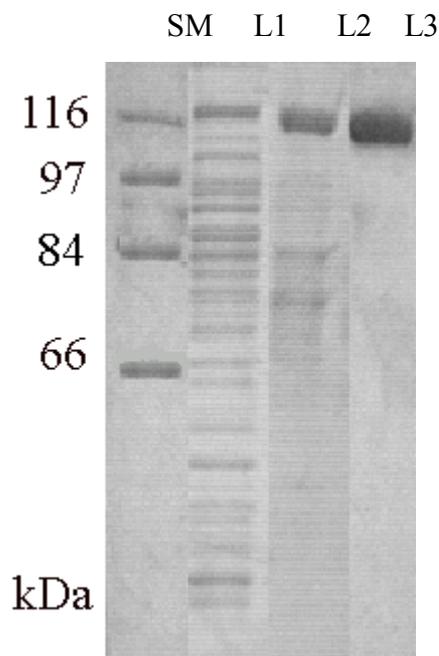

Figure 6. 10 % (w/v) SDS-PAGE gel, stained with Coomassie brilliant Blue, after individual purification steps. SM stands for the size marker from Sigma. L1 is the pool after the heat treatment at 65 °C for 30 min and successive Q-Sepharose anion exchange chromatography, L2 after the Hydroxyapatite column and L3 after Superose 6 gel filtration column.

min at 4 °C. The resulting supernatant was subjected to incubation for 30 min at 65 °C against the supernatant obtained from the above centrifugation. Denatured contaminating proteins were removed by additional centrifugation step at 25,000 rpm for 30 min at 4 °C. The cleared part was applied onto the Q-Sepharose column (1.8 × 20 cm, Pharmacia). Recombinant tricorn protease was eluted by a linear NaCl gradient from 0 to 500 mM in buffer A. Fractions, which have a molecular weight of 120 kDa on the SDS gel (Figure 6 L1), were pooled against buffer B (20 mM Potassium Phosphate pH 7.4 and 2 mM β-mercaptoethanol), and loaded onto the Hydroxyapatite column (1.8 × 20 cm, Pharmacia). Protein was eluted by potassium phosphate gradient from 20 to 400 mM at same pH (Figure 6 L2). Protease was concentrated with an Amicon pressurized device and ultra filtration, and filtered by centrifugation while the buffer was changed to buffer A including 100 mM NaCl. Further purification was done by applying the Superose 6 gel filtration chromatography (1.0 × 30 cm, Pharmacia; Figure 6 L3). Less than 2 mg of the tricorn protease was purified from 6 l culture. Purified protein was stored in buffer A.

4.2.2 N-terminal Sequencing

To obtain the peptide sequence of the purified recombinant protein, N-terminal sequences of first eight amino acids were determined by the Edmann degradation (*Dr. Karlheinz Mann)* on an Applied Biosystems sequencer model 492. About 24 amino acids of N terminus (sequence in the parenthesis in below results) were cleaved by contaminated proteases during purification, since catalytically inactive point mutants (S965A and H746N) have the same processing patterns.

N-terminal sequence results of the purified recombinant protein:

      (Met-Pro-Ser-Leu-Met-Ser-Phe-Gly-Ser-Cys-Gln-Trp-Ile-Asp-Gln-Arg-Phe-

      Ser-Arg-Ser-Leu-Tyr-Arg-Asn-)

      major 1: Phe-Lys-Thr-Phe-Lys-Leu-His-Glu

      major 2: Phe-Lys-Leu-His-Glu-Met-His-Gly

4.2.3 Peptidolytic Assays

Peptidolytic activities of the tricorn protease were checked by mixing enzyme and fluorogenic benzoyl (Bz)-Val-Gly-Arg-7-amino-4-methylcoumarin (AMC) substrates in a 100 µl reaction volume. Reaction mix was incubated for 30 or 60 min at 65 °C. After adding of 900 µl reaction buffer (20 mM Tris·HCl pH 7.5, 100 mM NaCl), fluorescences were measured at the wavelength of 460 nm by exciting UV at 360 nm.

4.2.4 Inhibitory Assays

Each 1 mM (final concentration) of below inhibitors was added to the enzyme solution and preincubated for 1 hour at 4 °C. After that, 10 mM (final concentration) of fluorogenic Bz-Val-Gly-Arg-AMC substrates was added. And then, the reaction mix was incubated for 30 or 60 min at 65 °C. After adding of 900 µl buffer fluorescence was measured at the wavelength of 460 nm by exciting UV at 360 nm. The inhibitory effects were taken to be valid only on the case that the enzyme preincubated with an 1 mM inhibitor (final concentration) exhibited less than relative residual activities of 15 %, compared with same amount of uninhibited enzyme.

Serine protease inhibitors: PMSF, benzamidine

Proteasome inhibitors: vinyl sulfone derivative, epoxymycin, Ac-Leu-Leu-methioninal

Chloromethyl ketone-based inhibitors: N-α-tosyl-$_L$-lysyl-chloromethyl ketone, N-tosyl-$_L$-phenylalanyl-chloromethyl ketone, D-Phe-Pro-Arg-chloromethyl ketone, H-Ala-Ala-Pro-Val-chloromethyl ketone, Decanoyl-Arg-Val-Arg-Lys-chlorometyl ketone, MeOSuc-Ala-Ala-Pro-Val-chloromethyl ketone, H-Asp-Gln-Thr-Gln-Lys-Gln-Tyr-Gln-Glu-Leu-Thr-Phe-Phe-chloromethyl ketone

Others: ethyl-2-oxo-4-phenylbutyrate, β-fluoropyruvate, ethyl-3,3,3-trifluoropyruvate, benzyloxycarbonyl (Z)-PheΨ[CO-CONH]Arg-Glu-Phe-OH)

4.2.5 Proteolytic Activity Assays

Oxidized insulin B-chain, fluorescein isothiocyanate (FITC)-labeled casein, secretin, or secretinamide were incubated with each purified wild-type and mutant protein sample at 65 °C in a reaction volume of 400 µl in a buffer composed of 20 mM Tris·HCl and 100 mM NaCl. For control reactions, a 100 µl aliquot was taken immediately after substrates and the protease were mixed. At different times, further 100 µl aliquots were taken and mixed with the H-Ala-Ala-Pro-Val-chloromethyl ketone (AAPVCMK) inhibitor to stop the further cleavage reaction.

4.2.6 Reverse-Phase HPLC Analysis of Degradation Products

80 µl of reaction mixture, which was taken from the above step, was analyzed by reverse-phase HPLC on a X-Terra-MS C8 reverse-phase column. The column was equilibrated with 5 % of acetonitrile (AcCN) and 95 % of 2 % H_3PO_4 and eluted with a linear gradient of 5 to 90 % of AcCN and 95 to 10 % of 2 % H_3PO_4 in 20 min at a flow rate of 1.5 ml/min. Degradation products of oxidized insulin B-chain, secretin and secretinamide were detected by UV at 210 nm and of FITC-labeled casein at 440 nm, respectively.

4.2.7 Alkylation

Excess amounts (about 10 times of the protein molar concentration) of N-ethyl maleimide (NEM) or fluorescein-5-maleimide (FMI) were added to the wild-type and the cysteine mutant enzymes, and modification reactions were carried out overnight at 4 °C. Unreacted maleimide derivatives were removed by overnight dialysis.

4.2.8 Oxidation

1 mM (final concentration) oxidized glutathione was added to the wild-type and the double cysteine mutant enzymes (R414C and A643C) and incubated overnight at 4 °C. Residual glutathione was removed by extensive dialysis.

4.3 Crystallography

4.3.1 Crystallization

Final concentration of the purified tricorn protease, 5 mg/ml, for crystallization was calculated by measuring the UV at 280 nm and converting the value considering the extinction coefficient (1.05 in case of the tricorn protease) calculated from the primary amino acid sequences of the tricorn protease. Initial screenings were performed at 18 °C by vapor-diffusion using Cryschem sitting-drop plates where 1 μl of protein solution was mixed with an equal volume of reservoir solutions of various screening kits purchased from Hampton (USA) and home-made grid screenings. Among several hundreds screening conditions, crystals were obtained in the condition of 20 % (v/v) isopropanol and 0.1 M MES, pH 6.0. Thin-plate clustered crystals were improved simply by modifying this initial condition. In consequence, crystals suitable for the X-ray diffraction were obtained using the hanging-drop vapor-diffusion method by mixing 1.5 μl of 5 mg/ml protein solution and 0.9 μl of the precipitant solution, which is composed of 10 % (v/v) isopropanol and 0.1 M MES, pH 6.4.
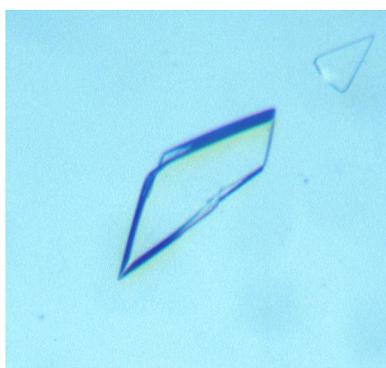
Figure 7. Crystals of the tricorn protease from *Thermoplasma acidophilum*. Crystals have grown at room temperature within two weeks to the maximal size of $0.4 \times 0.2 \times 0.05$ mm^3. Crystals diffracted to 3.5 and 2.0 Å resolution at home and at ESRF, respectively. They belong to the space group of $P2_1$ and contain one hexamer in the asymmetric unit.

Crystals were grown to a maximum dimension of $0.4 \times 0.2 \times 0.05$ mm^3 at room temperature within two weeks (Figure 7). The mixing ratio of the protein solution versus the precipitant solution was a critical factor in obtaining better crystals,

because many small crystals appeared during the overnight equilibration if protein and precipitant solution had been mixed with an equal or a precipitant excess ratio.

4.3.2 Data Collection

4.3.2.1 Mounting in Capillaries at Room Temperature

For checking the diffraction quality of the tricorn protease crystals, crystals were mounted in capillaries at first. Single crystal was transferred into the capillary at room temperature and the residual solution around the crystal was removed. Both ends of capillary were filled with the solution of crystallization condition, and sealed with bee-wax. Mounted crystals diffracted to around 3.5 Å resolution, but the diffraction intensities were very week. Furthermore, crystals had radiation damage during the data collection.

4.3.2.2 Cryo-Crystallography

Since crystals were very sensitive to the radiation at room temperature and the high resolution data were needed, cryo-conditions for the cryo-crystallography were searched against various organic solvents, polyethylene glycols of low molecular weight and some salts. Added 25 % (v/v) of glycerol (final concentration), which is one of the most commonly used cryo-protectants, to the crystallization condition seemed to give a good cryo-protectancy under the liquid nitrogen stream. This condition, however, made smear ice-rings around 3.7
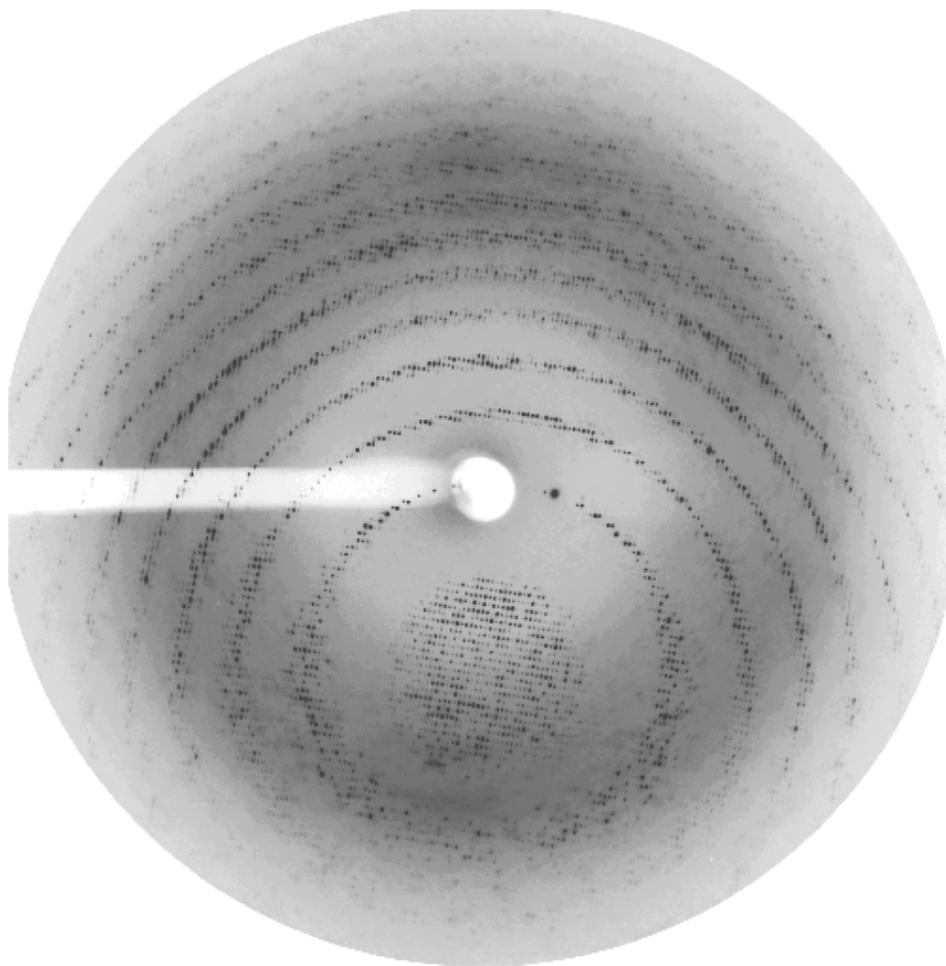
Figure 8. Diffraction image of the tricorn protease crystal. The image was recorded on a MarCCD detector on SLS protein beamline at PSI, Villigan with a rotation of 0.5° during exposure. The detector edge corresponds to a limiting resolution of 2.57 Å at an X-ray wavelength of 0.9798 Å.

Å resolution repeatedly. On the other hand, crystals were melted, when more than 25 % (v/v) of glycerol was added. Furthermore, the other commonly used cryo-protectants failed to protect crystals from the cold shock. Then, saturating erythritol (a four-carbon sugar) in the crystallization condition was checked as an alternative cryo-protectant, which protected crystals. However, crystals were broken when they were moved into the cryo-buffer in which erythritol was fully saturated. Therefore, erythritol was directly added as a salt into the drop, which contains a single crystal. Complete saturation was confirmed by a clear solution and no ice-ring under the cryo-stream. Erythritol as a cryo-protectant, however, did not work with crystals, which were obtained from different batches, especially with seleno-DL-methionine

incorporated crystals. In those cases, 25 % (v/v) of MPD (final concentration) was used for the cryo-crystallography.

4.3.2.3 Data Collection

Crystals were mounted at –173 °C in the nitrogen streams of Cryostream Cooler (Oxford Cryosystems, England) on a 345 mm imaging-plate detector mounted on a RIGAKU rotating anode X-ray generator (MAR Research, Hamburg), after being immerged in a suitable cryo-solution. Although crystals diffracted to around 3.5 Å resolution, weak diffraction intensities resulted in a high R*sym*. Therefore, high brilliant synchrotron beam was used for the better data (Figure 8). The high resolution and the tantalum bromide derivative data for the anomalous signal were collected at ID14-4 of European Synchrotron Radiation Factory (ESRF) in Grenoble (Table 2). Native crystals unisotropically diffracted to 1.8 Å resolution along one axis and to 2.0 along another axis. The crystals belong to the space group of $P2_1$ with an unit cell dimension of a = 96 Å, b = 250 Å, c = 160 Å, α = 90 °, β = 105 °, and γ = 90 °. Assuming one hexamer in the asymmetric unit leaded to a crystal volume per unit molecular weight (Vm) of 2.58 $\text{Å}^3/\text{Da}$, corresponding to a solvent content of about 50 % (Matthews, 1968).

4.3.2.4 Derivative Search

Initial trials to find a heavy-atom derivative were done against various heavy atom clusters, considering a huge molecular mass in the asymmetric unit (> 700 kDa). It covered 9 kinds of tungsten clusters, 8 kinds of mercury clusters, 2 kinds of gold clusters, 7 kinds of platinum clusters, 2 kinds of iridium clusters, old- and new-made tantalum bromide clusters, 1 tantalum-osmium mixed cluster, 1 rhenium cluster, 1 bismuth cluster and 1 nebidium cluster. Among them, the heavy atom position inside the crystal was possibly located with the tantalum bromide cluster.

4.3.2.4.1 $Ta_6Br_{12}^{2+}$

Old tantalum bromide cluster was readily soluble in the crystallization solution and made crystals green within 1 hour. Soaked crystals, however, were equilibrated at room temperature overnight for the data collection. Crystals diffracted to above 3.5 Å resolution under the synchrotron radiation. Nonetheless, the heavy atom positions were not possibly located. After that, a new tantalum bromide cluster was prepared, but it was not soluble into the crystallization solution. Some synthetic mother liquors were tested, in which crystals were stable and the new tantalum bromide cluster is readily soluble at the same time, by replacing a major precipitant, isopropanol, with alcohols, PEGs of low molecular weight and glycerol. Crystals were stable in about 15 % of ethanol (v/v), PEG400 (w/v), and glycerol (v/v) without isopropanol, respectively. However, tantalum bromide cluster was not soluble at pH 6.4 in the solutions of above compounds. The pH of a buffer was changed to 5.6, 7.5, 8.5, from 6.4 at the next step. Optimal condition was confirmed at 15 % of (v/v) glycerol or ethanol and 0.1 M Hepes, pH 7.5. Among them, glycerol-containing solution was adopted since glycerol itself might be used as a cryo-protectant. To minimize the crystal damage during the transferring step into a new synthetic buffer, crystals were moved, step-by-step, by increasing the glycerol content and decreasing the isopropanol content simultaneously, to 0 %. At the same time, the pH of solution was also increased by changing the mixing ratios of the MES, pH 6.4 buffer and the Hepes, pH 7.5 buffer, finally to the Hepes, pH 7.5 buffer. These sequential transferring steps into the ideal synthetic buffer occurred during 2 days. Then, tantalum bromide was added into the drop as a salt. Crystals were equilibrated overnight in that solution and turned completely green.

4.3.2.4.1.1 Data Collection for Phasing

The data for the tantalum bromide cluster-soaked crystal were collected at low resolution (to 6 Å) on a 345 mm imaging-plate detector mounted on a RIGAKU rotating anode X-ray generator (MAR Research) at the beginning and the position of this heavy atom cluster was located considering several aspects (will be described in 5.1 Structure Determination). For phasing with this derivative, data collections for the multi-wavelength anomalous dispersion (MAD) were performed. This crystal

diffracted and data were collected to 4.2 Å resolution at ESRF. The wavelengths at the K-shell edge of bromide (Table 2) were selected after the fluorescence was scanned, since it was not possible to locate the wavelengths to the LIII edge of tantalum on the ID14-4 beamline of ESRF. Nevertheless, the data at the remote wavelength could not be collected since this crystal was quite sensitive to the radiation. On the other hand, the data of same derivative at the LIII edge of tantalum (Table 2) were collected at BW6 of Deutsche Electonische Synchrotron (DESY) in Hamburg. Data were processed and merged by DENZO and SCALEPACK package (Otwinowski & Winor, 1997).

4.3.2.4.2 Selenomethionine-Incorporated Protein

Although a large number of heavy atoms were tried, only one heavy atom derivative was found. In order to get another heavy atom derivative for the better phasing and the more clear densities, and to be helpful for the sequence tracing during the model building, selenomethionine-incorporated protein expression and crystallization were adopted.

4.3.2.4.2.1 Expression

Seleno-DL-methionine isoform of the tricorn protease was expressed in *E.coli* strain B834(DE3), which is an auxotroph for methionine. A single colony of *E.coli* B834(DE3) carrying the expression vector of the tricorn protease, which was grown on the LB-agarose plate with ampicillin, was used to inoculate 100 ml of new minimal media with normal methionine and 0.1 mg/ml ampicillin (final concentration), in which some salts and trace elements were amplified (Budisa *et al.*, 1995). This culture was grown at 30 °C until the exponential phase reached, corresponding to an OD_{600nm} of 0.5. Bacterial cells were centrifuged at 5,000 rpm for 5 min at 4 °C. The pellet was washed two times by and resuspended in 60 ml of new minimal media with seleno-DL-methionine from Sigma. 6 l culture of supplemented new minimal media containing seleno-DL-methionine and 0.1 mg/ml ampicillin was inoculated with 60 ml of bacteria suspension and grown at 30°C. When the culture density had reached a cell density OD_{600nm} of 0.8, adding 1 mM of IPTG (final concentration) induced the selenomethionine-incorporated tricorn protease

expression. After overnight culture, cells were harvested by pelleting at 6,000 rpm for 30 min at 4 °C.

4.3.2.4.2.2 Purification and Crystallization

The selenomethionine-incorporated protein was purified by following the same purification steps as the wild-type enzyme. However, good crystals were not reproduced at the same crystallization condition as the wild-type protein. New screening was tried and the initial condition of 35 % (w/v) PEG550MME, 100 mM Tris·HCl and 10 mM $MgCl_2$ was found. Better crystals with a similar shape to, and same cell parameters as, the wild-type were obtained at the slightly modified conditions (25-30 % (w/v) PEG550MME, 50 mM Tris·HCl and 4 mM $MgCl_2$).

4.3.2.4.2.3 Data Collection

Two kinds of crystals for the selenomethionine-incorporated protein were prepared: one as an uninhibited form and the other as a TLCK inhibited cocrystal. Three data sets were collected on BW6 in DESY after fluorescences were scanned. Two data sets were from the inhibitor cocrystal at the wavelengths of the peak and the inclination point of selenium K-edge, since this crystal resulted in better statistics. Another data were from the uninhibited crystal at the peak wavelength.

4.3.3 Graphical Representation

Illustrations of structures were prepared using Molscript (Kraulis, 1991) or Bobscript (Esnouf, 1997) and in some cases rendered with Raster3D (Merrit & Bacon, 1997). Surface illustrations were prepared with GRASP (Nicholls *et al.*, 1993). Sequence alignment was made using the Fasta and plotted with Alscript (Barton, 1993).

4.3.4 Cocrystallization and Inhibitor Soaking

For the crystal structures of the tricorn protease in complex with inhibitors, protein solution was mixed with 1 mM of inhibitors (final concentration) and equilibrated for more than 1 hour. Complex crystals were obtained by a cocrystallization method or

soaking. In the cases of soaking, the inhibitor solution was added into the drop, which contains a single and good shaped wild-type crystal, or an inhibitor was directly added as a powder after the inhibitory efficacy was checked. The drop was equilibrated at room temperature.

5 Structure Description

5.1 Structure Determination

The structure was determined by a combination of SIRAS, MAD, and non-crystallographic symmetry (NCS) averaging. The molecular center was unambiguously determined by (a) inspection of the Harker plane calculated at low resolution (80 - 8 Å); (b) correlation calculations employing the local 3-2 symmetry; (c) analysis of a pseudo $C222_1$ symmetry at very low resolution (80 - 15 Å); and (d) packing considerations. The tantalum bromide derivative could be identified by the coincidence of the heavy atom local symmetry center (Figure 9) with that of the protein boundaries in agreement with that of protein using the program RSPS (Knight, 2000). Solvent flattening using Solomon (Abrahams & Leslie, 1996) identified boundaries in agreement with electron microscopic data (Tamura *et al.*, 1996a). A 3 degree tilt of the molecular orientation relative to the self-rotation result (GLRF) could be resolved by a 3-3 NCS correlation scan analysis of the cut out SIRAS electron density using routines in MAIN (Turk, 1992). For NCS averaging, a protein mask was built dynamically in a two step procedure by initially generating a protein mask using 35 % solvent (Solomon) and subsequentially restricting that mask to the 3-2 symmetric area (MAIN). The density improvement stalled at 5.5 Å resolution. Therefore, multi-crystal averaging using slightly non-isomorphous data sets was employed. The NCS and multi-crystal transformation were refined by rigid body minimizing a cut out electron density using appropriate routines in the program AMoRe in the CCP4 Program package. Real space electron density averaging was performed using MAIN in combination with CCP4 routines. The phases derived from the NCS averaged electron density were employed to calculate an anomalous difference fourier map using seleno-DL-methionine incorporated tricorn protein crystal data (Table 2). This map unambiguously identified 144 selene sites which allowed straightforward sequence alignment during model building using the program MAIN. The model including waters was refined by using X-PLOR package and CNS package (Brünger *et al.*, 1998). The quality of the final model was analyzed with PROCHECK (Laskowski *et al.*, 1993). Finally refined model shows favorable

geometries (Figure 11) except two residues, Ser965 and Asp1017, which are located in the disallowed region even though the densities are quite clear. The refinement statistics are quite good considering the huge molecular weight in the asymmetric unit (Table 3). Final model shows R_{cryst} = 24.8 %, R_{free} = 25.9 %, rms_{bond} = 0.08 Å, and rms_{angle} = 1.0 °, and surprises a total of 49,068 non-hydrogen protein atoms and 2394 water molecules with average crystallographic temperature factors of B_{prot} = 31.1 Å^2 and B_{wat} = 33.3 Å^2, respectively. The average deviation of NCS-related subunits is 0.01 Å (Table 3).

On the other hand, the crystal structures in complex with inhibitors were directly refined by CNS package (Brünger *et al.*, 1998) using the wild-type tricorn protease structure as a model.

A
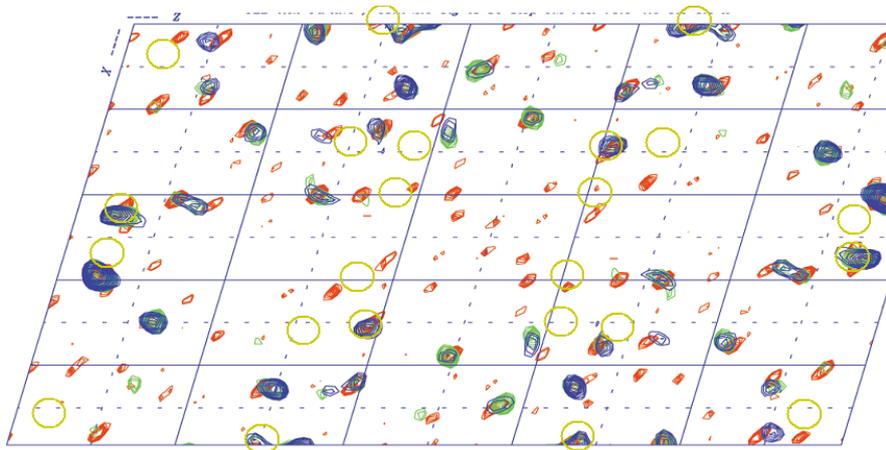
B

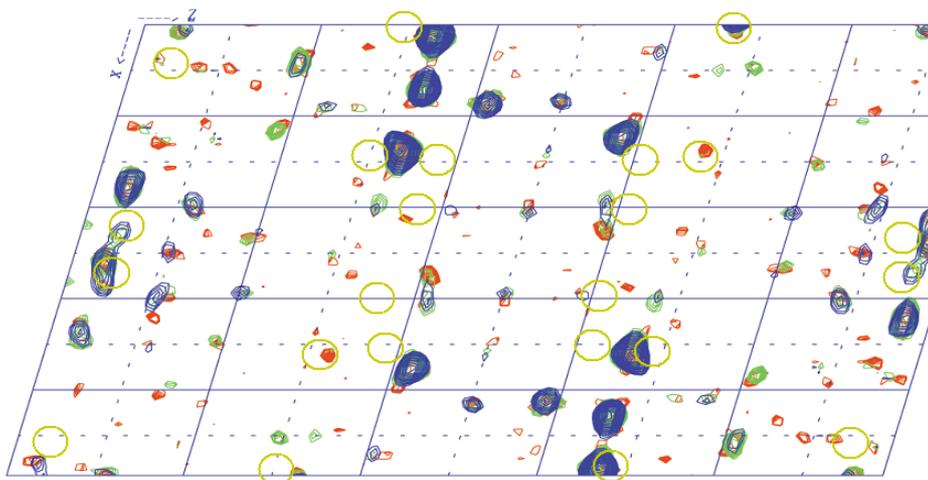
Figure 9. Plots of the *v*=1/2 Harker section of the isomorphous (A) and anomalous (B) difference patterson map for the tantalum bromide derivative of the tricorn protease from *Thermoplasma*

acidophilum, calculated with different resolution limits (red: 6.0 Å, green: 5.0 Å, blue: 4.5 Å). Structure factors were calculated with SFALL and Patterson maps with FFT. The plots were done with NPO.

Table 2. Data collection and phasing statistics.

Compound	Wave-Length (Å)	Resolu-tion (Å)	Number of observa-tions	Unique reflections	Complete-ness (%)	Rsym (%)[1]	Riso (%)[2]	Number of sites	Phasing power (a/c)[3]	Rcullis (a/c)[4]
Native (uninhibited)	0.9390	50-2.0	1040404	412325	83.0	9.7/50.4	-	-	-	-
Native (TLCK)	0.9390	50-2.7	545518	152288	97.9	16.2/44.0	-	-	-	-
Ta_6Br_{12}	1.5418	80-7.5	32675	7928	87.8	14.6/21.1	98.8	6	1.60/1.02	0.73/0.72
Ta_6Br_{12} (Br peak)	0.9184	80-4.2	171381	47744	97.4	15.0/51.3	62.4	6	2.61/1.61	0.57/0.68
Ta_6Br_{12} (Br incl.)	0.9202	80-4.2	171914	48103	97.5	17.8/55.0	60.7	6	2.57/1.69	0.58/0.66
Ta_6Br_{12} (Ta peak)	1.2548	80-4.2	102278	47247	91.4	14.3/34.0	35.9	6	1.35/1.00	0.81/0.81
SeMet (uninh., peak)	0.9793	20-2.6	1072179	405224	95.8	18.7/55.0	27.9	144	0.77/0.68	0.91/0.89
SeMet (TLCK, peak)	0.9793	20-2.9	520749	267077	85.4	11.2/20.4	51.7	144	0.69/0.84	0.89/0.88
SeMet (TLCK, incl.)	0.9799	20-2.9	517365	265754	85.9	11.0/20.7		144	0.91/0.83	0.89/0.90
Osmium (uninh.)	1.0011	20-2.9	449591	138034	99.3	10.4/23.3	37.4	n.s.[5]	-	-

[1] $Rsym = \Sigma_{hkl} \Sigma_j |I_j - <I>| / \Sigma_{hkl} \Sigma_j I_j$; where $<I>$ is the mean intensity of reflection hkl.

[2] $Riso = \Sigma_{hkl} |F_{PH} - F_P| / \Sigma_{hkl} F_P \times 100$; where F_{PH} and F_P refer to the derivative and the native structure factor, respectively.

[3] Phasing Power (acentric/centric) $= <|F_H|> / <|F_{PH} - |F_P + F_H||>$; where F_H denotes the heavy atom structure factor. The denominator is also known as residual lack of closure (E).

[4] $Rcullis = \Sigma_{hkl} |F_{PH} - |F_P + F_H|| / \Sigma_{hkl} |F_{PH} - F_P|$ = lack of closure / isomorphous difference

[5] No heavy atom sites were determined; this data set was used for multi-crystal averaging

Table 3. Refinement statistics for the crystal structure of the tricorn protease.

X-ray source	ESRF
Unit cell (Å, °)	95.86 245.998 159.04 90 105.30 90
Space group	P21
Resolution range (Å)	50-2.0
Total reflections	1040404
Unique reflections	412325
Redundancy	2.52
Completeness (%)	83.0
Rsym[1]	9.7
R$factor$[2]	24.8
$Rfree$[3]	25.9
No. of protein	49062
No. of water	2394
Rmsd for bonds (Å)	0.01
Rmsd for angles (°)	1.0
Rmsd for NCS (Å)	0.01
Average B-factors (Å^2)	
protein	31.1
solvents	33.3
Geometry	
most favored regions	86.4 %
additionally allowed regions	12.7 %
generously allowed regions	0.7 %
disallowed regions	0.2 %

ESRF, European Synchrotron Radiation Factory in Grenoble, France

[1] R$sym = \Sigma_{hkl} \Sigma_j | I_j\text{-}<I> | /\Sigma_{hkl}\Sigma_j I_j$: where $<I>$ is the mean intensity of reflection *hkl*.

[2] $R_{factor} = \Sigma_{hkl} | |F\text{obs}|\text{-}|F\text{calc}| | / \Sigma_{hkl} |F\text{obs}|$; where Fobs and Fcalc are respectively, the observed and calculated structure factor amplitude for reflections hkl included in the refinement.

[3] R_{free} is the same as R_{factor} but calculated over a randomly selected fraction (4.2 %) of reflection data not included in the refinement.
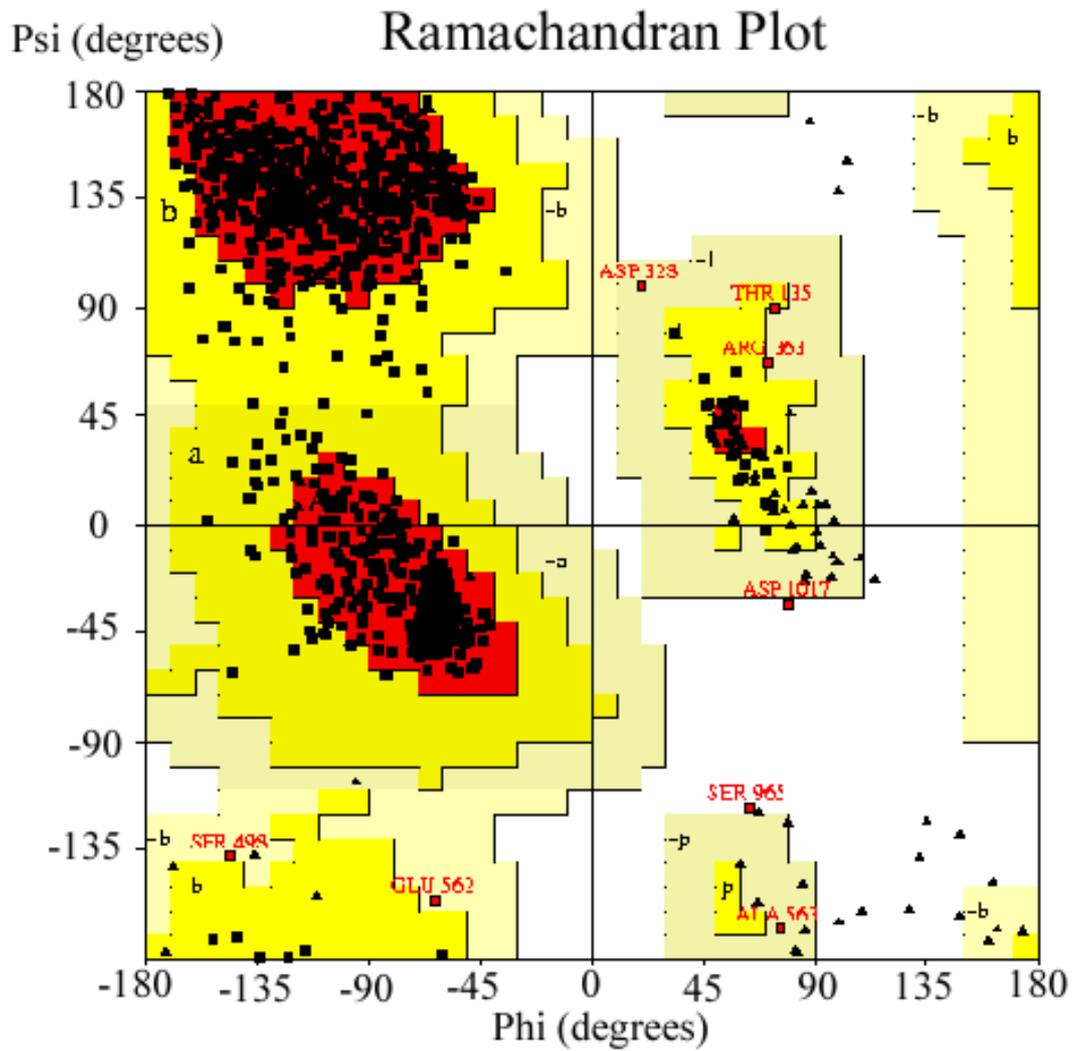
Rmsd; root-mean-square deviation.

Figure 10. Ramachandran plot of the refined tricorn protease crystal structure at 2.0 Å resolution: 86.4 % in the most favored region, 12.7 % in the additionally allowed region, 0.7 % in the generally allowed region and 0.2 % (Ser965 and Asp1017) in the disallowed region.

5.2 Structure Description

5.2.1 Overall Structure

The hexameric D3-symmetric tricorn protease is assembled by two perfectly staggered and interdigitating trimeric rings, with each subunit of one ring forming contacts almost exclusively with the two subunits of the other ring related by the molecular diads. The toroid structure has the shape of a distorted hexagon (Figure 11A). The dimers forming the long sides of the hexagon are intimately intertwined. By contrast, the interface of the subunits forming the short side is very flat. As a consequence, the two resulting subunit interfaces differ drastically in their respective contact area (6,233 and 1,683 Å^2) and assemble the tricorn protease as a trimer of dimers (Figure 11A). The overall dimensions of the molecule are 160 Å within the plane normal to the three-fold axis and 88 Å parallel to it (Figure 11B). The conically shaped central pore measures 45 Å in diameter at its entrance and 20 Å close to the center of the molecule where the pore connects with additional cavities formed by individual subunits like spokes of a wheel, Figure 12A, in qualitative agreement with electron microscopic data (Tamura *et al.*, 1996a; Figure 2). The internal oblate cavity has a volume of about 97,000 Å^3, measuring approximately 85 Å and 8 Å in its diameter and height, respectively.

The positions of all but the terminal amino acids are defined in the electron density (Met39-Asn1061). The hexameric ring formation is introduced by dimer-dimer interactions on the dimeric interfaces. Unlike the monomer-monomer interactions (will be described below) in dimer formation, the interactions on the dimer-dimer interfaces are mainly accomplished by a number of hydrogen bonds between two 6-bladed β-propeller domains, two C1 domains and two 7-bladed β-propeller domains, along with the insertion between strand 3 and 4 of blade 5 (Pro546-Tyr566).

Within the hexameric framework, three 7-bladed β-propeller domains are arranged around the molecular three-fold axis, where they line the entrance to the central pore (Figure 11A). Most of the contacts determining the relative arrangement of the 7-bladed β-propeller domain are mediated by the intervening PDZ domains of the diad-related subunits (Figure 11A). Each of the three short sides of the hexagon is formed by two 6-bladed β-propeller domains (Figure 11A). The 7-bladed β-propeller domain and the PDZ domain, together with two arms (Leu520-Val535 and Asn930-Asn949) that are interacting with a neighboring
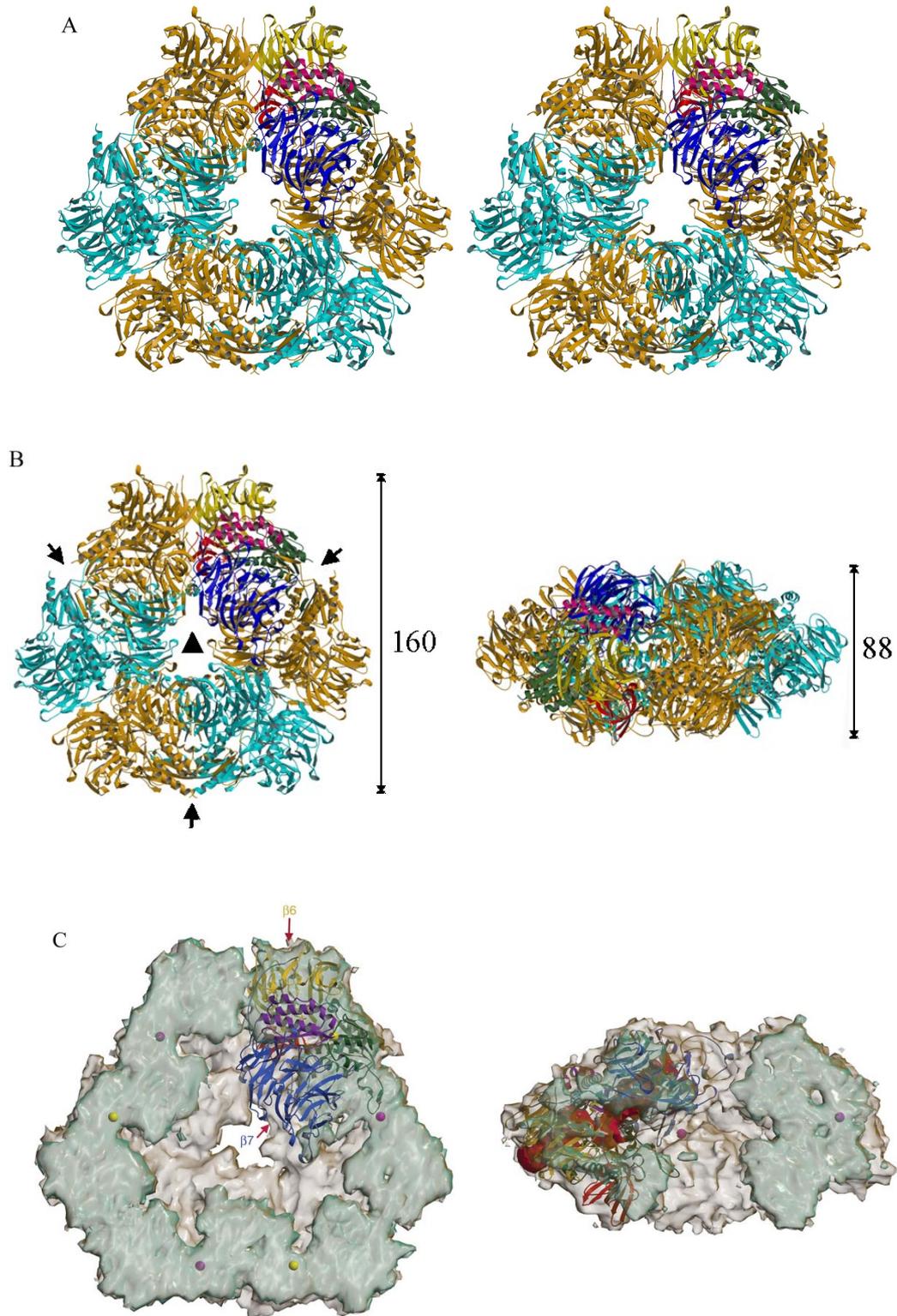
Figure 11. Structure of the tricorn protease from *Thermoplasma acidophilum*.

A) Stereoview of ribbon representations of the tricorn protease viewed along the molecular threefold axis. Individual subunits are distinguished by alternating colors.

B) Ribbon representations showing the overall dimensions of the hexameric tricorn protease (along the central threefold (left, short arrows means the molecular twofold axes), viewed from the side (right)). Individual subunits are distinguished by alternating colors. The overall dimensions of the hexameric molecule are 160 Å within the plane normal to the threefold axis and 88 Å parallel to it. The conically shaped central pore measures 45 Å in diameter at its entrance and 20 Å close to the center of the molecule.

C) Cut-open surface representation viewed along the threefold axis (left). The central pore, indicated by the gray surface, has a volume of about 97,000 $Å^3$, measuring about 85 and 8 Å in its diameter and height, respectively. It is connected by cavities to the six active sites, indicated as yellow and margenta spheres. Cut-open surface representation viewed from the side (right). The pseudo-symmetry axes of the 7-bladed β-propeller domain intersect the molecular threefold (running vertically) at an angle of 57 °, thus giving rise to the conical shape of the pore entrance. The red surface indicates internal cavities implying direct access for the substrate through the 7-bladed β-propeller domain to (blue), and product egress through the 6-bladed β-propeller domain from (yellow), the active site (right).

monomer, provide the contact interfaces in the tight dimer sub-complex of the tricorn protease (Figure 11A), and helical bundle is located between both propepper domain structures (Figure 12).

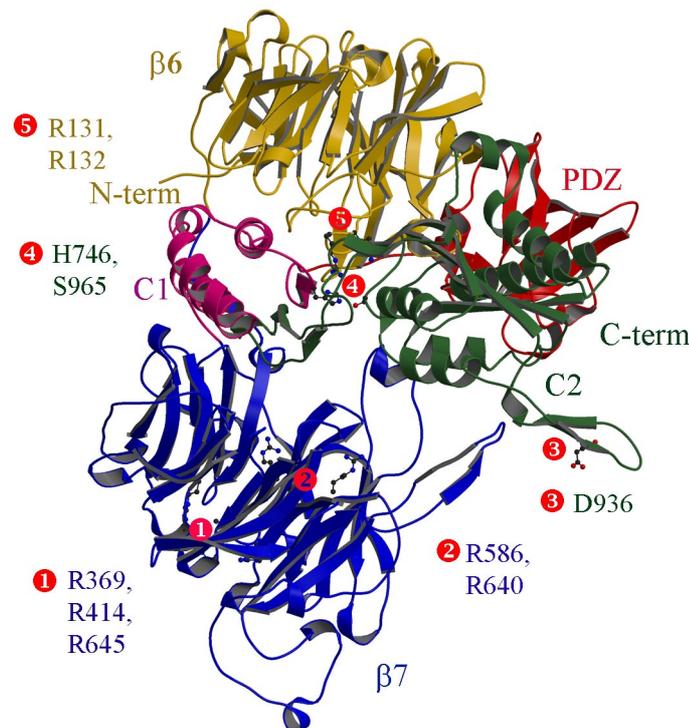## 5.2.2 Sub-Domains of Tricorn Monomer

Figure 12. Representation of the highlighted subunit of Figure 11 using identical color codings. The residues forming the propeller lids (Arg369, Arg414, and Arg645 in β7 (blue)), the residues transferring the substrate in the 7-bladed β-propeller domain (Arg586 and Arg640 in β7 (blue)), and the residues of the putative substrate-binding anchor (Arg131-Arg132, gold) are displayed as well as the catalytic residues, His746 (C1, margenta) and Ser965 (C2, dark-green). Asp936, positioned on the C2 domain, confers specificity for basic substrate residues in the active site of the neighboring monomer.

A single subunit can further be divided into five sub-domains. The amino-terminal part of the protein folds as a 6-bladed β-propeller domain (Met39-Asp310) and is followed sequentially by a 7-bladed β-propeller domain (Ala326-Lys675) and a C-terminal domain (Ser681-Asn1061). A-PDZ-like domain (Arg761-Asp855) is interspersed between the two carboxy-terminal mixed α-β domains (Ser681-Gly752, C1 and Arg856-Asn1061, C2) (Figure 12). Both the 6-bladed β-propeller and the 7-bladed β-propeller structure are topologically unclosed and oriented roughly anti-parallel to each other, thereby embracing the two loosely packed C-terminal domains, and partly the PDZ-like domains (Figure 12).

5.2.2.1 6- and 7-Bladed β-Propeller Domain

Two propeller structures are oriented roughly anti-parallel to each other, with their pseudo-symmetry axes enclosing an angle of approximately 160 °. Both the 6- and the 7-bladed β-propeller structures are topologically unclosed and organized in a clockwise direction with the pseudo-symmetry axis, defined by the innermost strands (Figures 13A, 13B). Such an open Velcro-like propeller structure has been observed only in the crystal structure of the prolyl oligopeptidase from bovine porcine (Fülöp *et al.*, 1998). The 6-bladed β-propeller domain axis is directed towards the active site of the tricorn protease (Figures 11A, 12). The otherwise direct connection from the chamber of the active site to the exterior through the 6-bladed β-propeller domain is obstructed by the arginine anchor (Arg131-Arg132), which is located on a helical loop containing three glycines (Gly126, Gly130 and Gly139) lacking any protein contacts (Figure 16B right). In the related tricorn protease homologue from *Sulfalobus solfataricus*, a segment of seven residues is deleted in this region. The 7-bladed β-propeller structure is stabilized by an external clamp, however, which is formed by two short anti-parallel β-strands (Ile318-Ser320 and Leu678-Ser680) on the in- and outgoing segments, respectively (Figure 13b). The 7-bladed β-propeller domain carries prominent insertions

between stand 2 and strand 3 on blades 4, 5, and 6 (His469-Val476, Asp521-Val535, Val602-Glu615) which mostly mediate contacts to the diad-related subunit. A further insertion between strand 3 and 4 of blade 5 (Pro546-Tyr566) is oriented towards the central channel where it also mediates subunit contacts. The 7-bladed β-propeller channel is wide open but capped on its outside by four basic residues (Arg369, Arg414, Arg645, Lys646) which are only partially charge compensated by one acidic residue (Asp456). This locally positive lid to the 7-bladed β-propeller domain is encircled by acidic residues (Asp333, Asp335, Asp372, Asp456, Asp506, Asp508, Asp592, Glu663). Excepting Glu663 that is located on the hairpin connecting strand 3 and 4 of blade 7, all these charged amino acids are positioned between strand 1 and strand 2 of the 7-bladed β-propeller domain (Figure 13C).
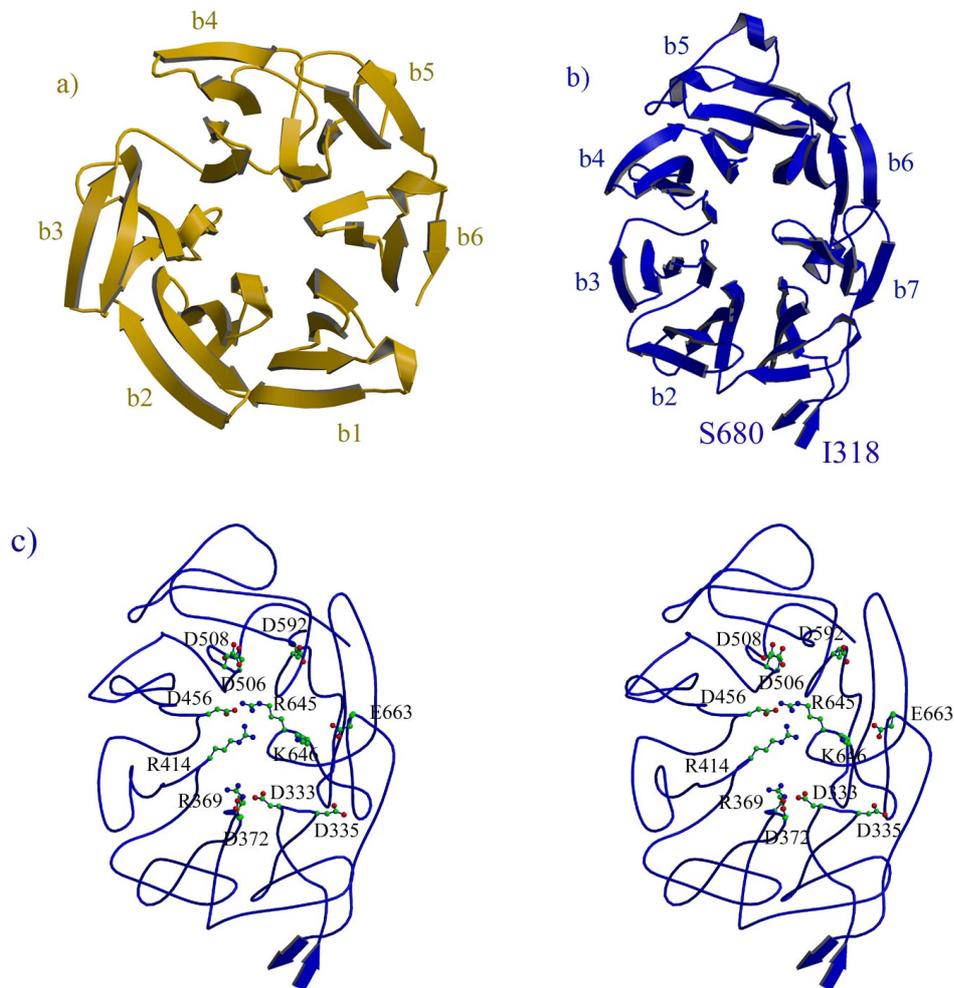

Figure 13. Two open Velcro propeller structures of tricorn monomer using the same color coding as Figure 11. (a) 6-bladed β-propeller structure. (b) 7-bladed β-propeller structure. Both propeller structures lack disulfide bridges and strand exchange between blades. b2, b3 *etc.* means blade 2, blade 3 *etc.* on a counter-clockwise direction, respectively. c) Stereoview of 7-bladed β-propeller domain showing charge distribution on the entrance surface.

5.2.2.2 C-Terminal Domain

Each of the three C-terminal domains (C1, PDZ, C2) is remarkably similar to the respective domains (A, B, and C) found in the D1 processing protease (D1P) of photosystem II (Liao *et al.*, 2000; Figure 14). The rms deviation between the Cα positions of these domains are 2.2 Å, 2.3 Å, and 2.7 Å with 84, 86, and 135 matching amino acids, respectively. A weak homology between these domains is recognizable in the primary sequences (11 %, 19 %, and 20 % identities). The relative arrangement of these domains, however, differs very much between the tricorn protease and D1P: With the C2 domain aligned to the C domain of D1P, the orientation of C1 domain differs from that of the D1P A-domain by 35 degrees. Analogously, the required transformation to align the PDZ-like domains includes a 96 degree rotation. The rotation axes of these transformations are unrelated to each other. In addition,
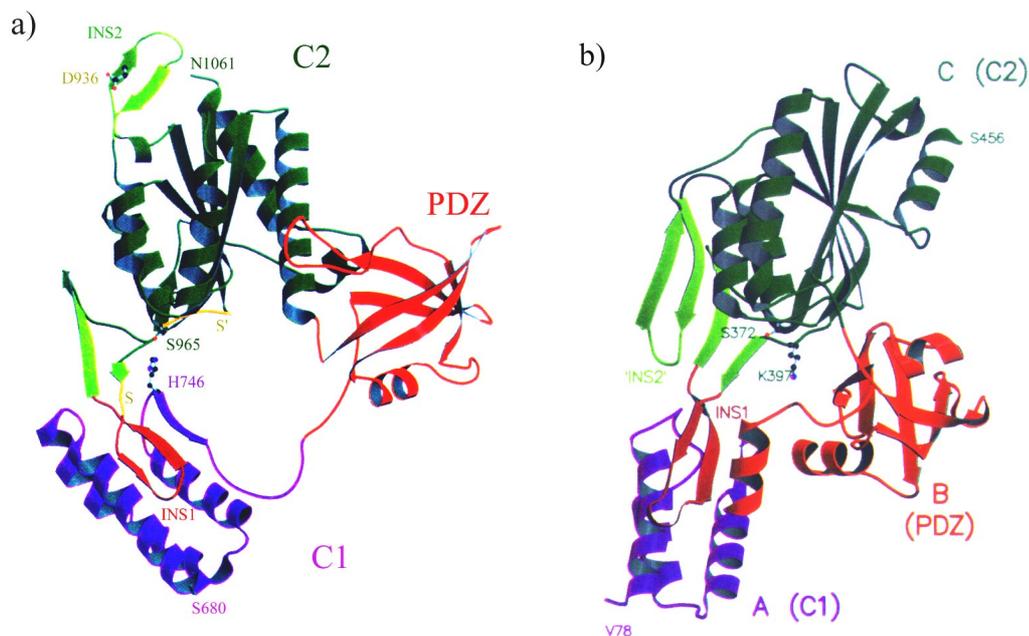

Figure 14. Ribbon representation of the tricorn protease C-terminal domain (a) and its functional analogous protein, D1P protease (1fc7, b). Individual domains are distinguished by alternating colors and labeled with same colors.

proper alignment of the PDZ domain requires a 30 Å translation. The catalytic serine residues (Ser965 and Ser372, respectively) are positioned on topologically equivalent positions at the helix entrance in the C2 (C) domain (D1P). Further, the amides forming the oxyanion hole (Gly918, Asp966, and Gly318, Ala373 in the tricorn protease and D1P, respectively)

superimpose to within 1 Å. As in other tail specific-like proteases, the residue serving as a general base in D1P is a lysine (Lys372) residing within the C domain of D1P, while it is a histidine in the tricorn protease (His746) that resides on the C1 domain of the tricorn protease.

5.2.2.2.1 C1 and C2 Domain

The core domains, C1 and C2, are closely associated by a small sequential insertion within C2 (Gly993-Pro1009), integrated into the C1 domain. A further prominent sequence insertion, Ile933-Pro947, is structurally integrated into the diad-related C2 domain (Figure 14A). It reflects the intimate dimer contact on a structural level. The C1 domain is built as an anti-parallel three-helix bundle, which is capped on its concave side by an anti-parallel three-stranded β-sheet. The αβ-mixed sandwich-like C2 domain consists of a five stranded pleated parallel β-sheet flanked by two and three helices on either side (Figure 14A). The C-terminal helix is oriented away from the central channel and positioned close to the diad axis.

5.2.2.2.2 PDZ Domain

The PDZ-like domain in the tricorn protease is consisting of twisted five-stranded anti-parallel β-strands and a two-turn and a single-turn α-helix (Figure 14A). Hydrophobic residues, Phe769, Ile765, Ile795, Ile807, Ile821, Leu825, Ile835 and L837, form the core of the module (Figure 15). The structure is well ordered with an average B-factor of 30.28 Å^2. It is positioned in the plane of the 6-bladed β-propeller domain (Figures 11A, 12). It deviates considerably from PDZ domains identified so far by its spatial architecture and its topological
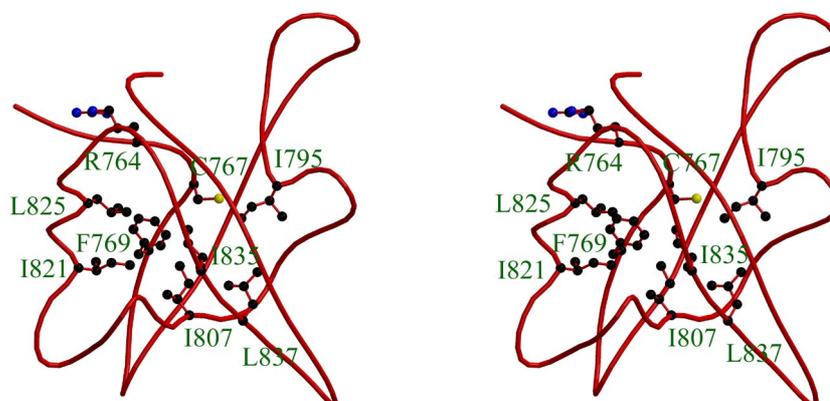
Figure 15. Stereoview of the PDZ domain that shows the conserved $R^{764}IAC^{767}$ motif and its central hydrophobic core. The PDZ segments were drawn as coils with the same color coding as Figure 11. The PDZ domain takes part in the dimerization (compare with Figure 11A).

alignment. The rms deviation to the closed related structure is 2.3 Å (PDB entry 1fc6; Figure 14), as determined by the 3-D alignment program DALI (Holm & Sander, 1993). The classic peptide recognition element including the GLGF motif (Doyle *et al.*, 1996; Eichinger *et al.*, 1999) is structurally conserved in the PDZ domain of the tricorn protease ($R^{764}IAC^{767}$).

However, it appears for a number of reasons unlikely that the PDZ domain of the tricorn protease will take part in substrate recognition as suggested for D1P (Liao *et al.*, 2000): (i) The putative substrate binding site as defined by the crystal structures of C-terminal peptide in complex with PDZ domains (Doyle *et al.*, 1996; Eichinger *et al.*, 1999) is partly occupied by other strands of blade 3 of 6-bladed β-propeller domain within the same subunit; (ii) the generally conserved arginine (Arg247) involved in recognition of the carboxylate of the peptide C-terminus corresponds to a hydrophobic residue in the tricorn protease (Ile851); (iii) the orientation and position of PDZ domain of the tricorn protease differs so strongly from that seen in D1P that any analogy based on the sequential domain arrangement is invalidated on the basis of their respective 3-dimensional domain arrangement. On the contrary, PDZ domain mainly serves to scaffold the sub-domains as described earlier and, in addition, might be involved in recognition of associating component proteins.
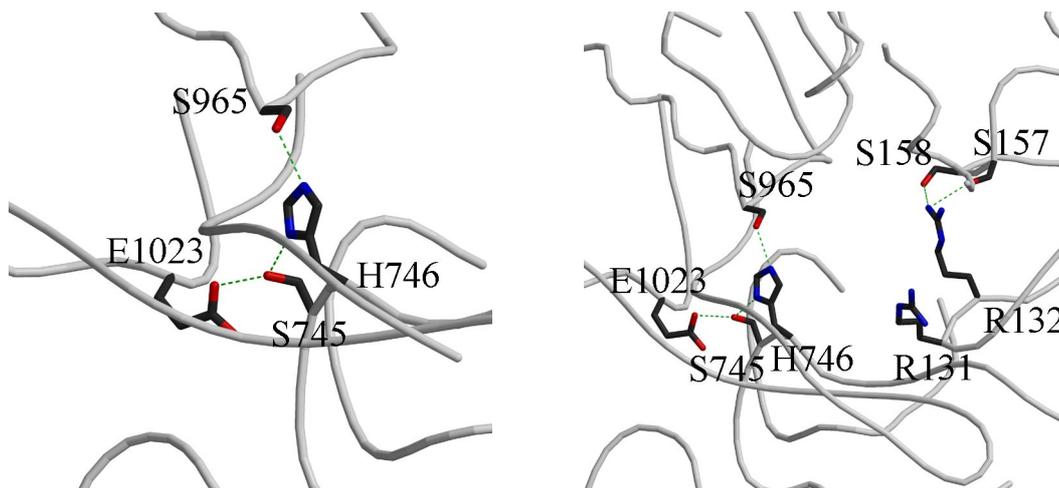
5.3 Active Site

A
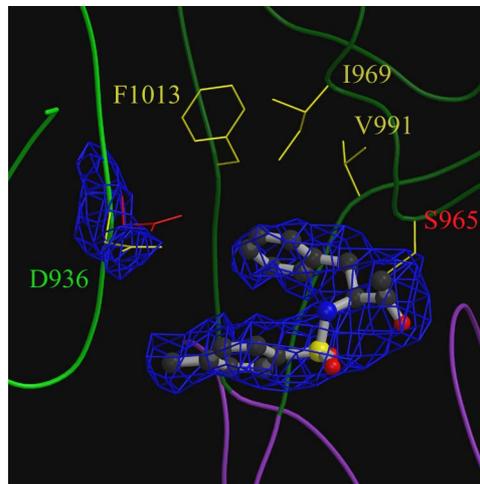
B

Figure 16. (A) Omit electron density at 3.0 Å resolution showing the TPCK inhibitor bound to the tricorn protease. Tricorn segments were drawn as coils with same color coding as Figure 11. The binding of the phenyl residue in P1 induces a conformational shift of Asp936 originating from a neighboring monomer, as compared with the position observed in the uninhibited and TLCK inhibited structure (red), thereby making the hydrophobic niche formed by Phe1013, Ile969, and Val991 more accessible.

(B) Representation of catalytic residues as stick models. Tricorn segments were drawn as grey coils and some residues around the active site as sticks. The catalytic Ser965 is activated by His746, which is oriented properly by the sequential hydrogen-bond networks of Ser745 and Glu1023 (left). Broad range of active site includes a putative substrate binding site (Arg131-Arg132), and its hydrogen-bond partners (Ser157-Ser158) (right).

The amino acid residues crucial for its catalytic activity were visualized with the complex crystal structures of chloromethyl ketone-based inhibitors (TLCK and TPCK). For both

inhibitors, continuous electron density was observed connecting to the side chain of Ser965, which was unambiguously fitted by the respective inhibitor (Figure 16A). Ser965 is positioned at the entrance to the helix within sub-domain C2 (Figure 14A). The uncapped amino group of Asp966 forms, together with that of Gly918, the oxyanion hole which is occupied by a water molecule in the uninhibited structure. His746, which is correctly oriented by the O^r of S745 polarized in turn by E1023, is positioned ideally to activate the catalytic Ser965 at a hydrogen-bonding distance of 2.7 Å (Figure 16B). However, a covalent linkage of His746 to the inhibitor was not observed in none of the both inhibitor complexes; this situation is paralleled, for instance, by cysteine proteases where chloromethyl ketone-based inhibitors react not with the catalytic histidine but instead only with the catalytic cysteine (Eichinger *et al.*, 1999). Trypsin-like serine proteases are irreversibly inhibited by chloromethyl ketone through an additional covalent bond formation of the catalytic histidine to the methyl-ketone (Bode *et al.*, 1989). The arrangement of Ser965, His746 and the oxyanion hole suggests that steps of peptide bond hydrolysis follow the classical sequence of the trypsin-like serine protease, namely the formation of the tetrahedral adduct, the acyl-enzyme complex, and hydrolysis.

5.4 Substrate Specificity at S1 Pocket

The tricorn protease has been shown to exhibit both tryptic and chymotryptic specificities (Tamura *et al.*, 1996a; Table 1). The X-ray crystal structure in complex with TLCK reveals that specificity for basic P1 residues is conferred by Asp936 which is provided by a neighboring monomer (Figure 16A). Hereby, the previously described structural linkage (trimer of dimers) is translated into functional cooperativity within the dimers. In the uninhibited 2.0 Å resolution crystal structure, the S1 specificity determinant residue Asp936 is however mobile. The carboxylate group of Asp936 has a crystallographic temperature factor of 37 Å^2 (overall 31 Å^2) and the rms deviation between the NCS related carboxylate atoms is 0.2 Å (overall deviation 0.1 Å). Consistent herewith, the side chain of Asp936 in the TPCK complex structure adopts an alternative rotamer to allow the TPCK phenyl ring to freely access the hydrophobic niche formed by Ile969, Val991, and Phe1013 (Figure 16A). Asp936 thus serves as a substrate specificity switch accommodating both hydrophobic and basic P1 residues.

6 Mutageneses and Inhibitor Syntheses

6.1 Site-Directed-Mutageneses

The tricorn wild-type construct (pRSet6c-Tricorn with the *Nde*I and the *Xho*I restriction enzyme sites for insertion) was used as a template for all performed mutations by the Quickchange™ site-directed mutagenesis kit from Stratagene (USA). Polymerase Chain Reaction (Innis *et al.*, 1988; Saiki *et al.*, 1988) was carried out with *Pfu Turbo* polymerase (Stratagene, USA) on a stratagene Robocycler Gradient 96 (Stratagene, Heidelberg, Germany) obeying the following steps:

- Initial denaturation step (1cycle):   95 °C, 5 min;
- Elongation step (16 cycles): denaturation 95 °C, 1 min

                                     annealing 55 °C, 1 min

                                     elongation 68 °C, 14 min;

- Termination step (1 cycle): 68 °C, 10 min.

And, sample reactions were prepared as indicated below:

$10 \times$ Reaction buffer: 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 100 mM Tris·HCl (pH 8.8), 20 mM $MgSO_4$, 1 % (w/v) Triton$^®$ X-100, 1 mg/ml nuclease-free BSA

    5 µl of $10 \times$ reaction buffer

    X µl (5-50 ng) of dsDNA template

    X µl (125 ng) of the first oligonucleotide primer

    X µl (125 ng) of the second oligonucleotide primer

    1 µl of dNTP mix which corresponds 0.5 mM of each nucleotide

    ddH_2O to a final volume of 50 µl

    and 40 µl of mineral oil was added on the top of the reaction mixture.

then, 1 µl of *pfuTurbo* DNA polymerase (2.5 units/µl) was added.

XL1-Blue Supercompetent cells

(Stratagene, USA)

Genotype: *sup*E44, *hsd*R17, *rec*A1, *end*A1, *gyr*A96, *thi*-1, *rel*A1, *lac*⁻ [F' *pro*AB⁺, *lac*I�q, *lac*ZΔM15, Tn10(*tet*ʳ)].

The parental supercoiled dsDNA was digested by incubating for 1 hr at 37 °C after 10 units of *Dpn I* restriction enzyme was added. The *Dpn I*-treated DNA was transformed to the XL1-Blue supercompetent cells by heat shock for 45 seconds at 42 °C.

6.2 Design of Primers for Mutants

Individual primers for required mutations were designed strictly following the instructions of QuickChange™ Site-Directed Mutagenesis kit from Stratagene (USA):

- Primer length: between 25 and 45 bases
- Melting temperature (T_m): should be greater than or equal to 78 °C.

The following formula was used for estimating the T_m of primers:

$$T_m = 81.5 + 0.41(\%GC) - 675/N - \% \text{ mismatch}$$

N is the primer length in bases.

6.2.1 at the Active Site

6.2.1.1 S965A

5'-AACGAGTATGCCGGA<u>GCT</u>GACGGTGATATCTTC-3' for forward
5'-GAAGATATCACCGTC<u>AGC</u>TCCGGCATACTCGTT-3' for reverse

6.2.1.2 H746N

5'-AACGAATACCGCACATCG<u>AAC</u>TCCTATGAAATGGGCGGC-3' for forward
5'-GCCGCCCATTTCATAGGA<u>GTT</u>CGATGTGCGGTATTCTCC-3' for reverse

6.2.2 at the 7-Bladed β-Propeller Domain

6.2.2.1 R414C

5'-CAATGGGTGTAGAC<u>TGC</u>AATGGCAAATTC-3' for forward
5'-GAATTTGCCATT<u>GCA</u>GTCTACACCCATTG-3' for reverse

6.2.2.2 A643C (template R414C)

5'-GATCTGAGGCTGTCC<u>TGC</u>GACAGGAAGACTG-3' for forward
5'-CAGTCTTCCTGTC<u>GCA</u>GGACAGCCTCAGATC-3' for reverse

6.2.3 at the Putative Substrate Binding Site (R131E-R132E)

6.2.4.1 R131E

5'-GGAAAGAGCACCGGGG<u>GAA</u>AGGATGTTCACCGATG-3' for forward
5'-CATCGGTGAACATCCT<u>TTC</u>CCCGGTGCTCTTTCC-3' for reverse

6.2.4.26.2.4.1 R132E (template R131E)

5'-GAAAGAGCACCGGGG<u>GAAGAG</u>ATGTTCACCGATCTTG-3' for forward
5'-CAACATCGGTGAACAT<u>CTCTTC</u>CCCGGTGCTCTTTC-3' for reverse

6.2.4 at the 6-Bladed β-Propeller Domain (L184C)

5'-GGCCGGCGACACATATA<u>TGC</u>TTCGCTGATGGAAGAAG-3' for forward
5'-CTTCTTCCATCAGCGAA<u>GCA</u>TATCTGTGTCGCCGGCC-3' for reverse

6.3 Sequencing

All mutations were confirmed by PCR-based DNA sequencing method (Innis *et al.*, 1988; Saiki *et al.*, 1988).

6.4 Inhibitor Synthesis

Both inhibitors were synthesized by Prof. Moroder group of MPI Biochemie.

6.4.1 Synthesis of H-Asp-Gln-Thr-Gln-Lys-Gln-Tyr-Gln-Glu-Leu-Thr-Phe-Phe-CH$_2$Cl (TridecaCMK)

The C-terminal H-Phe-CH$_2$Cl was obtained by reaction of Boc-Phe-OH via the mixed anhydride with trimethylsilyldiazomethane (Cesar *et al.*, 2001) followed by treatment of the resulting Boc-Phe-CHN$_2$ with 6 M HCl in dioxane. The side-chain protected N^α-Boc-dodecapeptide which was synthesized on chlorotrityl-resin by standard Fmoc/tBu procedures and cleaved from the resin by exposure to CH$_2$Cl$_2$/trifluoroethanol/AcOH (8:1:1), was coupled with H-Phe-CH$_2$Cl with EDC/HOAt/DIEA. Upon exposure of the fully protected intermediate to TFA/H$_2$O/TIPS (95:4:1 (2 hours, room temperature)), the crude compound was purified by preparative reverse-phase HPLC; ESI-MS: m/z = 1708.8 [M+H]$^+$; M$_r$ = 1708.3 calculated for C$_{77}$H$_{111}$N$_{18}$O$_{24}$Cl.

6.4.2 Synthesis of Z-PheΨ[CO-CONH]Arg-Glu-Phe-OH (ZDK)

2-(*R,S*)-Hydroxy-3(*S*)-amino-4-phenylbutanoic acid (Harbeson *et al.*, 1994) was converted to the N^α-benzyloxycarbonyl derivative Z-PheΨ[CH(OH)-CO]OH and then coupled with H-Arg(Pbf)-Glu(OtBu)-Phe-OtBu with EDC/HOBt. The resulting protected tetrapeptide derivative was oxidized at the hydroxy group with Dess-Martin periodinane reagent to produce the corresponding α-ketoamide derivative. Upon exposure to 95 % TFA for 2 hours at room temperature, the crude product was purified by preparative reverse-phase HPLC; ESI-MS: m/z = 760.2 [M+H]$^+$; M$_r$ = 759.8 calculated for C$_{38}$H$_{45}$N$_7$O$_{10}$.

7 Result

7.1 at the Active Site (S965A, H746N)

Two critical active site residues, Ser965 and His746, which were identified from the complex crystal structures of inhibitors with TLCK and TPCK (Figure 16A) and are located at the C-terminal half (Figure 12A), were mutated to alanine and asparagine (S965A, H746N), respectively. Both mutants showed completely abolished enzymatic activity against fluorogenic AMC substrates (Figure 17, Table 5).
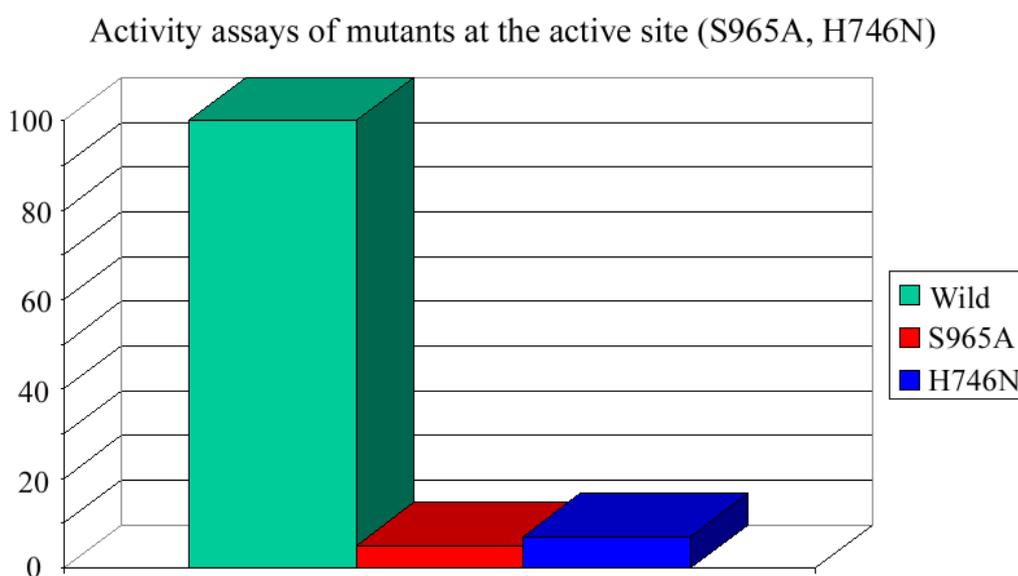

Figure 17. Activity assays of mutants at the active site. Wild-type tricorn protease is a reference for comparing the activities with each engineered point mutation. The enzymatic activity of each mutant was decreased to less than 10 % (no activity), compared with wild-type enzyme.

7.2 at the 7-Bladed β-Propeller Domain (R414C, A643C)

7.2.1 Assay Results

Although the channel through the 7-bladed β-propeller domain is covered by a basic lid at its outside (Arg369, Arg414 and Arg645; Figures 12, 13C), it provides a significantly shorter route from the outside of the protein to the catalytic chamber (60 Å) compared with the alternative route through the central pore (83 Å). The latter path to the active site has multiple

branches (Val600-Gly617, Leu520-Val535) and dead ends (Gly888-Phe895, Lys997-Val1005). To examine the role of the 7-bladed β-propeller domain, a mutant obstructing its central propeller channel was designed by replacing two residues with cysteines, which are located at the entrance to the 7-bladed β-propeller domain, R414C, A643C (Figures 12, 13C). Compared with the wild-type enzyme, this double mutation itself exhibited 70 % and 50 % residual activities towards fluorogenic AMC substrate and insulin B-chain, respectively, prior to any modification. After modification by FMI, the enzymatic activity with fluorogenic substrates was further decreased to 50 %, with insulin B-chain to less than 40 %, resulting in relative residual activities of 35 % and 20 % of the wild-type enzyme towards fluorogenic substrate and insulin B-chain, respectively. Oxidation of both cysteines exhibited similar results, with 30 % activities of the wild-type enzyme, although the formation of disulfide bond was not defined between these two introduced cysteine residues. By contrast, no effect on the wild-type protein was observed by alkylation and a little bit increased enzymatic activities by oxidation (Figure 18, Table 5).
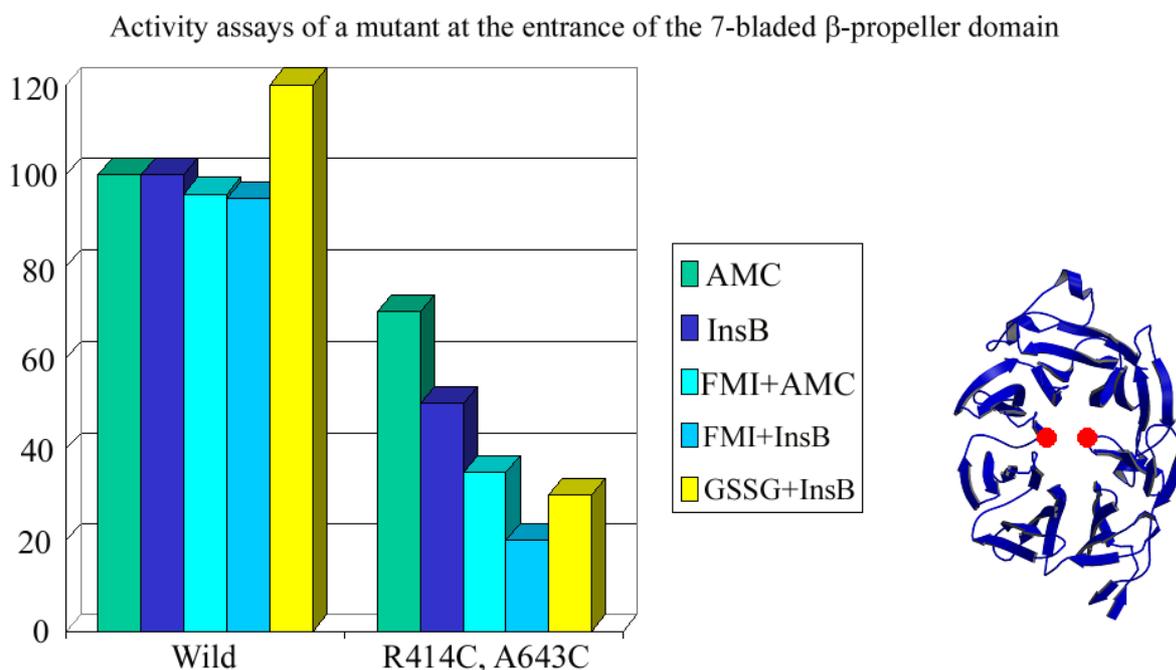

Figure 18. Activity assays of a double cysteine mutant at the entrance of the 7-bladed β-propeller domain (R414C, A643C). Compared to the wild-type enzyme, this double cysteine mutant enzyme exhibited residual activities of 35 % and 20 % towards fluorogenic substrate and insulin B-chain, respectively, after alkylation by FMI. Oxidation with the oxidized glutathione (GSSG) showed similar results, but the formation of disulfide bond was not defined. The red dots on the propeller structure (right) notify the mutated positions.

7.2.2 Crystal Structure of Inhibitor Complex

7.2.2.1 Decanoyl-Arg-Val-Arg-Lys-Chloromethyl Ketone

At the same time, the crystal structures in complex with peptide inhibitors of various lengths, whose C-termini are modified with chloromethyl ketone, were analyzed in order to directly visualize the role of the 7-bladed β-propeller domain in the proteolysis step. In the complex structure of a decanoyl-Arg-Val-Arg-Lys-chloromethyl ketone (RVRKCMK), electron density for the N-terminal deca-hydrocarbon tail was observed. However, it was not long enough to reach the 7-bladed β-propeller tunnel. The chloromethyl ketone head group alkylates the active site residue, Ser965, via an O^γ-methyl covalent bond, as did it in the inhibitor complex crystal structures of TLCK and TPCK. The carbonyl oxygen of P1 is trapped in the oxyanion hole (Figure 19). Two more extra hydrogen-bonds are observed in the Arg-P2 residue with the peptide oxygen of Tyr609 and in the side chain of the Arg-P4 residue with the side chain of Glu605.
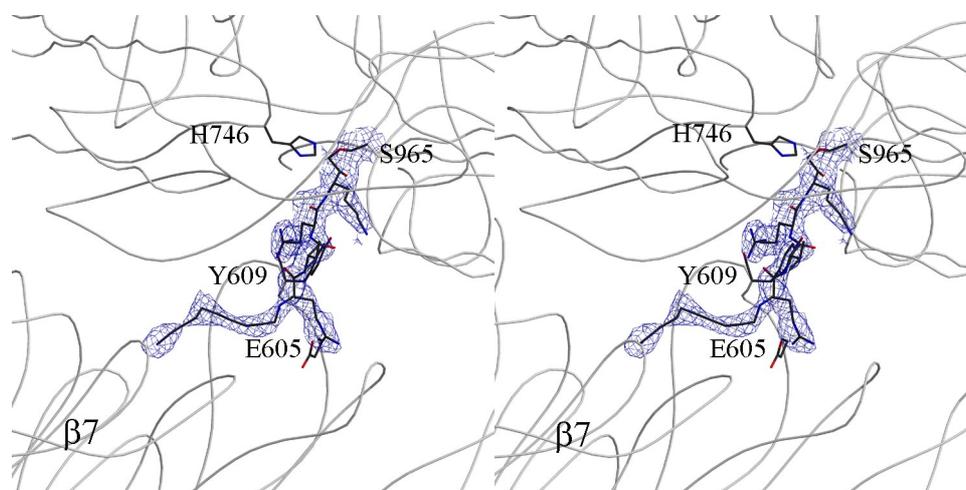

Figure 19. Stereoview of the averaged electron density map around the decanoyl-Arg-Val-Arg-Lys-chloromethyl ketone inhibitor complex at 2.8 Å. Tricorn segments are drawn as gray coil and the inhibitor, Ser965, and His746 as sticks. β7 in middle left marks the 7-bladed β-propeller domain. Like TLCK inhibitor complex crystal structure, the inhibitor is bound directly to Ser965, not to His746. The N-terminal deca-hydrocarbon tail is clearly seen, but it could not reach the 7-bladed β-propeller domain, although it directs to the 7-bladed β-propeller domain.

7.2.2.2 Tridecapeptide-Chloromethyl Ketone

A tridecapeptide-chloromethyl ketone (TridecaCMK), H-Asp-Gln-Thr-Gln-Lys-Gln-Tyr-Gln-Glu-Leu-Thr-Phe-Phe-CH$_2$Cl, was synthesized for the continuous trial to visualize the substrate entry part of the tricorn protease to the sequestered active site. The crystal structure of the tricorn protease in complex with this inhibitor clearly showed electron density for part of the peptide backbone trapped in the lower part of the 7-bladed β-propeller domain, even though the density of the inhibitor segment, Gln1106-Gln1108, was defined only for its main chain and that of its first two residues was not defined (Figure 20). The N-terminal part of this inhibitor is located in the funnel-shaped end of the 7-bladed β-propeller tunnel and forms a hairpin structure. The carbonyl oxygen of P1 is trapped in the oxyanion hole. This inhibitor is also bound directly to the catalytic Ser965 like other chloromethyl ketone-based inhibitors. Its residues from P1 to P4 are particularly well defined as shown by clear density in Figure 20.
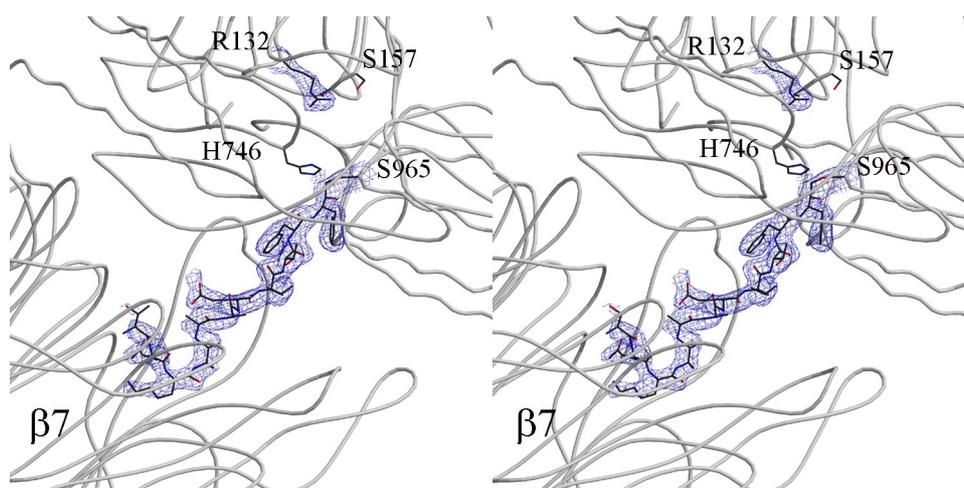

Figure 20. Stereoview of the averaged electron density map around the TridecaCMK inhibitor complex at 2.6 Å. Tricorn segments are drawn as gray coil and the inhibitor, Ser965, and His746 as sticks. β7 in middle left marks the 7-bladed β-propeller domain. The inhibitor is covalently bound to Ser965. Towards the N-terminus of the inhibitor, the density does not allow positioning of side chains, but the mainchain can be unequivocally traced. Undefined side chains have been truncated to alanines. The inhibitor conformation is extended from Glu1109 to Phe1113 and forms a hairpin-like structure from Gln1104 to Tyr1107.

7.3 at the Putative Substrate Binding Site (R131E-R132E)

7.3.1 Assay Results

A peptide loop across the inner funnel of the 6-bladed β-propeller domain would be the anchoring site for the substrate C-terminus on the bases of such as assay results that substrates with a negative charged N-terminal are not cleaved (Tamura *et al.*, 1996a, 1998), and structural features: (i) the substrate might be pre-oriented for the processive proteolysis from the beginning using several arginines (Arg369, Arg414 and Arg645; Figures 12, 13C) of the entrance part at the 7-bladed β-propeller domain; (ii) a short strand, Gly993-Thr995, was served as an unprimed substrate-docking strand using its main-chain atoms from the
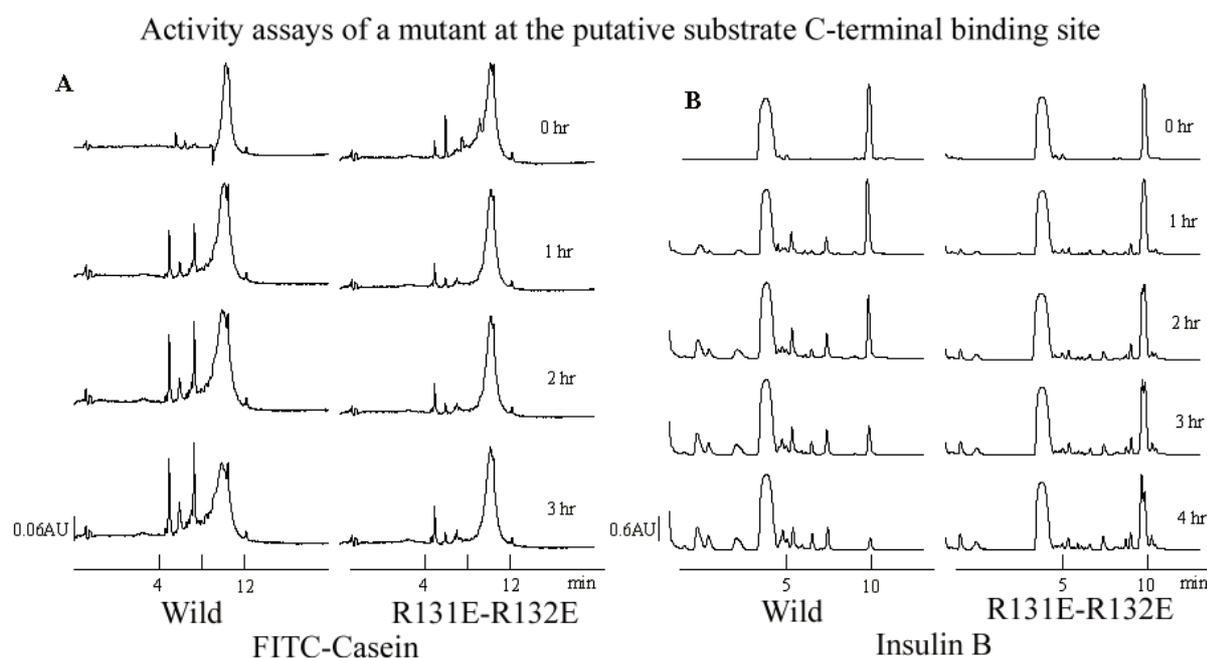

Figure 21. Protein substrate digestions with the wild-type and a charge reversal mutant at the putative substrate C-terminal docking site (R131E-R132E). Time dependent degradation results show that charge reversal mutant at the putative substrate C-terminus anchoring site has almost less than 10 % activities of wild-types with FITC-labeled casein (a), and less than 30 % against oxidized insulin B-chain (b). Furthermore, the cleavage patterns of charge reversal mutant are different from those of the wild-type, but remain time-invariant, indicating processivity. The high peaks in a) around 11 are from the undegraded FITC-labeled caseins. The uneven base lines of a) at 0 hr seemed to be derived from the insufficient column equilibration at the beginning. The high peaks in b) around 5 and 10 min are from the added AAPVCMK inhibitor to stop the further cleavage reaction after specific times and undegraded oxidized insulin B-chain, respectively.

inhibitor crystal structure (Figures 19, 20); (iii) another strand, Gly916-Gly918, can be extended as a primed substrate-docking strand; (iv) the topology and size of the inner cavities favor an extended conformation of the substrate and the C-terminus of the substrate will be attracted to basic residue(s), thereby presenting the scissile bond of substrate at the Ser965 for proteolysis. This peptide loop carries Arg131-Arg132 (Figures 12, 16B right). When a double charge reversal mutant was introduced to these residues (R131E-R132E), it exhibited a little bit increased activities towards fluorogenic substrates, compared with the wild-type enzyme (Figure 22). However, it showed remarkably lowered enzymatic activity for protein substrates, about 10 % for FITC-labeled casein compared with the wild-type enzyme, and about 30 %, for oxidized insulin B-chain (Figures 21, 22 and Table 5). This mutant enzyme degraded protein substrates in a processive manner (Figure 21), in the same way that has been described for the wild-type enzyme towards insulin B-chain (Tamura *et al.*, 1998). The cleavage patterns, however, are clearly different (Figure 21B).
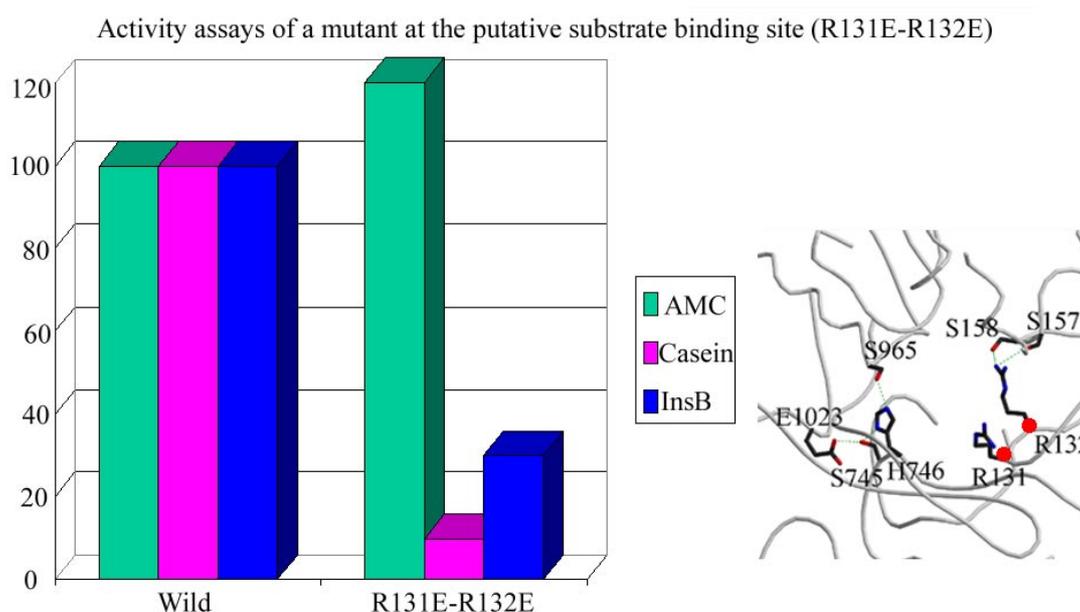

Figure 22. Summary of activity assays of mutant at the putative substrate binding site (R131E-R132E). The enzymatic activities of this double charge reversal mutant exhibited less than 10 % and 30 % towards the FITC-labeled casein and oxidized insulin B-chain, respectively. The red dots on the right figure notify the mutation points.

7.3.2 Assays with Substrates of a Free C-Terminal and of an Amidated C-Terminal

To confirm the role of Arg131-Arg132 on the substrate C-terminal anchoring, the degradation experiments of the wild-type enzyme were performed towards substrates of different C-terminal characteristics, a substrate with a free C-terminal and a substrate with an amidated one (secretin-OH and secretin-NH$_2$). Unexpectedly, the tricorn protease processed the amidated secretin at the C-terminus as well as the free carboxy secretin (Figure 23).

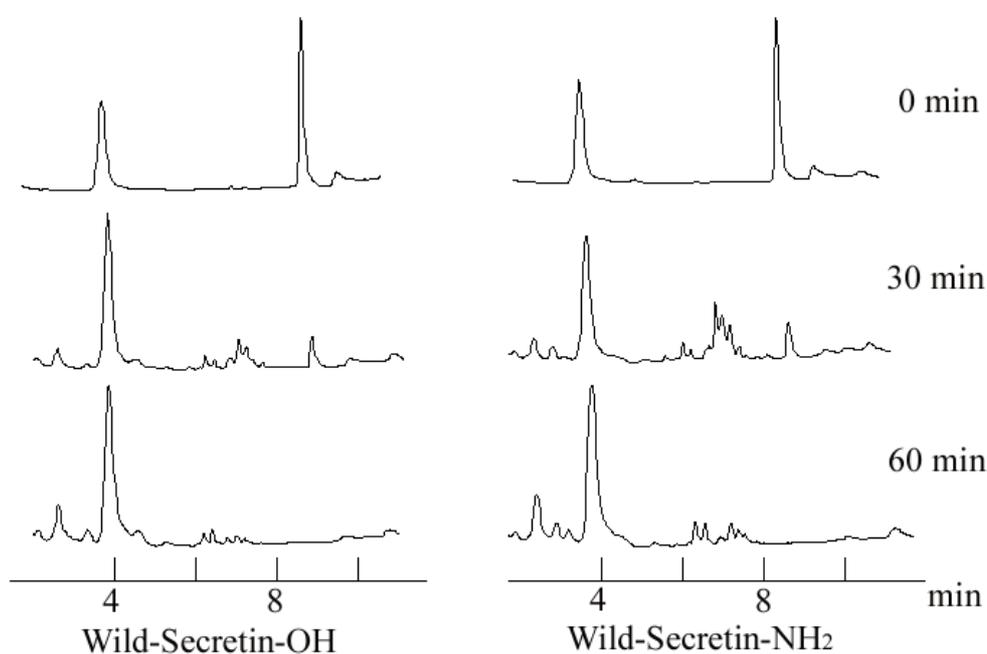
Figure 23. Degradation experiments of substrates with a free C-terminal and an amidated C-terminal. The wild-type tricorn protease also degrades substrates with an amidated C-terminal as well as substrates with a free C-terminal. The high peaks at 4 min and around 9 min are from the added AAPVCMK inhibitor to stop the further reaction after specific times and undegraded substrates, respectively.

7.3.3 Crystal Structure of the Z-PheΨ[CO-CONH]Arg-Glu-Phe-OH

In order to visualize the substrate binding mode at the primed site and explain the above two contradictory results (A charge reversal mutant at the putative substrate C-terminal binding site showed critically lessened enzymatic activities and the substrate without a free C-terminus was also degraded), one inhibitor, benzyloxycarbonyl (Z)-PheΨ[CO-CONH]Arg-Glu-Phe-OH (ZDK), with a diketo group at the cleavage site, was synthesized and the crystal structure of the tricorn protease in complex with this inhibitor was elucidated. The inhibitor complex crystal structure showed that Arg131-Arg132 are hydrogen-bonded to the carbonyl oxygen of the P2' and P3' peptide. While they displayed no hydrogen bonds to the free C-terminus, the electrostatic charges are partially balanced (Figure 24).
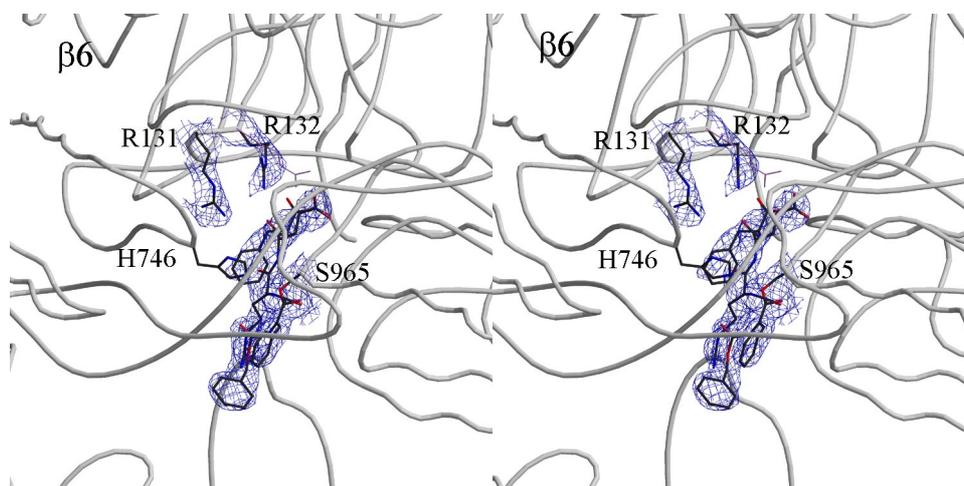

Figure 24. Stereoview of the averaged electron density map of the bound inhibitor with a diketo group at the cleavage site between the active site and the Arg131-Arg132 anchoring segment at 2.7 Å. Side chains of key residues around the active cleft (Ser965 and His746), the anchor segment (Arg131-Arg132), and the inhibitor were drawn by sticks. The side chain of Arg132 seen in the free enzyme and its hydrogen bond partners (Ser157-Ser158) were drawn in thin pink sticks. β6 in the right top stands for the 6-bladed β-propeller domain. The inhibitor is bound via the first carbonyl carbon of the diketo group to the O^γ of Ser965 as a tetrahedral adduct.

7.4 at the 6-Bladed β-Propeller Domain (L184C)

The comparison with prolyl oligopeptidase from bovine porcine, including the open Velcro topology (Fülöp *et al.*, 1998), suggests an important role of the 6-bladed β-propeller domain for the product exit from the active site. The role of the 6-bladed β-propeller domain was examined by a mutation of leucine to cysteine at 184. This residue is located midway in the tunnel of the 6-bladed β-propeller domain (Figure 25, 28). The mutant showed a doubling of peptidolytic and proteolytic activities. After modification with NEM, reduced activities against the wild-type enzyme were observed towards both fluorogenic substrates and insulin B-chain, respectively (Figure 25, Table 5), whereas no effect on the wild-type enzyme was observed.
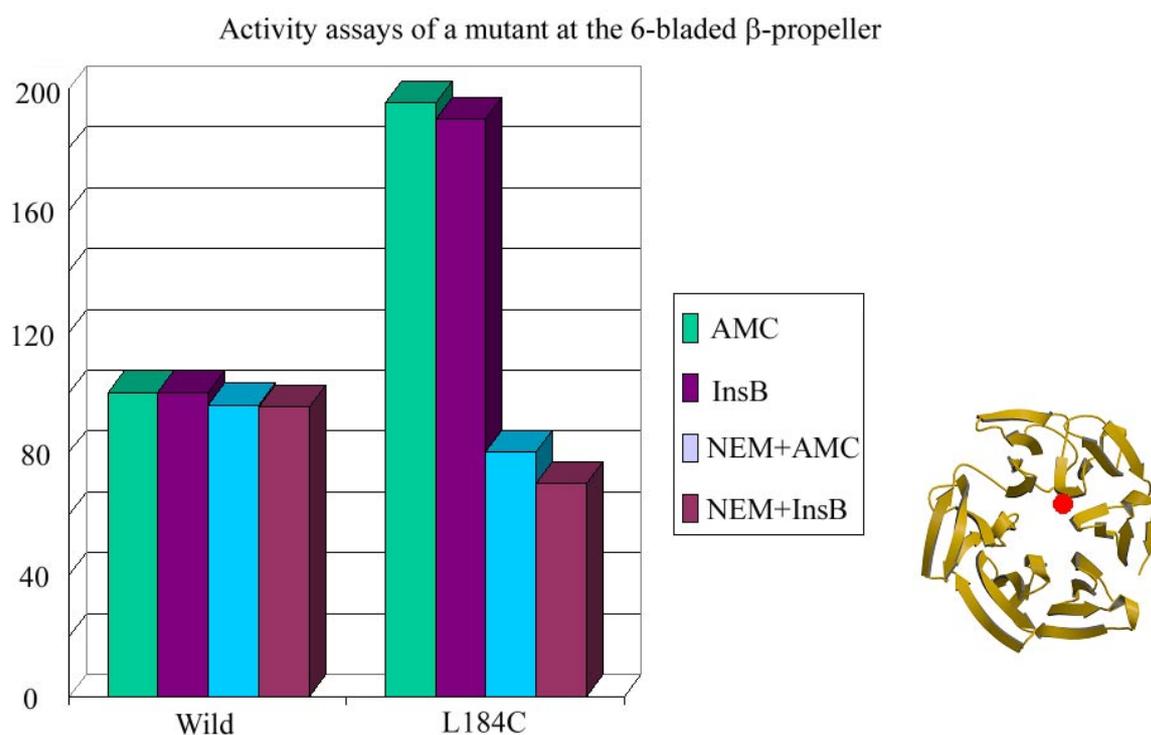

Figure 25. Summary of activity assays of a mutant at the 6-bladed β-propeller domain (L184C). This mutant exhibited a doubling of enzymatic activities, which was reduced to the normal value after alkylation. The red dot on the propeller structure (right) identifies the mutation point of L184C.

7.5 Two β-Strands Sandwich the Substrate at the Active Site

In order to elucidate the substrate-binding mode of the tricorn protease in more detail, several crystal structures in complex with inhibitors of different lengths and residue characteristics were examined. The tripeptide, D-Phe-Pro-Arg-chloromethyl ketone (PPACK), which is a specific inhibitor of thrombin, an RVRKCMK, a TridecaCMK, and the ZDK with a diketo group at the cleavage site clearly showed that two strands (Gly993-Thr995 and Gly916-Gly918) define both the unprimed and primed substrate recognition site at the active cleft. Upon binding, the substrate inserts between these strands to form a local anti-parallel β-sheet by providing hydrogen donors and acceptors to the protein matrix. The interactions are mainly being accomplished between the backbone atoms and not the side chains of the protease or the substrate (Figures 28, 29). Like the complex crystal structures of TLCK and TPCK, the inhibitors are bound directly to Ser965, not His746 (Figures 19, 20, 24). The carbonyl oxygen of the conserved Gly918 serves as a hydrogen bond acceptor of the backbone amide of the P1 residue. Ile994 is on the opposite side and acts as a hydrogen bond donor with its backbone amide, to the carbonyl oxygen of P2 and as an acceptor with its carbonyl oxygen, to the peptide amide of the substrate P2 residue (Figure 29).

The P1 residue of PPACK, Arg-P1, maintains the same conformation as the TLCK inhibitor in the S1 pocket. Pro-P2 and Phe-P3 are positioned inside neutral pockets while the carbonyl oxygen is trapped in the oxyanion hole (data not shown).

The P1 residue of RVRKCMK, Lys-P1, is interacting with Asp936 in the S1 pocket like a lysine residue in the TLCK inhibitor complex crystal structure. The Arg-P2 is interacting with the peptide oxygen of Tyr609 while its Val-P3 is trapped in the neutral S2 pocket of Tyr609, Phe919 and Phe1011. And, the Arg-P4 residue interacts with the side chain of Glu605 (Figure 19).

The TridecaCMK is bound to the active site keeping the same interactions between protein and ligand residues as the other chloromethyl ketone-based inhibitors. In particular, its P1, P2 and P3 residues (Phe-P1, Phe-P2 and Thr-P3) fit well into the corresponding S1, S2 and S3 pocket, as shown by the clear electron density in Figure 20.

At the primed sites observed in the complex structure of a ZDK inhibitor, only two hydrogen bonds contribute besides the interactions with Arg131-Arg132, (described in 7.3.3). The peptide nitrogen of the P2' is forming a hydrogen bond with the carbonyl oxygen of Gly916. Also, the side chain of the P1' arginine is fixed by the peptide oxygen of Tyr609 (Figure 29).

Table 4. Data collection and refinement statistics.

	RVRKCMK	TridecaCMK	ZDK
Radiation source	DESY	SLS	DESY
Wavelength, Å	1.05	0.9794	1.05
Space group	$P2_1$	$P2_1$	$P2_1$
Resolution, Å	30-2.8	30-2.6	20-2.7
Cell parameters			
a, Å	95.44	95.47	95.67
b, Å	245.43	245.10	244.55
c, Å	159.40	157.89	158.32
β, °	104.79	105.19	104.59
Unique/total reflections	172151/372512	209855/645775	178808/295654
Redundancy	2.16	3.07	1.65
Completeness, %	88.9(87.0*)	97.7(85.1*)	93.2(83.6*)
Rsym, %	9.0(39.3*)	8.4(30.5*)	6.6(26.3*)
R_{factor}, %	28.4	25.4	24.7
R_{free}[1], %	31.5	28.8	28.0
Rmsd bond lengths, Å	0.009	0.014	0.010
Rmsd bond angles, °	1.25	1.49	1.27
Rmsd NCS protein, Å	0.034	0.026	0.012

SLS, Swiss Light Source

DESY, Deutsche Elektronische Synchrotron

*: Values for the reflections in the highest-resolution shell (2.83-2.80, 2.64-2.60, 2.75-2.70 Å, respectively).

[1] R_{free} is calculated over a randomly selected fraction of reflection data not included in the refinement (8.8, 8.6, 8.5 %, respectively).

Table 5. Summary of activity assays of wild-type and mutants.

		Wild-type	active site		7-propeller		anchor	6-propeller
			H746N	S965A	R414C	R414C, A643C	R131-132E	L184C
Relative activities against wild type	AMC	****	-	-	NM	***	****	*****
	Insulin B	*****	NM	NM	NM	**	*	*****
	Casein	*****	NM	NM	NM	NM	-	NM
Relative activities after oxidation, AMC		****	NM	NM	NM	*	NM	NM
Relative activities after alkylation	AMC	****	NM	NM	**[1]	*	NM	***
	Insulin B[2]	****	NM	NM	NM	*	NM	***

*****: > 150 %, ****: 80 –100 %, ***: 55-80 %, **: 35-55 %, *: 15-35 %, -: < 15 %, NM: not measured

[1]: against non-treated R414C.

[2]: final results of alkylation against non-treated wild-type.

AMC substrate: Bz-Val-Gly-Arg-AMC, final 10 mM of 100 µl reaction volume.

Alkylating agent, NEM or FMI, overnight, 4 °C.

Oxidizing agent, 1 mM oxidized glutathione, overnight, 4 °C.

Fluorogenic AMC substrate was purchased from Bachem, FMI from Molecular Probes and FITC-labeled casein, NEM, oxidized insulin B-chain and oxidized glutathione from Sigma-Aldrich.

8 Discussion

8.1 Enzymatic Mechanism

β-propeller structures are generally rigidified by clamping their terminal blades either through the exchange of β-strands between them, or by a disulfide bond to covalently connect them only in the case of 4-bladed β-propellers (closed Velcro, Faber *et al.*, 1995; Neer *et al.*,
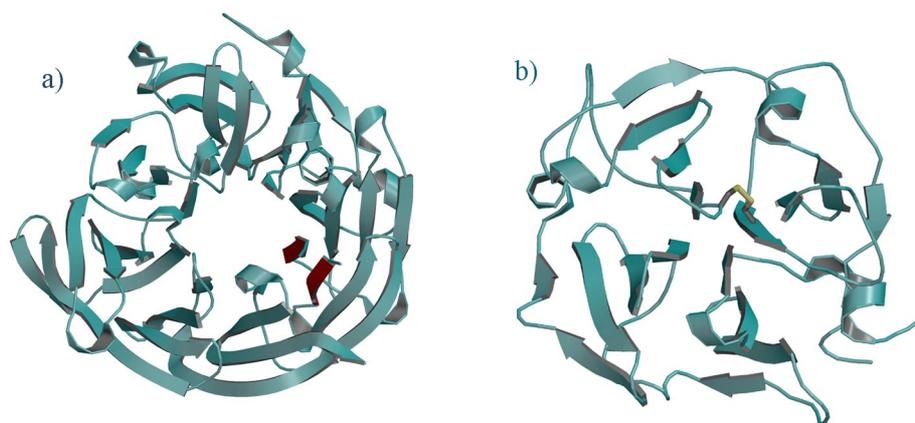

Figure 26. Structures of closed Velcro β-propeller structures. Velcro β-propeller structures are rigidified by clamping their terminal blades (a, 1qks) or forming a disulfide bond between neighboring blades (b, 1hxn). While strands for closing the Velcro were colored by margenta (a), a disulfide bond to connect blades was drawn as yellow sticks (b).

1996; Baker *et al.*, 1997; Figure 26). In contrast, prolyl oligopeptidase from bovine porcine has an open Velcro 7-bladed β-propeller domain with regular sequential strand arrangement and without a disulfide clamp. The lack of strand exchange between the terminal blades has been suggested to allow structural flexibility (open Velcro, Fülöp *et al.*, 1998; Figure 27).

Similarly, both the 6-bladed β-propeller structure and the 7-bladed β-propeller structure of the tricorn protease are also unconstrained (Figure 13). Their central pores provide the shortest path of substrate to, and product from, the active site (Figure 11c right). Arg414 and Ala643 are located at the pore entry to the 7-bladed β-propeller domain (Figure 28) and showed reductions of activity towards both fluorogenic and protein substrates, when mutated to cysteines. After alkylation or oxidation of both cysteines, the activity was strongly reduced. The introduction of bulky side groups into the lumen of the propeller pore is likely

to hinder substrate passage, as does disulfide bond formation. The latter modification might also reduce a breathing motion of the propeller blades by formation of the plausible disulfide
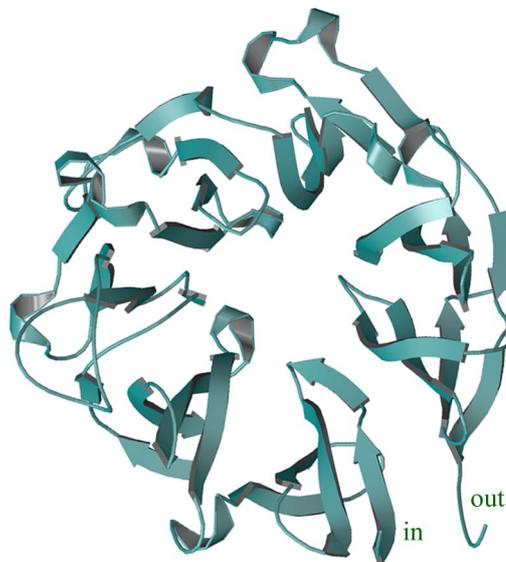

Figure 27. Open Velcro 7-bladed β-propeller structure part of prolyl oligopeptidase (1qfm). This β-propeller structure lacks the disulfide clamp and strand exchange between blades and substrate passes through the propeller structure to the catalytic site (Fülöp *et al.*, 1998).

bond between the two introduced cysteine residues or the mixtures of partial modifications like the alkylation to each introduced cysteine residue, even though the disulfide bond formation was not identified. As the channel diameter measured between the Cα carbons of Arg414 and Ala643 is only 6.4 Å in the free enzyme, transient widening would facilitate substrate passage. It is interesting to note that the entry part of the blade containing Ala643 (blade 7, Figure 13B) appears more flexible, as indicated by relatively high crystallographic B-values in the free enzyme (above 42 Å^2, compared to 30 Å^2 of the other blades). Complete inhibition of enzyme activity had been observed in prolyl oligopeptidase when a disulfide bond was introduced between the two terminal blades 1 and 7 of the open Velcro propeller structures (Fülöp *et al.*, 2000). The residual activity of the modified tricorn protease may be in part due to an alternative path through the central pore along the molecular 3-fold axis, which is much longer and tortuous compared to the direct pathway through the 7-bladed β-propeller domain. The notion that the pore of the 7-bladed β-propeller domain is the main substrate route to the active site is strongly supported by the structure of the TridecaCMK complex as it directly showed the peptide in the lower part of the lumen of the propeller pore (Figure 20).

The chloromethyl ketone-based inhibitors and the ZDK inhibitor are fixed by formation of a local anti-parallel β-sheet with the two strands (Gly993-Thr995 and Gly916-Gly918) at the unprimed and the primed site, respectively (Figures 28, 29, 31). The main interactions are accomplished between backbone atoms. Three hydrogen bonds are observed at the unprimed site, one at the primed site and two at the oxyanion hole. Two more hydrogen
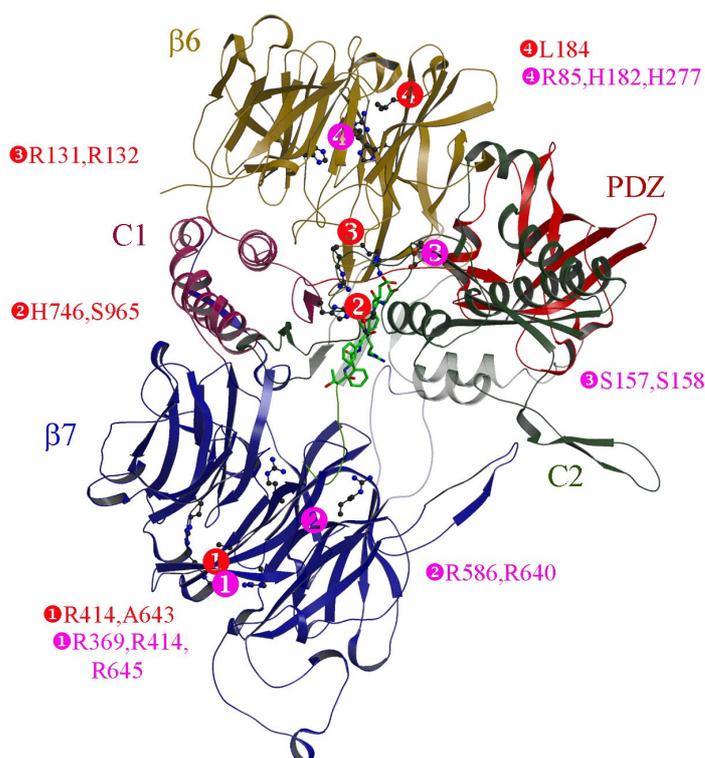

Figure 28. Substrate binding mode derived from two experimental data. The tricorn monomer with a model of a physiological substrate at the active site. A tricorn monomer is drawn in a ribbon representation. Each domain is displayed by different colors. The modeled substrate is highlighted by thick sticks, whereas its N-terminal tail that lacks defined interactions with the protease is drawn as a thin green coil in the 7-bladed β-propeller (β7). A helix (Phe606-Tyr610) was simplified as a blue coil. The substrate strand and strands for forming a local anti-parallel β-sheet with substrate (Gly993-Thr995 and Gly916-Gly918) were displayed transparently, to show their residues. Residues, which are participating in substrate recognition (Arg369, Arg414 and Arg645) at the entrance part of, and transfer (Arg368 and Arg640) within, the β7, were drawn by thin ball-and-stick models as well as residues involved in product release (Arg131, Arg132, Ser157, Ser158, His277, His182 and Arg85) through the 6-bladed β-propeller (β6). The substrate modeled on the basis of two inhibitor complexes is bound to the O$^\gamma$ of Ser965 and was represented as a tetrahedral adduct. The unprimed residues were derived from the TridecaCMK inhibitor complex crystal structure and the primed residues from the ZDK inhibitor complex, respectively.

bonds are provided by the Arg131 and Arg132 side chains to the peptide carbonyl oxygen of substrate to fix the last peptide group (Figure 24). Arg132 reorients away from Ser157 and Ser158 with which it forms hydrogen bonds in the inhibitor-free and the chloromethyl ketone-based inhibited enzyme (Figure 24). The substrate-enzyme interaction is dominated by mainchain interactions rather than sidechains, which explains the broad substrate specificities of the tricorn protease (Tamura *et al.*, 1998; Table 1).

The backbone interaction to the P1 residue provided by Gly918 resembles the anti-parallel S1-P1 strand recognition in the tricorn protease associated F1 protease by Gly37, which is strictly conserved in F1 homologous proteins (Goettig *et al.*, in preparation). The presence of the charged Arg131-Arg132 anchor site further reflects the structural analogy of these functionally interacting proteases. In F1, the two negatively charged residues (Glu213 and Glu245) serve to recognize the N-terminus of substrate, analogous to the recognition of the C-terminus of substrate by Arg131-Arg132 in the tricorn protease.

While the complex crystal structures of chloromethyl ketone-based inhibitors confirmed the overall backbone interaction pattern at the unprimed site, the side chain of the P2 arginine in the complex crystal structure of the RVRKCMK formed an additional hydrogen bond with the peptide oxygen of Tyr609 (Figure 19), which was defined as the S1' site in the complex crystal structure of the ZDK inhibitor (Fgiure 24). The crystal structure of TPCK showed that the geometry and charge of the S1 site is able to adapt via Asp936 of a neighboring monomer to the P1 residue. The complex crystal structure of a ZDK inhibitor indicates similar adaptation of the Arg131-Arg132 anchor to the primed substrate residues (Figure 24). The observed geometry and interaction explains how tri- and even tetrapeptide can bind at the primed side. In fact, it appears even better suited for dipeptides which are expected to directly bind with its C-terminus to the Arg131-Arg132 anchor.

The catalytic nucleophile, Ser965 activated by His746, attacks the peptide carbonyl carbon of substrate P1 residues. The exact cleavage positions are determined by the oxyanion hole of Gly918 and Asp966 and the S1 pocket (Figure 29). A tetrahedral adduct is formed and the peptide bond cleaved by subsequent acyl enzyme formation and hydrolysis (Figure 30).
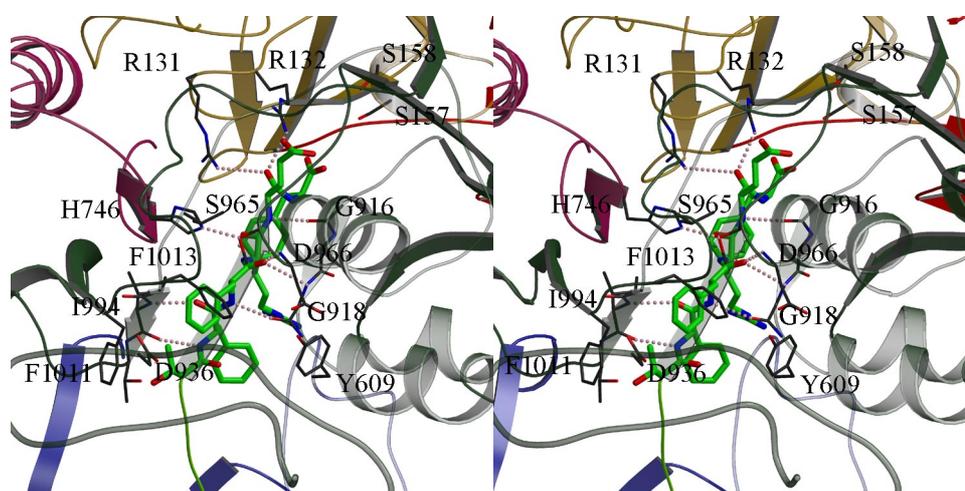
Figure 29. Stereo representation of a close-up view at the active site with the same colors as Figure 28. The arm of a neighboring monomer (Asn930-Asn949), which contributes Asp936 to the S1 pocket, is shown as a thick and dark grey coil. The substrate strand and strands for forming a local anti-parallel β-sheet with substrate (Gly993-Thr995 and Gly916-Gly918) were displayed transparently, to show their residues. Relevant hydrogen bonds between protein and substrate residues are indicated as dotted pink lines.

Once substrates are cleaved at the active site, products must dissociate from the active site for the next round of proteolysis or for the recovery of free enzyme state. A direct and short pathway of the di- or tripeptide product to the surface is offered by the central pore of the 6-bladed β-propeller domain. The Arg131-Arg132 anchoring segment spans across its inner entry (Figure 28). By slight rearrangements, a path is opened for the diffusion of the peptide product, which may be facilitated by His277, His182 and Arg85 lining the pore wall. The result of the mutation L184C is in accord with this model. Leu184 is located in the middle of the 6-bladed β-propeller domain (Figure 28). The smaller cysteine side group may facilitate product diffusion leading to strongly increased enzyme activities with protein substrates as well as fluorogenic substrates. Alkylation of the L184C mutant enzyme restricts the channel through the 6-bladed β-propeller domain, thus leading to reduced enzyme activities, slightly less than the wild-type (Figure 25, Table 5). Dissociation of the di- or tripeptide product generates the room necessary for the unprimed product to move forward by two or three residues towards the Arg131-Arg132 anchor.

Substrate preferences are weak as already indicated by the S1 site adaptable for hydrophobic and basic P1 residues. S1' seems to prefer basic residues which are long enough to interact with the peptide oxygen of Tyr609 as shown with the complex crystal structure of the ZDK inhibitor (Figure 24). Alternatively, this basic residue may be provided by the P2 position of the substrate, as seen with the crystal structure in complex with RVRKCMK. The tricorn protease also favors a basic residue at P4 site, as shown by the RVRKCMK crystal structure (Figure 19). The observed structural features explain the processive cleavage of the tricorn protease substrates from the carboxy- to the amino-terminus and classify the tricorn protease as a di- (tri-) peptidyl carboxypeptidase.

The processive cleavage was indicated by an invariant product profile of insulin B-chain as shown by the time course HPLC elution pattern (Tamura *et al.*, 1998; Figure 21B). Analysis of these degradation products further revealed the exopeptidase activity of the tricorn protease, with a clear preference for di- and tripeptidyl cleavage (Tamura et al., 1998). Similarly, the charge reversal mutant at the substrate C-terminal anchoring site, R131E-R132E, shows a time-independent product profile, which, however, differs from the wild-type products (Figure 21B). This observation assigns a marked role of the primed side in the alignment of the substrate with the active site.

Several independent lines of evidence prove the unidirectionality of the substrate-to-product flow. A first simple, yet fundamental argument is that the wider 7-bladed β-propeller domain is better suited to guide the larger substrate than the 6-bladed β-propeller domain, which reflects the size distribution of substrates (~10 amino acids) and products (2-3 amino acids). An opposite substrate flow where substrate enters the active site through the 6-bladed β-propeller domain is very unlikely already based on these size considerations. Second, in the case of high substrate excess, the catalytic turnover is expected to be limited by the product exit rather than the substrate entrance. This argument is strengthened by the processivity of the tricorn protease. In fact, only the widening of the 6-bladed β-propeller channel increased the catalytic turnover rate of the tricorn protease (Table 5). Third, the filtering mechanism of the 7-bladed β-propeller and the 6-bladed β-propeller channels is likely to distinguish substrates upon the size and the conformation, whereas it is impossible to distinguish di- or tripeptidic products upon these criteria. The mutation in the 7-bladed β-propeller domain is sensitive towards both protein and fluorogenic substrate, contrasting the mutation in the 6-bladed β-propeller domain (Table 5). A fourth, direct argument to support the role of 7-bladed β-propeller and 6-bladed β-propeller domain in substrate access and product egress is provided by the crystal structure of the tricorn protease in complex with a trideca-peptide substrate analogue, which clearly defines the substrate binding to the active site (Figure 20). The only exit route for the cleaved di- or tripeptidic product is through the 6-bladed β-propeller domain. As an alternative exit route, the 7-bladed β-propeller domain or the central pore would displaces the unprimed product, conflicting the processivity of the tricorn protease which implies the unprimed substrate to be contained at the active site until it is completely degraded.
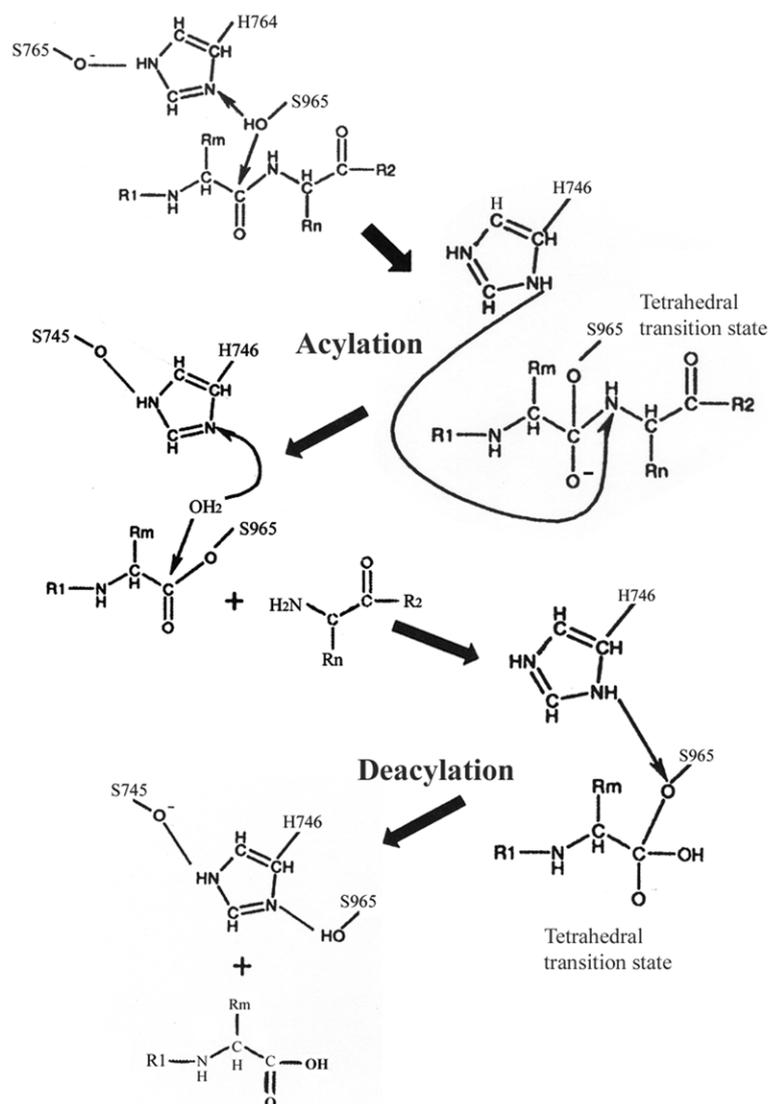
Figure 30. Schematic drawing of the hydrolysis. Ser965 activated by His746 via the hydrogen-bond networks of Ser745 and Glu1023, attacks the carbonyl carbon of the P1 residue. A tetrahedral adduct is formed and the peptide bond is cleaved. Ser965 restores its initial stage for the next round of cleavage reaction during the deacylation step. Di- or tripeptide further moves to the Arg131-Arg132 anchor.
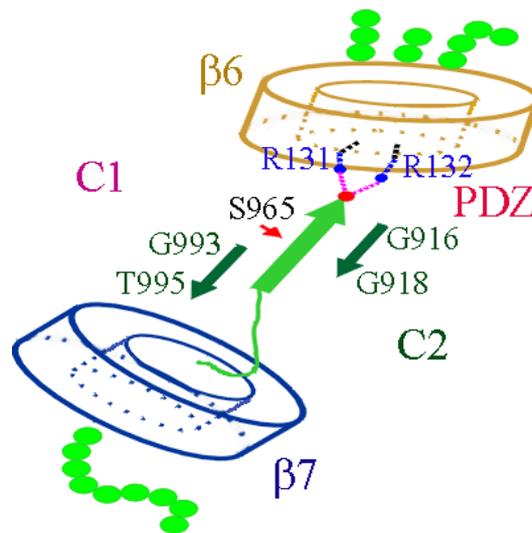
Figure 31. A schematic diagram of the tricorn protease during the proteolysis step. The substrates reach the active site through the 7-bladed β-propeller domain (β7), and are sandwiched by two short strands to form a local anti-parallel β-sheet with its C-terminal anchored to the Arg131-Arg132. The processed products diffuse away through the 6-bladed β-propeller domain (β6). Two cylinders, which drawn as similar colors as Figure 28, represent 7- and 6-bladed β-propeller domain, respectively. Substrate to the β7 and products from the β6 are displayed by continuous green dots in which a single green dot means a single amino acid. The substrate bound to the active site is indicated by a green strand leaving its N-tail coil in the channel of the β7. The red dot on the end of substrate strand stands for the substrate C-terminal. Ser965, the key residue for the substrate cleavage, is indicated as a red arrow at the cleavage site. Hydrogen bonds between the substrate C-terminal and Arg131-Arg132 was shown as dotted pink lines. Two β-strands, which sandwich the substrate, are displayed with the same color as C2. The rest parts of a tricorn monomer were symbolized as C1, PDZ and C2, respectively.

8.2 Prospects

8.2.1 Interaction with the Tricorn Protease Interacting Factors

The tricorn protease reportedly cooperates with three additional proteins, termed interacting factors F1, F2 and F3, to degrade oligopeptides sequentially to yield free amino acids (Tamura *et al.*, 1996a, b, 1998). The tricorn protease associated F1 is a prolyl iminopeptidase with 14 % sequence identity to the catalytic domain of prolyl oligopeptidase from bovine porcine, which has an additional propeller domain (Fülöp *et al.*, 1998). Guided by the structural scaffold of this latter structure, it was speculated that F1 docks onto the 6-bladed β-propeller structure of the tricorn protease pore protein. As in bovine prolyl oligopeptidase, substrate (product of the tricorn protease) enters F1 through the propeller channel in this

model. Previously, a physical interaction of F1 with the tricorn protease was reported (Tamura *et al.*, 1998). While F1 and the tricorn protease do not comigrate on a gel filtration chromatography under low ionic strength, preliminary data from dynamic light scattering do suggest a low affinity of three F1 proteases with one hexamer of the tricorn protease (Goettig *et al.*, in preparation), consistent with the EM micrograph of the capsid assembly which the contacts between the hexamers are mediated by three of the 6-bladed β-propeller domains (Figure 32).
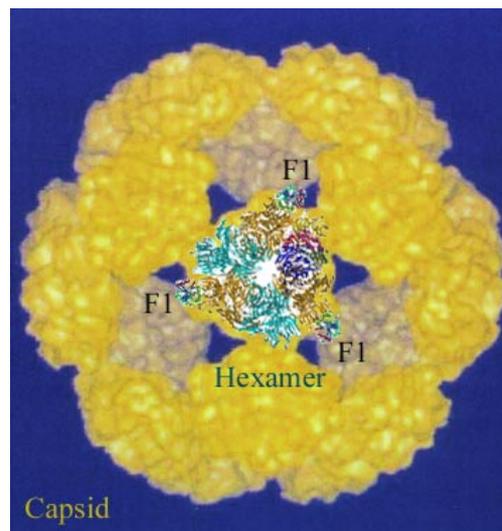
Figure 32. Schematic illustration which translates the functional interactions of the tricorn protease with its interacting factor 1 (F1). This model is hypothetical, putting the functional related proteins (the tricorn protease and F1 from *Thermoplasma acidophilum*) on a common scale.

8.2.2 Interaction with Proteasome

Similarly, there is an evidence for functional interaction of the tricorn protease with the proteasome (Tamura *et al.*, 1998). If this functional interaction is paralleled by a physical interaction, the exit of product from the proteasome should be aligned with the substrate entrance into the tricorn protease. This proposed interaction has to be weak, presumably transient (for example, substrate induced), or an auto-inhibitory proteasome-tricorn protease tunnel would result. While such a physical interaction would be consistent with the geometric dimensions of both molecules (Groll *et al.*, 1997; Figure 33), its existence needs to be experimentally confirmed and characterized.
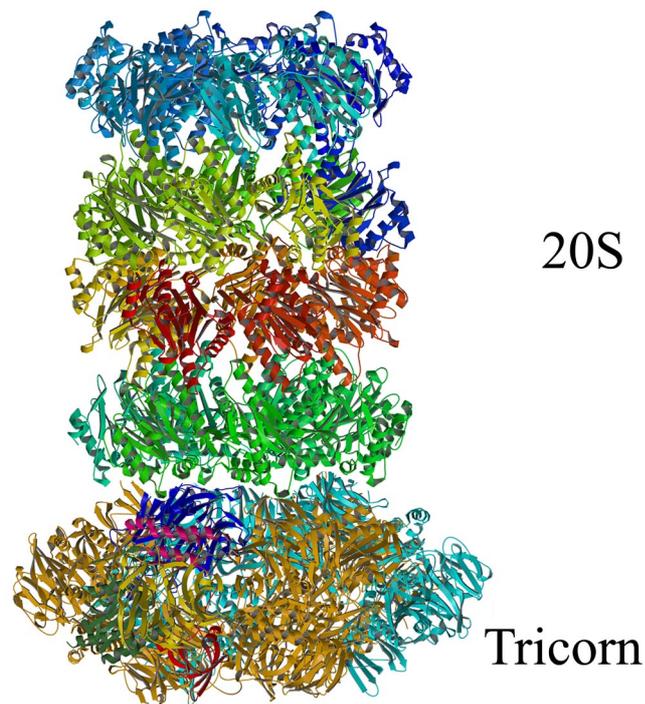
Figure 33. Schematic illustration which translates the functional interactions of the tricorn protease with the proteasome 20S from *Thermoplasma acidophilum* (Löwe *et al.*, 1995). This model is hypothetical, putting the functional related proteins (the tricorn protease and the proteasome 20S) on a common scale.

8.2.3 Eukaryotic Homologues

There is a circumstantial evidence for the presence of selected elements of the modular tricorn protease in humans, including their co-factor proteins (Stoltze *et al.*, 2000). Functional tricorn protease analogues might be difficult to detect in eukaryotes, however. Different from the proteasome, the tricorn protease is built not from a single folding domain, but from five, raising the possibility that the tricorn protease analogues might assemble non-covalently from different gene products, thereby possibly having additional functionalities. Because functional roles for the β-propellers was elucidated in substrate access and product egress, these modules are likely candidates to mediate the interaction with other factors, and might as well associate with these proteins as with the tricorn proteolytic core sub-domains. In fact, the eukaryotic D1P of photosystem II is the proof of the existence of the remaining isolated sub-domains (C1, PDZ and C2) (Liao *et al.*, 2000; Figure 14). This search for the presence of the

tricorn protease analogues in mammals could be continued, perhaps also assisted by selective inhibitors designed on the basis of the biochemical and structural principles presented here.

9 APPENDIX

9.1 Abbreviations

Å	Ångstrøm; 1 Å = 10^{-10} m
AMC	-7-amino-4-methylcoumarin
ATP	adenosine triphosphate
bp	basepair
BSA	bovine serum albumin
Bz	benzoyl
CMK	chloromethyl ketone
Da	dalton; 1 Da = $1g \cdot mol^{-1}$
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
ESI-MS	electron spray ionization mass spectroscopy
FITC	fluorescein isothiocyanate
FMI	fluorescein-5-maleimide
Hepes	N-2-[hydroxyethyl]piperazine-N'[-2-ethanesulfonic acid]
HPLC	high performance liquid chromatography
IPTG	isopropyl β-D-thiogalactoside
LB	Lauria-Bertani medium
MAD	multi-wavelength anomalous dispersion
MES	2-morpholinoethansulfonic acid
MME	monomethyl ether
MOPS	3-[N-morpholino]propanesulfonic acid
MPD	2-methyl-2,4-pentane diol
NEM	N-ethylmaleimide
NMM	new minimal media
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
PMSF	phenylmethylsulfonyl fluoride
rmsd	root-mean-square-deviation
rpm	revolution per minute
SDS	sodium dodecyl sulfate
Se-Met	seleno-DL-methionine
SIRAS	single isomorphous replacement with anomalous scattering
Suc	succinic
TFA	trifluoroacetic acid

Tris		α,α,α-trishydroxymethyl aminomethane			
v/v		volume per volume			
w/v		weight per volume			
Z		benzyloxycarbonyl			

Amino Acids

A	Ala	alanine	M	Met	methionine
C	Cys	cysteine	N	Asn	asparagine
D	Asp	aspartate	P	Pro	proline
E	Glu	glutamate	Q	Gln	glutamine
F	Phe	phenylalanine	R	Arg	arginine
G	Gly	glycine	S	Ser	serine
H	His	histidine	T	Thr	threonine
I	Ile	isoleucine	V	Val	valine
K	Lys	lysine	W	Trp	tryptophan
L	Leu	leucine	Y	Tyr	tyrosine

Nucleic Acid Bases

A	adenine	G	guanine
C	cytosine	T	thymine

Inhibitors

TLCK	N-α-tosyl-$_L$-lysyl-chloromethyl ketone
TPCK	N-tosyl-$_L$-phenylalanyl-chloromethyl ketone
PPACK	D-Phe-Pro-Arg-chloromethyl ketone
RVRKCMK	Decanoyl-Arg-Val-Arg-Lys-chlorometyl ketone
AAPVCMK	H-Ala-Ala-Pro-Val-chloromethyl ketone
TridecaCMK	H-Asp-Gln-Thr-Gln-Lys-Gln-Tyr-Gln-Glu-Leu-Thr-Phe-Phe-chloromethyl ketone
ZDK	Z-PheΨ[CO-CONH]Arg-Glu-Phe-OH

9.2 Index of Figures

9.3 Index of Tables

10 References

Abrahams, J. P. & Leslie, A. G. W. (1996). Methods Used in the Structure Determination of Bovine Mitochondrial F_1 ATPase. *Acta Cryst.* **D52**, 30-42.

Baker, S. C., Saunders, N. F. W., Willis, A. C., Ferguson, S. J., Hajdu, J. & Fülöp, V. (1997). Cytochrome cd_1 structure: unusual haem environments in a nitrite reductase and analysis of factors contributing to β-propeller folds. *J. Mol. Biol.* 269, 440-455.

Barret, A. J. & Rawlings, N. D. (1991). Types and families of endopeptidases. *Biochem. Soc. Trans.* **19**, 707-715.

Barton, G. J. (1993). ALSCRIPT. A tool to format multiple sequence alignments. *Prot. Engineering* **6**, 37-40.

Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*. San Diego, USA, Academic Press Inc.

Bochtler, M., Hartmann, C., Song, H. K., Bourenkov, G., Bartunik, H. D. & Huber, R. (2000). The structure of HslU and the ATP-dependent protease HslU-HslV. *Nature* 403, 800-805.

Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S. R. & Hofsteenge, J. (1989). The refined 1.9 Å crystal structure of human alpha-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J.* **8**, 3467-3475.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D 54*, 905-921.

Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J. & Huber, R. (1995). High-Level Biosynthetic Substitution of Methionine in Proteins By Its Analogs 2-Aminohexanoic Acid, Selenomethionine, Telluromethionine And Ethionine In *Escherichia Coli. Eur. J. Biochem.* **230**, 788-796.

Cesar, J. & Sollner-Dolenc, M. (2001). Trimethylsilyldiazomethane in the preparation of diazoketones via mixed anhydride and coupling reagent methods: a new approach to the Arndt-Eistert synthesis. *Tetrahedron Lett.* **42**, 7099-7102.

Collaborative Computational Project Number 4. (1994). The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst.* **D50**, 760-763.

Doyle, D. A., Lee, A., Lewis, J., Kim, E. J., Sheng, M. & MacKinnon, R. (1996). Crystal Structures of a Complexed and Peptide-Free Membrane Protein-Binding Domain - Molecular Basis of Peptide Recognition by PDZ. *Cell* **85**, 1067-1076.

Drenth, J. (1994). *Principles of protein X-Ray Crystallography.* New York, USA, Springer-Verlag.

Eichinger, A., Beisel, H. G., Jacob, U., Huber, R., Medrano, F, J., Banbula, A., Potempa, J., Travis, J., & Bode, W. (1999). Crystal structure of gingipain R: an Arg-specific bacterial cysteine proteinase with a caspase-like fold. *EMBO J.* **18**, 5453-5462.

Esnouf, R. M. (1997). An extensively modified version of Molscript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 132-134.

Etlinger, J. D. & Goldberg, A. L. (1997). A soluble ATP-dependent proteolytic system responsible for the degradation of abnormal proteins in reticulocytes. *Proc. Natl. Acad. Sci. USA* **74**, 54-58.

Faber, H. R, Groom, C. R., Baker, H. M., Morgan, W. T., Smith, A. & Baker, E. N. (1995). 1.8 Å structure of the C-terminal domain of rabbit serum hemopexin. *Structure* **3**, 551-559.

Fülöp, V., Böcskei, Z. & Polgár, L. (1998). Prolyl oligopeptidase: an unusual beta-propeller domain regulates proteolysis. *Cell* **94**, 161-170.

Fülöp, V., Szeltner, Z. & Polgár, L. (2000). Catalysis of serine oligopeptidase is controlled by a gating filter mechanism. *EMBO Rep.***1**, 277-281.

Glotzer, M., Murray, A. W. & Kirschner, M. W. (1991). Cyclin is degraded by the ubiquitin pathway. *Nature* **349**, 132-138.

Glusker, J. P., Lewis, M. & Rossi, M. (1994). *Crystal Structure Analysis for Chemists and Biologists.* Weinheim, Germany, VCH Verlagsgesellschaft GmbH.

Groll, M., Ditzel, L., Löwe, J., Stock, D., Bochtler, M., Bartunik, H. D. & Huber, R. (1997). Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* **386**, 463-471.

Groll, M., Bajorek, M., Kohler, A., Moroder, L., Rubin, D. M., Huber, R., Glickman, M. H. & Finely, D. (2000). A gated channel into the proteasome core. *Nat. Struct. Biol.* **7**, 1062-1067.

Hanahan, D. (1983). Studies on transformation of *E.coli* with plasmids. *J. Mol. Biol.* **166**, 557-568.

Harbeson, S. L., Abelleira, S. M., Akiyama, A., Barrett, R., Carroll, R. M., Straub, J. A., Tkacz, J. N., Wu, C. & Musso, G. F. (1994). Stereospecific synthesis of peptidyl alpha-keto amides as inhibitors of calphain. *J. Med. Chem.* **37**, 2918-2929.

Heukeshoven, J. & Dernick, R. (1985). Simplified method for silver staining of proteins in polyacrylamide gels and mechanism of silver staining. *Electrophoresis* **6**. 103-108.

Hiller, M. M., Finger, A., Schweiger, M. & Wolf, D. H. (1996). ER degradation of a misfolded luminal protein by the cytosolic ubiquitin-proteasome pathway. *Science* **273**, 1725-1728.

Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233(1)**, 123-138.

Innis, M. A., Myambo, K. B., Gelfand, D. H. & Brow, M. A. D. (1988). DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc. Nat. Acad. Sci.* **85**, 9436-9440.

Joshua-Tor, L., Xu, H. E., Johnston, S. A. & Rees, D. C. (1995). Crystal structure of a conserved protease that binds DNA: the bleomycin hydrolase, Gal6. *Science* **269**, 945-950.

Knight, S. D. (2000). RSPS version 4.0: a semi-interactive vector-search program for solving heavy-atom derivatives. *Acta Cryst.* **D52**, 42-47.

Kraulis, P. J. (1991). A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the bead of bacteriophage T4. *Nature* **227**. 680-685.

Laskowski, R., MacArthur, M., Hutchinson, E. & Thorton, J. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.

Liao, D. -I., Qian, J., Chisholm, D. A., Jordan, D. B. & Diner, B. A. (2000). Crystal structures of the photosystem II D1 C-terminal processing protease. *Nat. Struct. Biol.* **7**, 749-753.

Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W. & Huber, R. (1995). Crystal structure of the 20S proteasome from archeon *T. acidophilum* at 3.4 Å resolution. *Science* **268**, 533-539.

Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33,** 491-497.

McRee, D. E. (1993). *Practical Protein Crystallography.* New York, USA, Academic Press.

Merritt, E. A. & Bacon, D. J. (1997). "Raster3D: Photorealistic Molecular Graphics" *Meth. Enzymol.* **227**, 505-524.

Neer, J. N. & Smith, T. F. (1996). G-protein heterodimers: new structures propel new questions. *Cell* **84**, 175-178.

Nicholls, A., Bharadwaj, R. & Honig, B. (1993). GRASP-graphical representation and analysis of surface properties. *Biophys. J.* **64**. A166.

Otwinowski, Z. & Winor, W. (1997). Procession of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.

Palombella, V. J., Rando, O. J., Goldberg, A. L. & Maniatis, T. (1994). The Ubiquitin-proteasome Pathway Is Required For Processing the NF- B1 Precursor Protein and the Activation of NF-ĸB. *Cell* **78,** 773-785**.**

Ponting, C. P. & Pallen, M. J. (1999). β-propeller repeats and a PDZ domain in the tricorn protease: predicted self-compartmentalizytion and C-terminal polypeptide binding strategies of substrate selection. *FEMS Microbiol. Lett.* **179**, 447-451.

Rawlings, N. D. & Barret, A. J. (1993). Evolutionally families of peptidases. *Biochem. J.* **290**, 205-218.

Rock, K. L., Gramm, C., Rothstein, L., Clark, K., Stein, R., Dick, L., Hwang, D. & Goldberg, A. L. (1994). Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell* **78**, 761-771.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. H., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1998). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **233**. 1076-1078.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular cloning: A Laboratory Manual.* Cold Spring Habor Laboratory Press.

Schoepfer, R. (1993). The pRSET family of T7 promoter expression vectors for *Escherichia coli. Gene* **124**, 83-85.

Sommer, T. & Jentsch, S. (1994). A protein translocation defect linked to ubiquitin conjugation at the endoplasmic reticulum. *Nature* **365**, 176-179.

Sommer, T. & Wolf, D. H. (1997). Endoplasmatic reticulum degradation: reverse protein flow of no return. *FASEB J.* **11**, 1227-1233.

Stoltze, L., Schirle, M., Schwarz, G., Schröter, C., Thompson, M. W., Hersh, L. B., Kalbacher, H., Stevanovic, S., Rammensee, H. & Schild, H. (2000). Two new proteases in the MHC class I processing pathway. *Nat. Immun.* **1**, 413-418.

Tamura, N., Lottspeich, F., Baumeister, W. & Tamura, T. (1998). The role of tricorn protease and its aminopeptidase-interacting factors in cellular protein degradation. *Cell* **95**, 637-648.

Tamura, T., Tamura, N., Cejka, Z., Hegerl, R., Lottspeich, F. & Baumeister, W. (1996). Tricorn protease-The core of a modular proteolytic System. *Science* **274**, 1385-1389.

Tamura, T., Tamura, N., Lottspeich, F. & Baumeister, W. (1996). Tricorn protease (TRI) interacting factor 1 from Thermoplasma acidophilum is a proline iminopeptidase. *FEBS Lett.* **398**, 101-105.

Turk, D. in *Chemistry* (1992). Technische Universität München, München.

Walz, J., Tamura, T., Tamura, N., Grimm, R., Baumeister, W. & Koster, A. J. (1997). Tricorn protease exists as an icosahedral supermolecule in vivo. *Mol. Cell.* **1**, 59-65.

Walz, J., Koster, A. J., Tamura, T. & Baumeister, W. (1999). Capsids of tricorn protease studied by electron cryomicroscopy. *J. Struct. Biol.* **128**, 65-68.

Wenzel, T., Eckerskorn, C., Lottspeich, F. & Baumeister, W. (1994). Existence of a molecular ruler in proteasomes suggested by analysis of degradation products. *FEBS Lett.* **349**, 205-209.

Zheng, W., Johnston, S. A. & Joshua-Tor, L. (1998). The unusual active site of Gal6/bleomycin hydrolase can act as a carboxypeptidase, aminopeptidase, and peptide ligase. *Cell* **93**, 103-109.

11 Curriculum Vitae

Personal Data

Name: Jeong-Sun Kim

Date of Birth: 9. 24. 1969

Place of Birth: ChangSeong, Chonnam, Republic of Korea

EDUCATION

1999.7 – 2002.9	Ph.D., Max-Planck-Institute for Biochemistry, Martinsried, Germany
1997.2 –1999.2	M.S., Pohang University of Science and Technology, Pohang, Republic of Korea
1990.3 – 1997.2	B.S., Seoul National University, Suwon, Republic of Korea

MILITARY SERVICE

1990.7 – 1993.1   Military Service in the Korean Army, Kangwon, Republic of Korea

RESEARCH EXPERIENCES

1999.2 – 1999.5   Research Associate, Department of Life Science, Pohang
University of Science and Technology, Pohang, Republic of Korea

PUBLICATIONS

1. High-resolution crystal structure of delta5-3-ketosteroid isomerase with and without a reaction intermediate analogue.
Kim SW, Cha SS, Cho HS, **Jeong-Sun Kim**, Ha NC, Cho MJ, Joo S, Kim KK, Choi KW, Oh BH
Biochemistry, 1997, Nov 18; 36(46): 14030-14036.

2. High resolution crystal structure of a human necrosis factor-alpha mutant with low systemic toxicity.
Cha SS, **Jeong-Sun Kim**, Cho HS, Shin NK, Jeong W, Shin HC, Hahn JH, Oh BH
J. Biol. Chem. 1998 Jan 23; 273(4): 2153-2160.

3. Determination of the limited trypsinolysis pathways of tumor necrosis factor-alpha and its mutant by electrospray ionization mass spectrometry.
Kim YJ, Cha SS, **Jeong-Sun Kim**, Shin NK, Jeong W, Shin HC, Oh BH, Hahn JH
Anal. Biochem. 1999 Feb 15; 267(2): 279-286.

4. Crystal structure of a maltogenic amylase provides insights into a catalytic versatility.
Jeong-Sun Kim, Cha SS, Kim HJ, Kim TJ, Ha NC, Oh ST, Cho HS, Cho MJ, Kim MJ, Lee HS, Kim JW, Choi KY, Park KH, Oh BH
J. Biol. Chem. 1999 Sep 10; 274(37): 26279-26286.

5. Roles of active site aromatic residues in catalysis by ketosteroid isomerase from Pseudomonas putida biotype B.
Kim DH, Nam GH, Jang DS, Choi G, Joo S, **Jeong-Sun Kim**, Oh BH, Choi KY
Biochemistry 1999 Oct 19; 38(42): 13810-13819.

6. Contribution of the Hydrogen-Bond Network Involving a Tyrosine Triad in the Active Site to the Structure and Function of a Highly Proficient Ketosteroid Isomerase from Pseudomonas putida Biotype B.
Kim DH, Jang DS, Nam GH, Choi GD, **Jeong-Sun Kim**, Ha NC, Kim MS, Oh BH, Choi KY
Biochemistry 2000, April 39(16): 4581-4589.

7. Role of Glutamate 332 Residue in the Transglycosylation Activity of Thermus Maltogenic Amylase.

Kim TJ, Park CS, Cho HY, Cha SS, **Jeong-Sun Kim**, Lee SB, Moon TW, Kim JW, Oh BH, Park KH

Biochemistry 2000, June 39(23): 6773-6780.

8. Crystal structure of the tricorn protease reveals a protein disassembly line.

Jeong-Sun Kim, Brandstetter H, Groll M, Huber R

Nature 2001, November 22;414(6862):466-470.

9. Structural Basis for the Processive Protein Degradation by Tricorn Protease

Brandstetter H, **Jeong-Sun Kim**, Groll M, Göttig P, Huber R

Biological Chemistry 2002, July/August 383: 1157–1165

10. Structures of the tricorn-interacting aminopeptidase F1 with different ligands explain its catalytic mechanism

Goettig P, Groll M, **Jeong-Sun Kim**, Huber R, Brandstetter H

EMBO J. 2002 Oct 15;21(20):5343-5352.

11. Navigation inside a protease: substrate selection and product exit in the Tricorn protease from *Thermoplasma acidophilum*

Jeong-Sun Kim, Groll M, Musiol HJ, Behrendt R, Kaiser M, Moroder L, Huber R, Brandstetter H

J. Mol. Biol. in press

12. Purification, crystallization and preliminary X-ray diffaction analysis of fervidolysin, a keratinolytic zymogen from *Fervidobacterium pennivorans*

Jeong-Sun Kim, Kluskens L, van der Oost J

in preparation

13. Crystal structure of a keratinolytic zymogen from *Fervidobacterium pennivorans*

Jeong-Sun Kim, Kluskens L, van der Oost J, Huber R

in preparation