The Structure of Human Procathepsin S:

Crystallographic investigations on the functional role of the propeptide

Dissertation

zur Erlangung des akademischen Grades

Doctor rerum naturalium (Dr. rer. nat.)

Vorgelegt dem

Rat der Biologisch-Pharmazeutischen Fakultät

der Friedrich-Schiller-Universität Jena

von Dipl. Biol. Guido Kaulmann

geboren am 20.01.1966

in Bielefeld

Gutachter :

1.   **Prof. Dr. Bernd Wiederanders**

2.   **Prof. Dr. Rolf Hilgenfeld**

3.   **Prof. Dr. Dieter Brőmme**

Contents   Page

Footnotes

[a] $R_{\text{merge}} = \sum_{hkl} \sum_i | I_i(hkl) - I(hkl) | / \sum_{hkl} \sum_i I_i(hkl)$, where index *hkl* sums over all reflections and *i* sums over all equivalent and symmetry-related reflections (Stout & Jensen, 1968).

[b] R_{rim} is the redundancy-independent merging *R*-factor (Weiss & Hilgenfeld, 1997), which is identical to R_{meas} of Diederichs and Karplus (1997). $R_{rim} = \sum_{hkl} \{N / (N-1)\}^{1/2} \sum_i | I_i(hkl) - I(hkl) | / \sum_{hkl} \sum_i I_i(hkl)$, with *N* being the number of times a given reflection has been observed.

[c] R_{pim} is the precision-indicating merging *R*-factor (Weiss & Hilgenfeld, 1997). $R_{\text{pim}} = \sum_{hkl} \{1 / (N-1)\}^{1/2} \sum_i | I_i(hkl) - I(hkl) | / \sum_{hkl} \sum_i I_i(hkl)$.

[d] $R = \sum_{hkl} || F_{obs} | - k | F_{calc} || / \sum_{hkl} | F_{obs} |$, where *k* is a scale factor.

[e] $R_{free} = \sum_{hkl \subset T} || F_{obs} | - k | F_{calc} || / \sum_{hkl \subset T} | F_{obs} |$, where *hkl⊂T* indicates all reflections belonging to a test set T of unique reflections. The refinement is carried out without reflections in the test set, making the R_{free} value an unbiased indicator of the refinement process.

Abbreviations

APS	ammonium persulfate
B factor	mathematical expression of the isotropic temperature factor
CD	circular dichroism
dNTP	desoxy nucleotide triphosphate
E. coli	*Escherichia coli*
FCS	fetal calf serum
kDa	kilo Dalton
LB	Luria Bertoni
M	mole/litre
MES	2-morpholinoethanesulphonic acid
MHC II	major histocompatibility complex II
mM	millimolar
NADH	nicotineamide adenine dinucleotide, reduced
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
RMSD	<u>R</u>oot <u>M</u>ean <u>S</u>quare <u>D</u>eviation
SDS	sodium dodecyl sulphate
TEMED	N, N, N, N-tetramethylethylenediamine
Tris	Tris(hydroxymethyl)aminomethane
v/v	volume/volume
w/v	weight/volume

Zusammenfassung

Lysosomale Proteinasen werden in der Zelle als Vorläufermoleküle mit einer N-terminalen Prodomäne (Propeptid) synthetisiert, die die katalytische Aktivität des Enzyms blockiert. Am Beispiel des menschlichen Cathepsin S konnte gezeigt werden, dass diese Propeptide darüberhinaus als Faltungskatalysator für die Enzymdomänen agieren. Diese Funktionen der L-ähnlichen Cathepsin Propeptide, zu denen auch Cathepsin S gehört, sind mit der Ausbildung einer Raumstruktur dieser Propeptide verbunden, deren strukturelle Besonderheit – die terminale Mini-Domäne – auf diese Unterfamilie beschränkt ist. Diese Vermutung wurde sowohl *in-vitro* durch Faltung und Inhibition von Cathepsin S in Gegenwart seines Propeptids und einiger Propeptidmutanten mit ausgetauschten konservierten Aminosäuren, als auch *in-vivo* mit gleichermassen mutierten Zymogenen erhärtet. Darüberhinaus deuteten spektroskopische Untersuchungen eine von den Enzymdomänen unabhängige Ausbildung einer tertiären Struktur des Propeptids an. Die Struktur des isolierten Propeptids ist nicht bekannt. Bislang galt dies auch für die Struktur des Zymogens (Procathepsin S), allerdings deuteten Vergleiche der Proteinsequenzen innerhalb der L-ähnlichen Cathepsine auf eine hohe Übereinstimmung der Procathepsin S Struktur mit anderen Vertretern dieser Unterfamilie hin. Das Cathepsin S-Propeptid zeigte als Inhibitor bislang keine signifikante Selektivität gegenüber verwändten L-ähnlichen Cathepsinen. Umgekehrt gilt dies nur eingeschränkt. So ist beispielsweise die Hemmung des Enzyms durch das Cathepsin L-Propeptid deutlich vermindert, was auf Cathepsin S spezifische Interaktionen zwischen Propeptid und Enzym hinweist.

Zum besseren Verständnis der Interaktionen zwischen Cathepsin S und seinem Propeptid bezüglich Faltung, Inhibition und Selektivität war die Aufklärung der Raumstruktur des Zymogens unumgänglich. Darüberhinaus sollte der Frage nachgegangen werden, welche Konsequenzen die eingeführten Mutationen auf die Propeptidstruktur und seine Verankerung an die Enzymoberfläche haben könnten. Desweiteren wurde der Versuch unternommen, die unbekannte Struktur des isolierten Propeptids aufzuklären, um sie anschliessend mit dem equivalenten Element des Zymogens zum besseren Verständnis des Faltungsprozesses vergleichen zu können.

Im Rahmen dieser Arbeit konnte die Struktur einer kristallisierten menschlichen Procathepsin S Mutante (Cys25→Ala), in der das Cystein im aktiven Zentrum zur Herstellung einer inaktiven Mutante ausgetauscht wurde, bei einer Auflösung von 2.1 Å aufgeklärt werden. Im Vergleich mit anderen Proenzymen der L-ähnlichen Cathepsinen konnte innerhalb der Propeptidstruktur eine verlagerte Position der dritten Helix (α3p), die oberhalb des aktiven Zentrums die Substratbindungstasche S´1 abdeckt, festgestellt werden. Die veränderte Stellung dieser Helix wird durch eine Cathepsin S spezifische Wechselwirkung zwischen dem konservierten Met83 im Loop oberhalb dieser Helix und Phe146 des PBL (prosegment binding loop) des Enzyms beeinflusst. Phe146 ist in dieser Position einzigartig in der L-ähnlichen Cathepsin Unterfamilie und hat neben der in nächster Nähe befindlichen Aminosäure Arg141 einen entscheidenen Einfluss auf die Selektivität für andere Cathepsin-Propeptide. Darüberhinaus scheint Phe146 an der Ausbildung der Substratbindungstasche S´1 beteiligt zu sein. Im C-terminalen Bereich des Propeptids konnte ebenfalls eine derartige Wechselwirkung zwischen dem Propeptid über Pro98 und der Enzymoberfläche über Phe70 festgestellt werden. Der aromatische Seitenkettenrest von Phe70 wurde bisher der S3 Bindungstasche zugeordnet. Aus der Wechselwirkung mit dem Propeptid lässt sich eine mögliche Beteiligung an der Ausformung von S2 ableiten, welche für Kollagen als Substrat von Bedeutung ist.

Für die Stabilisierung der Propeptidstruktur und seiner Verankerung an der Enzymoberfläche zeichnen sich die drei Tryptophane 28, 31 und 52 aus, die durch ihre Stellung zueinander (Trp28 und Trp31 von Helix α1p; Trp52 von Helix α2p) einen hydrophoben Kern bilden. Die beiden Seitenkettenreste der Aminosäuren Val59 und Asn63 von Helix α2p tragen zur Verankerung des Propeptids an der Oberfläche des Enzyms über den PBL (prosegment binding loop) bei. Die Proenzymstruktur zeigt, dass der jeweilige Austausch dieser Aminosäuren durch Alanine zum Verlust bzw. zur Reduktion von Kontaktflächen zwischen den beiden o.g. Helices und zwischen Propeptid- und Enzymoberfläche führt. In allen Fällen führte dies zu einer signifikanten Abnahme der Hemm- und Faltungskompetenz der jeweiligen Propeptidmutanten. Drei der mutierten Aminosäuren (Trp52, Val59 und Asn63) gehören zum konservierten ER(F/W)N(I/V)N Motiv. In Verbindung mit Trp28 und Trp31, die mit Trp52 den hydrophoben Kern bilden, handelt es sich hierbei um ein für das Propeptid funktionell bedeutsames Motiv für die Faltung und Inhibition.

Das pH-Optimum von Cathepsin S liegt im sauren Bereich. Im Gegensatz zu anderen verwandten Cathepsinen ist Cathepsin S dennoch über ein breites pH-Spektrum aktiv und erweist sich selbst unter basischen Bedingungen als äusserst stabil. Dies kann mit dem elektrostatischen Potential des Enzyms erklärt werden, das, im Gegensatz zu den unter basischen Bedingungen sehr instabilen verwandten Enzymen, mehr neutralen Charakter hat. Dies trifft vor allem für das aktive Zentrum und seine unmittelbare Umgebung zu.

Als eine grosse Herausforderung stellte sich die Kristallisation des Cathepsin S Propeptids heraus. Die geringe Löslichkeit dieses Proteins im gefalteten Zustand erforderte die Entwicklung einer neuen Kristallisationsstrategie, die aus einer Kombination der konventionellen Methode der Dampfdruckdiffusion und einem durch einen schnellen pH-Wechsel evozierten Faltungsschritt der denaturierten Propeptidmoleküle besteht. Diese Strategie könnte in Zukunft auch für die Kristallisation von anderen Proteinen oder Peptiden mit ähnlichen Eigenschaften interessant sein.

Auch wenn die Struktur des Cathepsin S Propeptids noch nicht gelöst werden konnte, kann seine Kristallisation als zusätzliches Indiz für eine tertiäre Struktur angesehen werden. Diese Struktur sollte sich nur unwesentlich von dem equivalenten Element des Proenzyms unterscheiden. Die Basis für eine zukünftige Aufklärung der Propeptidstruktur ist gelegt. Zukünftige Untersuchungen an kristallisierten Komplexen zwischen Cathepsin S und Propeptid sollten weiterhelfen den Faltungsprozess und die darauf basierenden Interaktionen zwischen Propeptid und Enzym besser zu verstehen.

1.   Introduction

1.1   Cathepsin S: tissue distribution and function

Cathepsin S (EC 3.4.22.27) is a highly active lysosomal cysteine proteinase that was originally purified from bovine lymph nodes (Turnsek *et al*., 1975) and bovine spleen (Kirschke *et al*., 1986). The enzyme is a papain-family (C1) (Rawlings & Barret, 1993, 1994) member of the clan CA of cysteine peptidases (Kirschke *et al*., 1995), which are widely expressed throughout the animal and plant kingdoms and have also been identified in viruses and bacteria. The cathepsins, which belong to the superfamily of the papain-like proteinases, are located intracellulary in the lysosomal compartment and play an important role in protein degradation. Cathepsin S is found mainly in lymph nodes, spleen and macrophages, including rat brain glial cells, but is absent or expressed only at very low levels in kidney, liver, heart, pancreas, ovary, and small intestine (Kirschke *et al*., 1986, 1989, Qian *et al*., 1991, Petanceska & Devi, L, 1992, Petanceska *et al.,* 1994, 1996). The main occurrence of the enzyme in lymphatic tissue, especially in cells presenting antigens bound to MHC class II heterodimers, prompted speculation concerning a special function for cathepsin S in normal and diseased tissues. Cathepsin S is involved in the stepwise processing of the invariant chain li (Riese *et al*., 1996), thus liberating the binding groove for antigenic peptides to be presented by MHC class II proteins. Selective irreversible inhibition of cathepsin S resulted in a delayed immune response (Riese *et al*., 1998, Saegusa *et al*., 2002). Therefore, cathepsin S could be a target for treatment of immune-related disorders. Adenovirus mediated antisense gene transfer experiments in rats proved a further physiological role of cathepsin S for normal retinal function (Lai *et* al., 2000). Cathepsin S has a strong elastin-degrading activity (Brömme *et al*., 1996, Shi *et al*., 1992) suggesting a role in the progressive erosion of the elastic arterial walls during atherosclerosis. A selective inhibitor of cathepsin S blocked approximately 80% of the elastolytic activity of IL-1ß or IFN-γ activated primary cultures of smooth muscle cells (Sukhova *et al*., 1998). Furthermore, it shows collagenolytic activity (Kirschke *et al*., 1989). In addition to a decreased processing of the invariant chain li, Nakagawa *et al*. (1999) demonstrated a diminished collagen erosion in arthritic, cathepsin S deficient mice. Cathepsin S could be immunohistochemically localized in cortical neurons and glial cells in brains of patients with Alzheimer's disease (Munger *et al*., 1995). In addition, a high concentration of cathepsin S in rat brain glial cells (Petanceska *et al*., 1994, 1996) was shown. An overexpression of cathepsin S in HEK 293 cells led to an increased secretion of β-amyloid fragments. Like cathepsin B, L, and H, cathepsin S has also been correlated with tumor

progression (Kos *et al*., 2000). In contrast to most other lysosomal cysteine peptidases, cathepsin S is stabile at neutral and even slightly alkaline pH (Kirschke *et al*., 1989). Neither the structural basis of this stability nor its possible functional importance have been elucidated up to now, although the X-ray structure of the enzyme has been published twice (McGrath *et al*., 1998; Turkenburg *et al*., 2002). For more information the reader is referred to the reviews by Turk *et al*. (2000) and Lecaille *et al*. (2003).

1.2   Synthesis, transport and activation

As most other peptidases, cathepsin S is synthesized as an inactive precursor containing an additional N-terminal domain, the so-called propeptide or prodomain. After the cotranslational transport of the precursor into the endoplasmatic reticulum (ER), the propeptide is N-glycosylated at Asn104, a canonical glycosylation site with sequence Asn-X-Ser/Thr. The precursor is transported into the *cis*-Golgi compartment, where the oligosaccharide chain is processed by a cleavage of sugar residues, and one or more terminal mannose residues are phosphorylated (Kornfeld & Kornfeld, 1985). In this way, the recognition signal for the membrane spanning mannose-6-phosphate receptors in the *trans*-Golgi system is created (Cantor *et al*., 1992), and the precursor is transported via the known mannose-6-phosphate-way into the lysosomal compartment (von Figura & Hasilik, 1986, Griffiths *et al*., 1988, Kornfeld & Mellman, 1989). In the lysosomes, the activation or so-called maturation (processing) to the active enzyme is invoked by the cleavage of the propeptide at acidic pH (Brown & Swank, 1983, Samarel *et al*., 1989, Delbruck *et al*., 1994). Beside this way, cathepsin S can be transported into the extracellular room via the secretory way. Nissler *et al*. (1998, 2002) demonstrated a reuptake of cathepsin S into cells, and the transport of non-glycosylated proenzyme (procathepsin S) into the lysosomes suggesting an additional sorting motif in the cathepsin S molecule besides the mannose-6-phosphate residues.

1.3   Structure of cathepsin S

In 1992, human cathepsin S was cloned, and the sequence described by Shi *et al.* (1992) and Wiederanders *et al.* (1992). The precursor procathepsin S contains a propeptide of 114 amino acid residues, which include an N-terminal signal sequence of 16 amino acid residues. The domains of the mature enzyme contain 217 amino acid residues. At DNA level a high homology was shown for mature cathepsin S to other members of the papain like family. This has recently been confirmed by the determination of a cathepsin S-inhibitor complex (McGrath *et al.*, 1998) as well as of an inactive cathepsin S Cys25→Ser mutant (Turkenburg *et al.*, 2002) at structural level. As found for papain-like cysteine proteinases, cathepsin S consists of two domains, referred to as L (left) and R (right) domain (see Figure 4-20). The central helix, about 30 residues long, is the most prominent feature of the L-domain, whereas the fold of the R-domain is based on a β-barrel motif. The two domains interact through a large planar interaction surface forming a 'V'-shaped substrate-binding sites cleft with the three active site residues Cys25, His164, and Asn184. In the active form of the enzyme, the charged active site residue His164 forms an ion pair with Cys25 and Asn184. The proenzyme has a calculated molecular mass of 37,479 Da, and the active enzyme of 23,993 Da.

1.4   The propeptide of cathepsin S

1.4.1 Primary structure and homology

At DNA level, the N-terminal propeptide of cathepsin S is highly homologous to the propeptides of many other papain-family members, which are classified as cathepsin L-like endopeptidases (Karrer *et al.*, 1993, Kirschke *et al.*, 1989) (see Figure 1-2). The tertiary structure of two zymogens of this subfamily has already been elucidated, namely procathepsins L (Coulombe *et al.*, 1996) and K (Sivaraman *et al.*, 1999; LaLonde *et al.*, 1999) (see Figure 1-1). Their propeptide structure is similar: A mini-domain, the backbone of which is built by two crossed helices, α1p and α2p (suffix p stands for propeptide), anchores in the corresponding binding loop of the enzyme and directs the adjacent extended part of the propeptides in a non-productive orientation along the active site cleft between the two domains of the mature enzyme. A third helix (α3p) covers the substrate binding sites (S´) above the active-site cleft. Thus, access of substrate proteins to the catalytic site is prevented. In the protein sequence of the cathepsin S propeptide, two highly conserved motifs are found,

which have previously been identified in the same propeptide regions of cathepsin L, K, and papain: **ER(F/W)N(V/I)N** (Karrer *et al*., 1993) and **GNFD** (Vernet *et al*., 1995). The conserved residues of these sequences are involved in salt bridges and hydrogen bonds contributing to the stabilization of the tertiary propeptide structure. Therefore, the packing of the helices α1 and α2 is determined by the interactions of three aromatic residues, in the case of procathepsin L and K three tryptophanes. These aromatic residues form a hydrophobic core. In the cathepsin S propeptide sequence, the three tryptophanes are located in the same positions as found in the propeptide regions of the enzymes mentioned above. The conserved character and position of the residues supports the idea, that the structure of the cathepsin S propeptide is similar to other L-like cathepsin propeptides (Maubach *et al.,* 1997, Schilling *et al*.,2001).
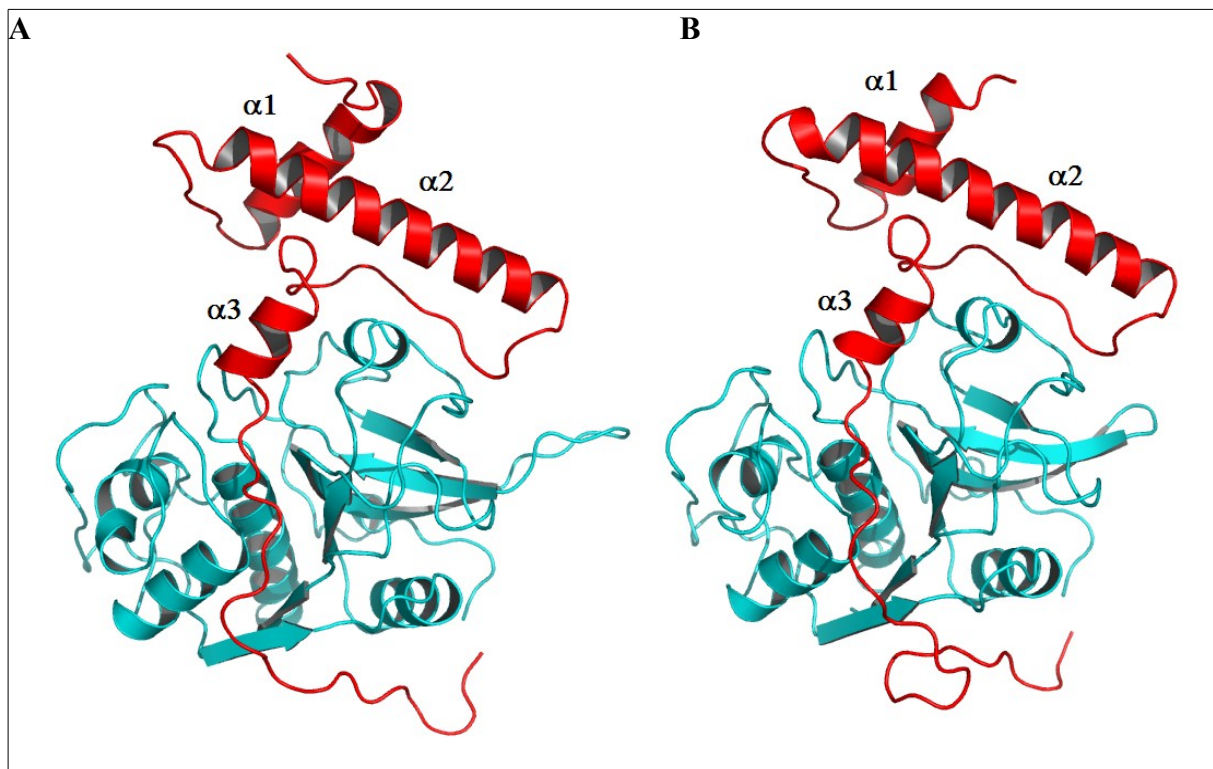
Figure 1-1 **Structures of procathepsins L and K**

Cartoon presentation of procathepsin L (picture A, PDB entry 1CS8, Coulombe *et al*., 1996) and K (picture B, PDB entry 1BY8, LaLonde *et al*., 1999). The enzyme domains are coloured in turquois, and the propeptides in red. Picture drawn with Molscript (Kraulis, 1991).

1.4.2 The propeptide as inhibitor

The inhibitory action of the propeptide was shown with recombinant propeptides, added in *trans* to the mature enzymes (Taylor *et al*., 1995; Völkel *et al*., 1996; Carmona *et al*., 1996; Maubach *et al*., 1997). Inhibition studies with cognate, non-cognate, and even chimeric propeptides have shown that the propeptide of cathepsin S is a rather non-specific inhibitor, compared to the more selective interaction of the propeptides of cathepsin L and K, respectively (Carmona *et al*., 1996; Maubach *et al*., 1997; Guo *at al*., 2000). Primary structure alignment does not reveal any reason for this observation, but tertiary structure superposition of the various propeptides may be elucidating.

1.4.3 Folding competence

An interesting aspect is the propeptide-dependent folding of cathepsin L-like subfamily members, which has been hypothesized for about one decade (Smith & Gottesman, 1989), but has been proven only recently (Yamamoto *et al*., 1999; Wiederanders, 2000; Schilling *et al*., 2001). As primary structure defines all higher structural levels of a protein, such an observation seems to be trivial (Anfinsen & Scheraga, 1975). However, in the case of α-lytic protease (Sauter *et al*., 1998), it was shown that the propeptide forms a well-structured template stabilizing a critical folding transition state of the adjacent enzyme part. This way, protease folding is kinetically controlled and thus contributes substantially to the stability of the enzyme. Such propeptides may be considered single turnover catalysts and are termed foldases or intramolecular chaperones. It is suggested that this function requires a tertiary structure of the isolated propeptide, which folds independently of the other parts of the molecule. Such an ordered structure was first indicated by various spectrospopic and fluorometric investigations on recombinant cathepsin S propeptides (Maubach *et al*., 1997), which fold via pH shift from acidic to neutral pH. Recently, the catalyzed folding of cathepsin S as a member of the papain-like cysteine proteases was studied in more detail. An alanine scan of more than ten highly conserved amino acids in the proregion revealed their functional importance, at the zymogen level *in vivo* (Kreusch *et al*., 2000) as well as *in vitro*, when the interaction of the recombinant propeptide mutants with the mature enzyme in *trans*-experiments was studied (Schilling *et al*., 2001, Pietschmann *et al*., 2002).

```
          --------α1p-----            -----------α2p-----------
             +   +                              +
ProcS  20  KDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHNLEHSMG 69
ProcK  18  PEEILDTHWELWKKTHRKQYNNKVDEISRRLIWEKNLKYISIHNLEASLG 67
ProcL  21  FDHSLEAQWTKWKAMHNRLYG-MNEEGWRRAVWEKNMKMIELHNQEYREG 69
Papain 40  STERLIQLFESWMLKHNKIYKNIDEKIYRFEIFKDNLKYIDETN----KK 85
                                      *   *  *   *   *   *
             --α3p---
ProcS  70  MHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRI--- 114
ProcK  68  VHTYELAMNHLGDMTSEEVVQKMTGLKVPLSHSRSNDTLYIPEWEGR- 114
ProcL  70  KHSFTMAMNAFGDMTSEEFRQVMNGFQN--RKPRKGKVFQEPLFYE-- 113
Papain 86  NNSYWLGLNVFADMSNDEFKEKYTGSIAGNYTTTELSYEEVLNDGDVN 133
```

Figure 1-2 Sequence alignment of propeptides

Sequence alignment of the propeptides of cathepsins S, K, L, and papain. Numbering is given for the pre-procathepsins and pre-propapain. Residues of the cathepsin S propeptide, substituted by scanning mutagenesis and their homologues, are highlighted in red. The residues belonging to the ER(W/F)N(I/V)N-motif are marked with asterisks. The aromatic residues belonging to the hydrophobic core are marked with crosses. The sequence regions belonging to the helices (procathepsin K) are labeled.

1.4.4 The ER(W/F)N(I/V)N-motif and the hydrophobic core: functional importance

In previous studies, the conserved residues in procathepsin S were subsequently replaced by alanins. In order to elucidate the functional importance of the conserved motif, human embryonic kidney cells were transfected with the mutants. Only those mutants, in which the three tryptophanes of the hydrophobic core (Trp28p, Trp31p and Trp52p) were exchanged, showed dramatic effects. They were not processed to mature cathepsin S, nor were they phosphorylated or secreted into the culture medium (Kreusch *et al*., 2000). Accompanying to these studies, cathepsin S propeptides with the same mutations were expressed (*E. coli* inclusion bodies) and refolded. As shown by CD and fluorescence emission spectroscopic stopped-flow and steady state measurements, the exchange of the tryptophanes of the hydrophobic core disturbed the folding process of the propeptide structure. Only partially folded propeptide structures were found, whereas the spectra of the wild-type propeptide point to a proper tertiary structure (Maubach *et al*., 1997, Schilling *et aöl*., 2001). Another consequence of these mutations was the loss in folding competence of the propeptide as foldases for cathepsin S, as shown in *in-vitro* refolding experiments. Therefore, a dramatic decrease of the inhibitory properties of the cathepsin S propeptide mutants was seen in inhibition studies (Schilling *et al*., 2001, Pietschmann *et al*., 2002).

1.5 Aim of the thesis

The aim of this project was to elucidate the structure of human procathepsin S. The strategy consisted of the cloning and expression of an inactive proenzyme mutant, which allows the crystallization without degradation of the propeptide by activated enzyme molecules.

The structure of the proenzyme will give insight into the interactions between the N-terminal propeptide and the enzyme domains, and will help to understand better aspects such as inhibition and folding of cathepsin S by its propeptide. The comparison of the procathepsin S structure with other related cathepsin proenzymes should help to explain the oberservations concerning the noticeable non-selective inhibition of the cathepsin S propeptide towards other related cathepsins as well as the different inhibitory behavior of several cathepsin propeptides towards cathepsin S. In the recently determined structures of human cathepsin S several cathepsin S-specific amino acids nearby the active-site have been found. Their functional role cannot not be elucidated because an interacting partner of a peptide/substrate is missing. The positions of the cathepsin S-specific residues nearby the substrate-binding sites prompted speculations concerning their role in the propeptide enzyme interactions.

Therefore, the crystallization of the expressed cathepsin S propeptide was undertaken in the hope of a successful structure determination. The comparison of the isolated propeptide structure with the respective structural element within the entire procathepsin may yield a better insight into the action of the propeptide as foldase and inhibitor. These functions of the propeptide require a very similar conformation of both, the isolated propeptide and the propeptide part within the entire procathepsin.

2. *MATERIALS*

2.1 *Molecular biology*

2.1.1 *Vectors and plasmids*

pTZ 18hCS with full-length insert encoding human procathepsin S (wildtype) in pTZ18 vector (Invitrogen); provided by G. Maubach

pcDNA3.1-3/-3$^+$ with full-length insert encoding human procathepsin S (Cys25$\rightarrow$Ala mutant) in pcDNA3.1 vector (Invitrogen); provided by S. Kreusch

pcDNA3.1-2/3$^+$ with full-length insert encoding human procathepsin S (Cys25$\rightarrow$Ser mutant) in pcDNA3.1 vector (Invitrogen); provided by S. Kreusch

pASK75-hppCS with insert encoding the cathepsin S propeptide (wildtype) in pASK75 vector (Biometra); provided by G. Maubach

pBlueBac4.5 Baculovirus Transfer Vector (Invitrogen)

2.1.2 *Bacterial strains, media*

Strain *E.coli* TOP 10 F ´ competent cells   Invitrogen

LB-Medium 5g/l yeast extract, 10g/l casitone, 10g/l NaCl, pH 7.4

LB-Ampicilin plates 5g/l yeast extract, 10g/l casitone, 10g/l NaCl, 15g/l bacto-agar, 100µg/ml ampicillin, pH 7.4

All media were autoclaved at 121°C (20 min). Ampicillin was passed through a sterile filter (0.22µm) and added after autoclaving.

2.1.3 *Kits*

Plasmid isolation Qiaspin Miniprep Plasmid Kit   Qiagen

Purification of PCR products High Pure PCR Purification Kit   Boehringer

Gel-elution Nucleotrap® DNA Extraction Kit Macherey und Nagel

Sequencing SequiTherm™ Long-Read™ Cycle sequencing kit EpicentreTechnologies

2.1.4   Enzymes and buffers

DNA-Polymerases	Pfu	Stratagene
	Taq	Gibco BRL
Ligase	T4 DNA Ligase	Gibco BRL
Restriction endonucleases	*Eco*RI	Gibco BRL
	*Bam*HI	Gibco BRL
	Dpn I	Boehringer
	*Xba*I	Gibco BRL
	*Hin*dIII	Gibco BRL
Nucleotides	dNTP-Mix	Boehringer
Buffer and additives	Buffer React 3	Gibco BRL
	Buffer H	Boehringer
	Buffer A	Boehringer
	$MgCl_2$	Gibco BRL
	Ligase Buffer	Gibco BRL

2.1.5   Oligonucleotides

The primers used for the generation of PCR products encoding human procathepsin S corres-pond to positions 158 and 1134 of the cDNA sequence (modified bases are underlined):

- **sense 5´- TCACAGGATCCATGAAACGGCTGGTTTG - 3´**

- **anti-sense 5´- ACGGGGAATTCCTAGATTTCTGGGTAAGAGG - 3´**

Primer used for site directed mutagenesis:

Cys25 →Ala :

- **sense 5´- TTGTGGTGCTGCCTGGGCTTTCAGTGCTGTGGG - 3´**

- **antisense 5´- AAAGCCCAGGCAGCACCACAAGAACCTTGATATTTCAC - 3´**

Cys25 →Ser:

- **sense 5´- TTGTGGTGCTTCCTGGGCTTTCAGTGCTGTGGG - 3´**

- **antisense 5´- AAAGCCCAGGAAGCACCACAAGAACCTTGATATTTCAC - 3**

2.2 Bacterial expression

2.2.1 Strain and inducer

Bacterial strain   *E.coli* BL21 (DE3) cells for expression   Novagen

Inducer   AHT (anhydrotetracycline) 50µg/l   Merck

2.2.2 Buffers for cell lysis

Buffer W   100mM Tris/HCl pH 8.0, 1mM EDTA, 0.02% (w/v) NaN_3
Buffer X   100mM Tris/HCl pH 8.0, 1mM EDTA, 1% Triton X-100
Buffer S   250mM sucrose, 100mM Tris/HCl pH 8.0, 1mM EDTA
Buffer V   1mM Tris/HCl pH 8.0, 1mM EDTA
Buffer D   50mM NaCl, pH 4.5, 6M GdnHCl
Buffer D1   50mM NaCl, pH 4.5, 0.5M GdnHCl, 15% acetonitril

2.3 Baculovirus expression

2.3.1 Insect cell culture

Insect cells   Sf9 (*Spodoptera frugiperda*)   Invitrogen

                          HighFive™ (*Trichoplusia nigra*)   Invitrogen

2.3.2 Bac-N-Blue™ Kit (Invitrogen)

The Bac-N-Blue™ Kit (Invitrogen) was used for the infection of insect cells. The contents are

listed below.

Cell-specific liposomes   InsectinPlus™ (1% v/v)

Transfer vector   pBlueBac4.5 (Baculovirus Transfer Vector)

Viral DNA   Bac-N-Blue™ DNA

2.3.3 Media And supplements

Media   EXCELL 400 insect media   JRH Biosciences

Antibiotics   Gentamycin (10µg/ml final concentration)   Gibco BRL

Serum   fetal calf serum (10% v/v)   Biochrom

2.3.4   Culture equipment

Culture flasks   25 - 200 cm² flask   Greiner

2.4   SDS-Page and Western blot

2.4.1   Antibodies

Primary antibody   Polyclonal antisera from rabbit against human
                              cathepsin S, obtained by immunization with recom-
                              binant human cathepsin S (Sascha), provided by E.
                              Weber (University Halle)

Secundary antibody   Anti-IgG Rabbit, coupled with peroxidase   Sigma

2.4.2   Buffers and solutions

SDS-Page   **Loading buffer:** (5x) 2,5ml SDS (20%), 1ml glycerol,
                    1,5ml 1M Tris/HCl pH 6.8, 0.55ml β-mercaptoethanol,
                    bromphenolblue
                    Running buffer: Tris/glycine-buffer (5x): 15g/l Tris,
                    72g/l glycin, 5g/l SDS
                    Washing buffer: (5x) 12g/l Tris, 45g/l NaCl,
                    2.5ml/l Triton X-100, 7.2ml/l HCl, pH 7.5 - 7.6

Western blot   **Blot buffer:** (10x) 30.3g/l Tris, 144g/l glycin, pH 8.3
                    Block solution: 5% milk powder in washing buffer
                    Staining solution: 6mg nitrobluetetrazoliumchloride,
                    20mg NADH, 10µl Phenol, 10µl H_2O_2
                    in 20ml washing buffer

Molecular weight   Prestained SDS-PAGE Standards Low Range   Bio-Rad
standard   Benchmark ™ Prestained Protein Ladder   Gibco BRL

2.4.3   Protein purification

Affinity chromatography   Concanavalin A Sepharose   Pharmacia
                    Washing buffer: 50mM Tris/HCL, 0.5M NaCl,
                    3mM NaN_3, pH 7.5
                    Binding buffer: 50mM Tris/HCL, 0.5M NaCl,
                    3mM NaN_3, 1mM $CaCl_2$, 1mM $MnCl_2$, 1mM
                    $MgCl_2$, pH 7.5
                    Elution buffer: 50mM Tris/HCL, 0.5M NaCl,
                    3mM NaN_3, pH 7.5
                    + 0.25M methyl-α-D-mannopyranoside   Merck

Gelfiltration	Superdex G75 26/60 Column	Amersham
	Washing and Elution buffer: 50mM Tris/HCl, 0.5M NaCl, 3mM NaN$_3$, pH 7.5	
Determination of protein concentration	**Bio-Rad-Protein-Assay** method according to Bradford (Bradford, 1976)	Bio-Rad
Protein concentrator	Amicon Concentrator	Millipore
	Centricon Plus-20 (5000 NMWL) Biomax-5	Millipore

2.5   *Crystallization materials and cryo tools*

Centricon Plus-20 centrifugal filter devices and Ultrafree-MC filter units	Millipore
Glass sample capillaries	GLAS
Highly liquid paraffin oil	Merck
High vacuum grease	Dow Corning
Magnetic base crystal caps	Hampton Research
mounted cryoloops	Hampton Research
24-well Linbro plates and VDX plates	Hampton Research
22mm circular siliconized coverslips	Hampton Research
sealing wax	Hampton Research
Crystal Screens 1 and 2	Hampton Research
crystal storage vials	Hampton Research
cryo canes	Hampton Research
magnetic crystal wands	Hampton Research
curved vial clamps and microtools	Hampton Research

2.6   Additional equipment

Electrophoresis	Mini-Protean II	Bio-Rad
Protein-Blot	Trans-Blot Semi-Dry	Bio-Rad
Sequencer	Li-Cor 4000L	MWG Biotech
Ultrasonication	Bandelin Sonoplus HD 70	Bandelin Electronics
Thermocontroler	PTC-100	MJ Research, Inc.
Thermocycler	Progene	Techne
Centrifuges	MC 12V	Sorvall
	Megafuge 1.0R	Heraeus Instruments
	L70	Beckman
	RC5B	Sorvall
	Heraeus Labofuge 400R	Heraeus Instruments
	Heraeus Biofuge plus	Heraeus Instruments
Spectrophotometer	UV Vis Spekol	Zeiss
	UV 1202 Photometer	Labsystems
Fluorimeter	Fluoroscan II-reader	Sartorius
Analytical balance	Sartorius BP 210 D	Sartorius
Table balance	Sartorius portable PT2100	Sartorius

pH meter	Sartorius portable PT2100	Millipore
Water purification	Milli-Qplus 185	Olympus
Microscope	Olympus SZH10 binocular	Zeiss
	Zeiss Stemi 1000 binocular	

Data collection

X-ray generator	Rotating anode Nonius FR591	Nonius
Image plate detector	Mar 300	Mar Research
	Mar 345	Mar Research
Image plate detector	Dip 2030K	Nonius
Cryostat	Oxford controller 600 series	Oxford Cryo
Air stream cooler	FTS TC-84	FTS systems
X-ray mirror system	MAC-XOS	MacScience
Synchrotron	Joint IMB Jena-University of Hamburg-EMBL synchrotron beamline X13 and EMBL beamline X11 at DESY Hamburg	
Computing	Indy workstation	SGI
	Onyx graphics workstation	SGI
	O2 graphics workstation	SGI
	Indigo2 graphics workstation	SGI

2.7   *List of suppliers in alphabetical order*

Amersham	Uppsala, Sweden
Bandelin Electronics	Berlin, Germany
Beckman	München, Germany
Biochrom	Terre Haute, Indiana, USA
Biometra	Göttingen, Germany
Bio-Rad	München, Germany
Boehringer	Mannheim, Germany
Dow Corning	Midland, USA
Epicentre Technologies	Madison, USA
FTS systems	Stone Ridge, USA
Gibco BRL	Karlsruhe, Germany
GLAS	Berlin, Germany
Greiner	Frickenhausen, Germany
Hampton Research	Laguna Niguel, USA
Heraeus Instruments	Hanau, Germany
JRH Biosciences	Lenexa, USA
Invitrogen	Leek, The Netherlands
Labsystems	Egelsbach, Germany

Macherey und Nagel	Düren, Germany
MacScience	Yokohama, Japan
Mar Research	Hamburg, Germany
Merck	Darmstadt, Germany
Millipore	Eschborn, Germany
MJ Reseach, Inc.	San Francisco, USA
MWG Biotech	München, Germany
Nonius	Delft, The Netherlands
Novagen	Lund, Sweden
Oxford Cryo	Oxford, United Kingdom
Olympus	Hamburg, Germany
Pharmacia	Freiburg, Germany
Qiagen	Hilden, Germany
Sartorius	Göttingen, Germany
Schott	Jena, Germany
SGI	Mountain View, USA
Shimadzu	Duisburg, Germany
Sigma	Taufkirchen, Germany
Sorvall	Bad Homburg, Germany
Stratagene	Amsterdam, The Netherlands
Techne	Cambridge, United Kingdom
Zeiss	Jena, Germany

3.   Methods

3.1   Cloning

3.1.1   Cloning of the human procathepsin S gene

3.1.1.1 Plasmid preparation

Plasmid DNA was purified and desalted using Qiaspin Miniprep Plasmid Kit (Quiagen) according to the manufacturer´s instruction.

3.1.1.2 Amplification of target DNA using polymerase chain reaction (PCR)

The plasmids pTZ 18hCS (wildtype), pcDNA3.1-3/-3$^+$ (Cys25$\rightarrow$Ala), and pcDNA3.1-2/3$^+$ (Cys25$\rightarrow$Ser), which contain the cDNA encoding the preproenzyme of human cathepsin S were used as templates for the PCR. The complete sequence was amplified using the forward primer 5´-TCACA*GGATCC*<u>ATG</u>AAACGGCTGGTTTG-3´ and the reverse primer 5´-ACGGG*GAATTC*CTAGATTTCTGGGTAAGAGG-3´ for PCR. The primers were deduced from positions 158 and 1134 of the cDNA sequence. The sense primer contains a *Bam*HI site (italics) and a codon for one additional N-terminal methionine (underlined), the anti-sense primer an *Eco*RI site (italics). The reaction mixture was assembled in a single tube, and then placed in the thermal cycler.

Amplification of human the gene encoding wildtype procathepsin S (pTZ 18hCS):

PCR-conditions

Reaction volume is 50μl: 10 - 20 ng DNA template
20 pmol primer
1.5 μl 50mM MgCl$_2$
2.0 μl 10mM dNTP Mix (Boehringer)
5.0 μl 10x PCR buffer for Taq polymerase (Gibco BRL)
2 U Taq polymerase (Gibco BRL)

Thermocycler programme

	Initial Denaturation Step:	95°C	3 min.
	Hotstart (adding of polymerase):	80°C	
27 cycles	Denaturation Step:	94°C	30 sec.
	Primer Annealing Step:	54°C	30 sec.
	Extension Step:	72°C	60 sec.
	Final Extension Step:	72°C	10 min.

Amplification of the genes encoding mutated procathepsin S (pcDNA3.1-3/-3$^+$ and pcDNA3.1-2/3$^+$):

PCR conditions

Reaction volume of 50µl:	80-90ng DNA template
	20pmol primer
	1.0µl 10mM dNTP-Mix (Boehringer)
	5.0µl 10x PCR-buffer for Pfu polymerase (Stratagene)
	1.25U Pfu polymerase (Stratagene)

Thermocycler program

	Initial Denaturation Step:	95°C	3 min.
	Hotstart (adding of polymerase):	80°C	
27 cycles	Denaturation Step:	94°C	30 sec.
	Primer Annealing Step:	57°C	30 sec.
	Extension Step:	72°C	60 sec.
	Final Extension Step:	72°C	10 min.

3.1.1.3 Site-directed mutagenesis

The plasmids pcDNA3.1-3/-3$^+$ and pcDNA3.1-2/3$^+$ containing the inserts encoding Cys25$\rightarrow$ Ala/Ser mutants of human procathepsin S were constructed by Stefan Kreusch in order to eliminate autocatalytic processing of the protein. Substitution of the active-site cysteine with alanine produced a catalytically inactive protein that was unable to independently process itself to the mature enzyme. The substitution was introduced by PCR-based site- directed mutagenesis according to the ExSite^{TM} protocol (Stratagene).

PCR conditions

Reaction volume is 25μl: 300ng DNA template
20pmol primer
3.0μl 50mM MgCl$_2$
3.0μl 10mM dNTP-Mix (Boehringer)
2.5μl 10x PCR-buffer for Pfu polymerase (Stratagene)
1.25U Pfu polymerase (Stratagene)

Thermocycler program

	Initial Denaturation Step:	95°C 3 min.
	Hotstart (adding of polymerase):	80°C
4 cycles	Denaturation Step:	94°C 30 sec.
	Primer Annealing Step:	58°C-0.2°C/cycle 30 sec.
	Extension Step:	72°C 8 min.
12 cycles	Denaturation Step:	94°C 20 sec.
	Primer Annealing Step:	58°C-0.2°C/cycle 30 sec
	Extension Step:	72°C 6 min.
	Final Extension Step:	72°C 10 min.

3.1.1.4 Selection of PCR products for ligation and transformation

The PCR-amplified DNA fragments were separated using agarose gel electrophoresis (1% agarose, 1x TAE-buffer, 100 V at room temperature) and cut out while exposing to UV light. The DNA was eluted using the Nucleotrap®DNA Extraction Kit (Macherey und Nagel) according to manufacturer´s instruction. After extraction, the DNA was used for ligation setups. Another way consists of the selective digestion with the restriction endonuclease

*Dpn*I, which cuts the DNA sequence G- ^{m}A-T-C to produce flush ends. This enzyme requires the methylation of adenines in both strands of the DNA in order for cleavage to occur. This way, digestion of the original template can be achieved, and only mutated PCR products will be isolated, since the non-methylated PCR-products cannot be cleaved by *Dpn*I.

Procedure: Reaction volume of 100µl: 50µlPCR reaction assay
                                                  5µl10x restriction buffer A (Boehringer)
                                               1µl *Dpn*I (2U) (Boehringer)
                                             44µl Aqua bidest.
Incubation overnight at 37°C.

3.1.1.5 Desalting of plasmid DNA

To avoid unfavourable ligation conditions, the PCR products were purified from contaminants like dNTPs, primers, and short restriction fragments using the High Pure PCR Purification Kit (Boehringer).

.

3.1.1.6 Ligation of PCR products with the baculovirus transfer vector

The PCR products were ligated into the multiple cloning site of the pBlueBac4.5 baculovirus transfer vector (Invitrogen).

Ligation over night at 14°C	wild-type	Cys25→Ala	Cys25→Ser
PCR-product	3µl / 15ng	6µl / 15ng	15µl / 15ng
10x Ligation buffer (Gibco)	4µl	4µl	4µl
PBlueBac4.5-vector (cleaved at *BamHI* / *EcoRI* sites)	15µl / 60ng	15µl / 60ng	15µl / 60ng
T4-DNA Ligase (Gibco)	1µl / 4U	1µl / 4U	1µl / 4U
Aqua bidest.	17µl	14µl	5µl
Reaction volume	50µl	50µl	50µl

3.1.1.7 *Transformation of competent E. coli cells*

Aliquots of the ligation reaction mixtures were transformed into *E. coli* TOP 10 F´ competent cells (Invitrogen) according to the manufacturer's instruction. 100 and 150 µl of the transformed cultures were plated on LB plates containing 50µg/l ampicillin. and incubated at 37°C overnight.

3.1.2 *Cloning of the cathepsin S propeptide*

The vector pASK75-hppCS contains the cDNA encoding the propeptide of human cathepsin S (amino acids 16 to 114), which had been cloned into the pASK75 expression vector (Biometra) by G. Maubach (Maubach *et al.* 1997). Downstream of the tet-promoter region, the plasmid contains the signal sequence of ompA, which enables the transport of recombinant proteins into the periplasma. A C-terminal fusion tag, the Strep-tag, encodes a ten amino acid sequence that binds streptavidin. The affinity of the tag for streptavidin allows a one-step purification of recombinant proteins. The expression of the recombinant cathepsin S propeptide into the bacterial periplasm as well as the purification of tagged soluble protein using Streptavidin-sepharose failed. In order to express and purifiy the protein from inclusion bodies, the ompA sequence and the Strep-tag were deleted using site-directed mutagenesis. The plasmids were transformed into *E.coli* BL21 (DE3) cells for expression. 100 and 150µl of the transformed cultures were plated on LB plates containing 50 µg/ml ampicillin. and incubated at 37°C overnight.

3.1.3 *Sequencing*

All inserts were sequenced to check the correct amplification and ligation using the `cycle-sequencing´ method, a combination of the Sanger dideoxy method and PCR. Following the standard protocol, the SequiTherm™ Long-Read™ Cycle sequencing kit (Epicentre Technologies) was used. For the sequencing a LI-COR 4000L Sequencer was used.

Thermocycler program

	Initial Denaturation Step:	95°C	2 min.
31 cycles	Denaturation Step:	94°C	15 sec.
	Primer Annealing Step:	50-64°C	15 sec.
	Extension Step:	72°C	30 sec.
	Cooling Down Step	4.0°C	10 min.

3.2 Expression

3.2.1 Expression of recombinant human procathepsin S in insect cells

3.2.1.1 Cell culture

Spodoptera frugiperda Sf9 and *Trichoplusia nigra* HighFive™ cells (Invitrogen) were grown at 27°C in EXCELL 400 insect medium (JRH Biosciences) supplemented with heat-inactivated 10% fetal calf serum (Biochrom) and 10 µg/ml gentamycin (Gibco BRL) as antibiotic. The cells were grown as suspension cultures in 175 cm² flasks (Greiner). At a cell density of 2.5×10^6 cells/ml the cells were diluted back to 1.0×10^6 cells and distributed into new flasks. The number of cells was counted microscopically using a `Neubauer Zählkammer´.

3.2.1.2 Transfection of insect cells and homologous recombination between the transfer vector and linearized Bac-N-Blue AcMNPV DNA

For the transfection of the recombinant DNA into Sf9 cells, the Bac-N-Blue™ Kit (Invitrogen) was used according to the manufacturer´s instructions. Viral Bac-N- Blue™ DNA, recombinant transfer DNA (pBlueBac4.5-vector containing the insert encoding human procathepsin S wild-type and mutants), and cell specific liposomes InsectinPlus™ were mixed together in EXCELL 400 insect cell media (without FCS and supplements) and incubated with freshly seeded insect cells.

Transfection mixture:

In a 1.5 ml microcentrifuge tube containing 10µl (0.5µg) Bac-N- Blue™ DNA, the following

reagents were added:	Recombinant transfer plasmid (1µg/µl)	4µl (4µg)
	EXCELL 400 media (without FCS)	1ml
	InsectinPlus™ Liposomes	20µl

After vortexing, the transfection mixture was incubated at room temperature for 15 minutes and then added dropwise onto the monolayer of Sf9 cells, which had been grown adherently grown in a 60mm dish. The dishes were incubated at room temperature for 4 hours on a rocking platform. One dish without added transfection mixture was incubated as cell-only control. After an incubation period of 4 hours, 1ml complete EXCELL 400 medium (including 10% FCS and 10 µg/ml gentamycin) was added to each dish. The dishes were sealed in plastic bags and incubated at 27°C for 72 hours. After approximately 72 hours, budded virus was released into the medium. This transfection supernatant was harvested and assayed for recombinant plaques (the transfection viral stock). 2 ml of the medium from each dish of transfected cells were transferred into a sterile 15ml polypropylene tube and stored at + 4°C. To control the success of transfection, 3 ml of fresh complete medium were added to the transfected cells. The cells were incubated for a further 7 days and visually checked for the following events, which indicate a successful transfection:

Early phase: Increased cell diameter by 25 - 50%.

Increased size of cell nuclei, which may appear to fill the cells.

Late phase: Cells appear to stop growing when compared to a cell-only control.

Granular appearance. Signs of viral budding; vesicular appearance
to cells. Viral occlusion bodies in a few cells, which appear as refractive crystals in the nucleus of the insect cell.

Very late phase: Cell lysis of a few dead cells filled with occluded virus

3.2.1.3   Isolation of recombinant virus (Plaque assay)

To avoid the dilution of the recombinant virus by wild-type virus with time, which infects and replicates more efficiently than recombinant virus, the recombinant virus titer was purified from uncut viral DNA background (occlusion body, occ$^+$) and illegitimate (non-homologous) recombinants not containing the human procathepsin S gene (wild-type and mutants). The plaque assay was performed according to the manufacturer´s (Invitrogen) instructions as briefly described below:

- Sf9 insect cells were seeded on 100mm plates at 5×10^6 cells/plate and grown till 50% confluence in order to allow cells to double before growth ceases.

- The cells were infected with the 10^{-2}, 10^{-3}, and 10^{-4} dilutions of transfection viral stock (coding for human procathepsin S wild-type, Cys25→Ala, and Cys25→Ser mutants) for 1 hour at 27°C.

- The medium was removed. The cells were overlayed with a mixture of 2.5% low melting Baculovirus Agarose (Invitrogen), complete EXCESS 400 medium (+ 10% FCS and 10µg/ml gentamycin), and 50 µg/ml X-Gal (for blue-white screening of positive plaques). After polymerization of the agarose/medium mixture the plates were incubated at 37°C until plaques were formed.

After infection of a few cells, these cells lysed and released virions, which infected nearby cells, that lysed too. This procedure led to the release of more and more virions, which subsequently infected other neighbouring cells. The recombinant virus produced ß-galactosidase, which resulted in blue plaques of infected cells because of the hydrolysis of the chromogenic substrate. After 5-6 days, blue plaques were detected and isolated for further analysis.

3.2.1.4 PCR analysis of recombinant viral DNA

The PCR allows a quick, safe and non-radioactive method to determine the presence of the insert encoding procathepsin S in the recombinant virus and to confirm the isolation of pure, recombinant plaque. The procedure was performed according to the manufacturer´s instructions:

- A 12-well microtiter plate was filled with 2ml of complete EXCELL 400 insect medium (Invitrogen) per well. 5×10^5 log-phase Sf9 cells (1 ml) were seeded per well.
- Using a Pasteur pipette, recombinant plaques (blue coloured) were picked as agarose plugs. One agarose plug was transferred into each well. One well was used for the occ$^+$ plaque assay (wild-type Bac-N- Blue™ DNA) and another as a cells-only control.
- The microtiter plate was sealed with parafilm and incubated at 27°C for 3 days.
- After 3 days, the wells were visually screened for the presence of occlusion bodies. (occ$^+$).

- 750µl of the infected cell suspension of occ⁻ wells (without occlusion bodies in the cells) and the one occ⁺ cell suspension (wild-type Bac-N- Blue™ DNA) were used for DNA isolation.

In the next step the recombinant and wild-type virus DNA was isolated as described below:

- Removal of the cell debris by centrifugation in a microcentrifuge at 5000 rpm for 3 minutes.
- After transfer of the supernatant into a fresh tube, 750µl of cold (4°C) 20% polyethylene glycol (PEG) in 1 M NaCl were added. The solution was mixed and incubated at room temperature for 30 minutes.
- After incubation and centrifugation of the mixture at maximum speed for 10 minutes at room temperature, the supernatant was carefully removed from the pellet.
- The pellet was resuspended in 100 µl sterile water, 10 µl proteinase K (5mg/ml) was added, and the mixture was incubated at 50°C in a heat block for 1 hour.
- After extraction with an equal volume (110 µl) of phenol-chloroform (1:1), the mixture was centrifuged at maximum speed in a microcentrifuge for 5 minutes at room temperature, resulting in a phase separation. The upper aqueous phase was transfered to a fresh, sterile tube.
- The DNA was precipitated by adding 1/10 volume (11 µl) of 3 M sodium acetate and 2 volumes (220 µl) of 100% ethanol. The mixture was incubated for 20 minutes at -80°C.
- After centrifugation of the mixture at maximum speed for 15 minutes at 4°C the supernatant was removed from the pellet. The pellet was washed with 70% ethanol followed by centrifugation. All traces of ethanol were removed from the pellet.
- The pellet was resuspended in 10 µl sterile water and used for PCR setups.

Commercially available primers (Invitrogen) were used for PCR with virus DNA. The forward and reverse primer had been designed to flank the polyhedrin region. These primers are compatible with all polyhedrin promoter based baculovirus transfer vectors. The forward primer binds from -44 (nt 4049) to -21 (nt 4072) downstream of the polyhedrin gene, using the nomenclature of O´Reilly *et al*. (1992). The reverse primer binds at +794 (nt 4886) to +774 (nt 4866), upstream of the polyhedrin gene.

- **Forward primer (-44):**   5´- TTTACTGTTTTCGTAACAGTTTTG - 3´

- **Reverse primer (+794):**   5´- CAACAACGCACAGAATCTAGC - 3´

Since the size of the foreign gene insert is known, the size of the expected PCR product can be determined and wild-type contamination detected:

- **Size of expected PCR fragments (wild-type Bac-N- Blue™ DNA):**   **839 bp** (control)

- **Size of expected PCR fragments (pBlueBac4.5 Transfer Vector):**   **435 bp** (control)
- Size of human procathepsin S (wt and mutants) insert: 1018 bp
- **Size of expected PCR fragment cloned into pBlueBac4.5 Transfer Vector:**   435 bp
   + 1018bp
   = **1453 bp**

PCR conditions

Reaction volume of 50µl:

10ng	DNA template
20pmol	Primer
1.5µl	50mM $MgCl_2$
2.0µl	10mM dNTP Mix (Boehringer)
5.0µl	10x PCR buffer for Taq polymerase (Stratagene)
1.5U	Pfu polymerase (Stratagene)

Thermocycler programme

	Initial Denaturation Step:	95°C	3 min.
	Hotstart (adding of polymerase):	80°C	
4 cycles	Denaturation Step:	94°C	1 min.
	Primer Annealing Step:	55°C	1 min.
	Extension Step:	72°C	3 min.
30 cycles	Denaturation Step:	94°C	1 min.
	Primer Annealing Step:	58°C	1 min.
	Extension Step:	72°C	3 min.
	Final Extension Step:	72°C	7 min.

After the PCR prodedure the samples were analyzed on a 1% agarose gel.

3.2.1.5 Preparation of high-titer viral stocks

Corresponding to the positive plaque clone, the pure recombinant virus stock (**P-1 viral stock**) was identified. In the last step, a high-titer viral stock was generated as described below:

- 2 cm² flasks were seeded with 2 x 10⁶ log-phase Sf9 cells (98% viable, doubling every 18-24 hours) in 5 ml complete EXCELL 400 medium. 20µl of the P-1 viral stock (human procathepsin S wild-type and Cys25→Ala/Ser mutants) were added to each flask and incubated at 27°C until the cells were 100% lysed. The medium containing the lysed cells (10 ml) is the **P-2 viral stock**, a small-scale, high-titer stock. 1 ml of this stock was stored at -80°C for long-term storage.

- 5ml of the **P-2 viral stock** were added to 500ml suspension of log-phase Sf9 cells seeded at a density of 2.5 x 10⁶ cells/ml. The cells were incubated for 12 days on a rocking platform at 27°C. After cell lysis, the suspension was centrifugated at 1000 rpm for 20 minutes. The supernatant was transferred to a sterile bottle and stored at 4°C. This **P-3 viral stock**, a large-scale, high-titer stock was used for the infection of HighFive™ insect cells and the expression of the recombinant protein.

3.2.1.6 Infection of HighFive™ insect cells and expression of recombinant protein

HighFive™ insect cells (Invitrogen) were used for the expression of recombinant human procathepsin S and its mutants. These oocytes from *Trichoplusia nigra* can be cultivated under serum-free conditions. The recombinant proteins are released into the culture medium, thereby facilitating the subsequent purification of the expressed protein. At first, the expression rate was tested:

- Log-phase HighFive™ insect cells (5 x 10⁶ cells/ml in 20ml EXCELL 400 medium, serum-free) were seeded in 200 cm² flasks.
- 2 ml of the **P-3 viral stock** (human procathepsin S wild-type, Cys25→Ala, and Cys25→ Ser mutants) were added. The suspension was incubated 1 hour on a rocking platform at 27°C.

- After 1 hour, the suspensions were centrifuged (10 minutes at 27°C, 1000 rpm) and the supernatants containing the virus stocks from the pellets removed.

- The pellets were resuspended in 20 ml fresh EXCELL 400 medium.

- Incubation for 60 hours on a rocking platform at 27°C.

- After incubation the suspensions were centrifuged (15 minutes at 4°C, 5000 rpm) and the supernatants containing the recombinant proteins were aliquoted and stored at -20°C.

3.2.1.7 Large-scale expression of the human procathepsin S Cys25→Ala mutant

The large-scale expression was performed by Dr. G. Schmid (Fa. HoffmannLaRoche, Basel, Switzerland) in an Airlifter system. 1.3×10^6 HighFive^{TM} cells/ml were seeded in 24 liters serum-free medium. After inoculation with 480 ml virus titer, the suspension was incubated for 60 hours at 27°C. The cells were spun down in a big kontifuge at 15,000 rpm. The supernatant was ten-fold concentrated (ca. 2.5 liter) and stored in aliquots at -20°C.

3.2.2 Expression of the cathepsin S propeptide in E.coli

As previously described by Maubach *et al.* (1997), the following steps were carried out:

- 1 ml of a glycerol stock of *E. coli* BL21 (DE3) cells (Novagen) transformed with the pASK75-hppCS were added to 100 ml LB medium and incubated on a rocking platform at 30°C. Cell proliferation was monitored until the culture had reached an absorbance of 0.5 at 550 nm.

- The expression was induced by addition of 50μg anhydrotetracycline/l culture medium. After 3 hours of further incubation, the cells were harvested by centrifugation at 6000 rpm for 12 minutes.

- The cells containing the recombinant propeptide in inclusion bodies were resuspended in **Buffer X** and disrupted by sonification (3 cycles, 1 min.., 70 W).

- After centrifugation at 14,000 rpm for 12 minutes, the pellet was washed in **Buffer X** and then resuspended in **Buffer S** for subsequent purification using sucrose density-gradient centrifugation.

3.3   *Purification*

3.3.1   *Purification of the human procathepsin S (Cys25→Ala) mutant*

An XK 16/20 column (Amersham Pharmacia Biotech) was packed with 50 ml concanavalin A sepharose (Amersham Pharmacia Biotech) resin and equilibrated with 0.1 M Tris · HCl buffer, pH 7.4, containing 0.5 M NaCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$ and 1 mM $MnCl_2$. The supernatant was loaded onto the column at a flow rate of 2 ml/min. After washing the column with 3 column volumes of 0.1 M Tris · HCL buffer, pH 7.4, containing 0.5 M NaCl, the human procathepsin S mutant was eluted with the washing buffer containing 0.25 M methyl-α-D-mannopyranoside. The final purification of the recombinant human procathepsin S mutant to electrophoretic homogeneity was achieved by gel filtration on a HiLoad Superdex-75 16/60 column (Amersham Pharmacia Biotech). The protein solution was loaded onto the column at a flow rate of 1 ml/min. The collected fractions, containing procathepsin S, were pooled and concentrated to 7.0 mg/ml by ultrafiltration (Amicon Centricon Plus-20, 5000 NMWL Millipore-Biomax 5).

3.3.2   *Purification of the human cathepsin S propeptide*

According to the previously described procedure (Maubach *et al*., 1997), the inclusion bodies were purified by sucrose density gradient centrifugation (30 ml, 40-70% (w/w) sucrose in 1mM Tris/HCl (pH 8.0), 1 mM EDTA) at 28,000 rpm for 2 hours at 4°C. Fractions containing the propeptide were pooled, washed with **Buffer V** and then solubilized with 6 M GdnHCL in 50mM sodium-acetate buffer (pH 4.5) over night. In a final step, the protein was purified by gel filtration chromatography. 1-2 ml of concentrated protein solution were loaded onto a HiLoad Superdex-75 16/60 column (Amersham Pharmacia Biotech). **Buffer D1** was used as the running buffer. The gel filtration chromatography was carried out at a flow-rate of 1 ml/min. The collected fractions containing the cathepsin S propeptide were pooled and concentrated to 7.5 mg/ml by ultrafiltration (Amicon Centricon Plus-20, 5000 NMWL Millipore-Biomax 5). The buffer was replaced by 50mM sodium-acetate (pH 4.5) containing 15 % acetonitrile.

3.3.3 Protein analysis

Both protein samples were analysed using SDS-PAGE gel electrophoresis, Western-Blot, and

Coomassie Blue staining.

- SDS-polyacrylamide gel electrophoreses was carried out at denaturing conditions
  (Laemmli, 1970) using 12.5% or 15% gels for separation.

- For immunoblotting, the gel was transferred onto nitrocellulose (0.45 μm Amersham or
  Sartorius) using a semi-dry electrotransfer unit at 15 Volt for 20 minutes. For
  immunostaining the membrane was saturated for 1 hour with block solution at room
  temperature. The diluted primary antibody (Sascha) was added and the incubation
  continued overnight. The membrane was then washed three times (15 minutes) and
  incubated with the secondary antibody (Anti-IgG Rabbit, coupled with peroxidase) for 1
  hour. After additional washing (5 x 5 minutes), the membrane was incubated with the
  staining solution. The reaction was stopped by washing with water.

- For Coomassie staining, the gel was immersed in Coomassie Blue solution (450 ml
  methanol, 90ml acetic acid, 2.5ml Coomassie Blue, 450 ml Aqua bidest.) for 1 hour and
  destained overnight in destaining solution (520 ml H_2O, 400 ml methanol, 80 ml acetic
  acid).

- The concentration of procathepsin S was measured at 595nm using the Bradford / Bio-Rad
  - Protein Assay dye solution against a BSA standard (Bradford, 1976). The concentration
  of the propeptide was calculated from A_{280} measurements in the presence of 15%
  acetonitrile, using a molecular coefficient of 27,880 cm² M^{-1} as calculated from the amino
  acid composition.

3.4   *Crystallization*

Crystallization experiments were carried out using the hanging drop vapor diffusion method (McPherson *et al*., 1982, 1999), initially with Hampton Research crystal screens I and II. Further screens were continued with solutions prepared in the own lab. Also, the gel acupuncture approach / counter diffusion technique (Garcia-Ruiz *et al*., 1994) was used in order to grow cathepsin S propeptide crystals (see **Results**).

3.5   *Collection of diffraction Data*

The crystals were flash-cooled in a 100 K nitrogen stream generated by an Oxford Cryostream cooling system (Oxford Cryosystems, Oxford, United Kingdom). The diffraction data were initially collected using an FR591 rotating copper anode X-ray generator (Nonius, Delft, The Netherlands) equipped with a 30-cm Mar Research image plate (X-ray Research, Hamburg, Germany).

3.5.1   *Data collection of the cathepsin S propeptide crystals*

Full datasets were collected using the Joint IMB Jena-University of Hamburg-EMBL synchrotron beamline X13 at Deutsches Elektronen-Synchrotron, Hamburg, at a wavelength of 0.803 Å and the beamline X11 (EMBL/DESY) at a wavelength of 0.9101 Å to maximum 3.5 Å resolution.

3.5.2   *Data collection of procathepsin S crystals*

Diffraction data were initially collected using an FR591 rotating copper anode X-ray generator (Nonius, Delft, The Netherlands) equipped with a 30-cm Mar Research image plate (X-ray Research, Hamburg, Germany). No further cryoprotectant was needed in addition to the polyethylene glycol in the mother liquor. Data extending to 3.0 Å were collected over a crystal rotation range of 105°, as recommended by the program STRATEGY (Ravelli *et al*., 1997). Subsequently, a full dataset to 2.1 Å resolution was collected using the Joint IMB Jena-University of Hamburg-EMBL synchrotron beamline X13 at Deutsches Elektronen-Synchrotron, Hamburg, at a wavelength of 0.803 Å. Diffraction intensities were recorded using a Mar Research CCD detector (X-ray Research, Hamburg, Germany).

3.6   *Data processing*

The diffraction data were indexed, integrated, and scaled using the HKL package (Otwinowski *et al*., 1997). TRUNCATE (CCP4, 1994) was used to derive structure amplitudes from the measured intensities.

3.7   *Structure solution and refinement*

3.7.1 *Procathepsin S Cys25→Ala mutant*

The alignment of amino-acid sequences of different members of the papain-like proteinases, especially the L-type cathepsins yielded the highest homology between procathepsin S and procathepsin K (~55 %), which had been crystallized in two crystal forms (Sivaraman *et al*., 1999, LaLonde et al., 1999). Based on the assumption of a high structural similarity between procathepsin S and K, molecular replacement was carried out using the program AMoRe (Navaza, 1994), with procathepsin K (LaLonde *et al*., 1999) as a search model (PDB entry code 1BY8). The rotation function was calculated with reflections between 10 and 4 Å. The model was initially refined using CNS (Brünger *et al*., 1998). After rigid body refinement, simulated annealing, and energy minimization in torsion angle space, the refinement was continued against the synchrotron data that had become available in the meantime. 5% of the unique reflections were set aside for the calculations of R_{free} (Brünger, 1992). Several cycles of refinement with CNS (Brünger *et al*., 1998), which at this stage consisted only of minimization and *B*-factor optimization, followed by refitting on a graphical workstation using the program O (Jones *et al*., 1991).

3.7.2 *Cathepsin S propeptide*

Attempts, carried out with molecular replacement are described under ***Results***.

3.8   Chemical geometry and interactions

The chemical geometry of the structure was checked by Ramachandran plots (Ramachandran & Sasisekharan, 1968), calculated with the program PROCHECK (Laskowski *et al.*, 1993). In addition, the secondary structure was checked by using the program DSSP (Kabsch & Sander, 1983). Hydrogen bonds, salt bridges and van der Waals contacts were identified and their lengths determined with the program CONTACTS (CCP4, 1994) and further inspected using the program O (Jones *et al.*, 1991). CH-π-interactions were identified with programs written by M. Brandl (Brandl *et al.*, 2001) and D. Pal. The selection criteria used to identify a CH-π-interaction are shown in Figure 3-1 and described in detail by Brandl *et al*. (2001).
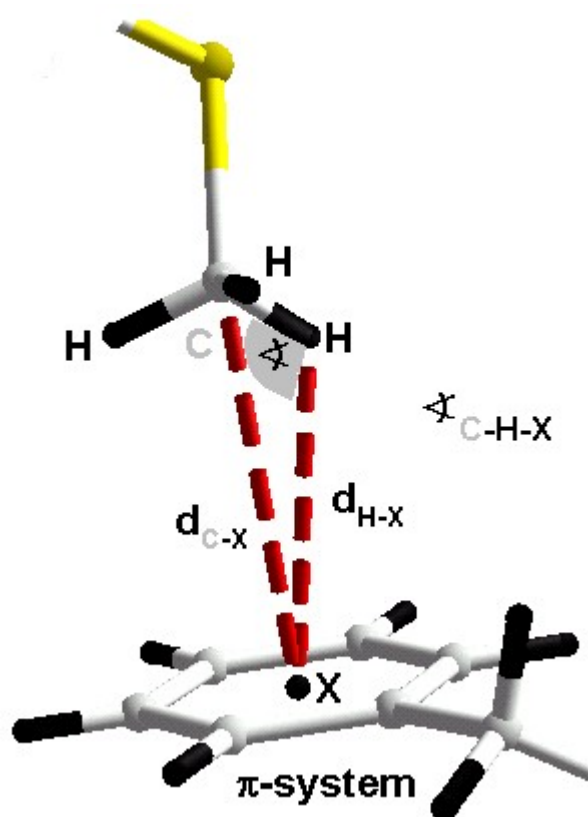
Figure 3-1

Definition of CH-π-interactions.

The π-system shown as aromatic ring of a phenylalanine. A methionine points towards the aromatic ring. The atoms are coloured: C (grey), S (yellow) and H (black).

The center-of- mass is indicated by the point X. The geometrical parameters used for identifying CH-π-interactions:
The distance between the carbon atoms and the center of the π-system (d_{C-X}). The maximum value was set to P 4.5 Å.
The angle at the hydrogen (ρ_{C-H-X}) with ρ Σ 120°.

The picture was prepared with the program SwissPdbViewer (Guex & Peitsch, 1997)

3.9   Electrostatic potential

The electrostatic potential distribution on the surface of the mature cathepsin S molecule as well as for the propeptide was calculated using the program Swiss-PdbViewer (Guex & Peitsch, 1997). The isoelectric point (pI) for each molecule including the proenzyme was calculated with COMPUTE pI/MW (Wilkins *et al.*, 1998).

4.   Results

4.1   From the gene to the purified recombinant protein

4.1.1   Expression of human procathepsin S and its Cys25→Ala/Ser mutants

4.1.1.1 Some details on the baculovirus expression system

A very thoroughly investigated eukaryotic expression system is the baculoviruses (O´Reilly *et al.*, 1992, Summers & Smith, 1988, Webb & Summers, 1990). These viruses only infect arthropods and even if they are highly virulent for some insects, they are neither virulent for vertebrates nor for plants. One of the most studied baculoviruses is the *Autographa california* Nuclear Polyhedrosis Virus (AcNPV) that has been isolated from insect larvae. Like other baculoviruses, it has a biphasic life cycle that comprises two different forms of virus:

1. extracellular virus that buds from the cells is responsible for cell-to-cell infection in cultured cells or in the insect host.

2. virus particles embedded in occlusions formed by a virus encoded protein (named polyhedrin) assure the horizontal transmission of the infection. The occlusions appear 24 hours after infection and are released into the environment after the death of the infected insect larvae. They protect the virus until they are taken up by other insect larvae and dissolved by the alkaline intestinal secretions, thus releasing the virus that subsequently enters another infectious cycle.

In cell cultures, the gene coding for polyhedrin is not essential for the infectious cycle and as polyhedrin is produced to a very high amount in the infected cells (up to 1mg per 1 to 2 x 10^6 cells), this gene has been chosen to be replaced by a foreign gene for the expression of recombinant proteins (Smith *et al.*, 1983). The foreign gene is first cloned into a plasmid transfer vector. By recombination with the wild-type virus DNA recombinant baculovirus DNA is obtained, that can be used for its high level of protein expression (Smith *et al.*, 1985, Matsuura *et al.*, 1987). In addition, expressed proteins undergo eukaryotic-specific posttranslational modifications such as glycosylation and polymerization (Luckow & Summers, 1988). Different proteinases have been successfully expressed in the baculovirus system. Baculoviruses have also been used successfully for the expression of several cathepsins (Steed *et al.*, 1998, Bossard *et al.*, 1996, Brömme & McGrath, 1996, Brömme *et*

al., 1996, Maubach *et al*, 1997, Mason *et al*, 2001). The release of the recombinant protein into the culture medium facilitates the subsequent purification. This aspect makes the expression system attractive for a large-scale expression of the protein of interest.

4.1.1.2 Cloning strategy, preparation and analysis of recombinant virus

Three full length procathapsin S encoding fragments (including the Cys25→Ala and Cys25→ Ser mutants) were cloned into the pBlueBac4.5 baculovirus vector (Invitrogen). This vector contains the early-to-late (ETL) promoter and the polyhedrin promoter from AcMNPV (see Figure 4-1).

Figure 4-1:

Cloning strategy for the fragment encoding human procathepsin S Cys25→Ala mutant.
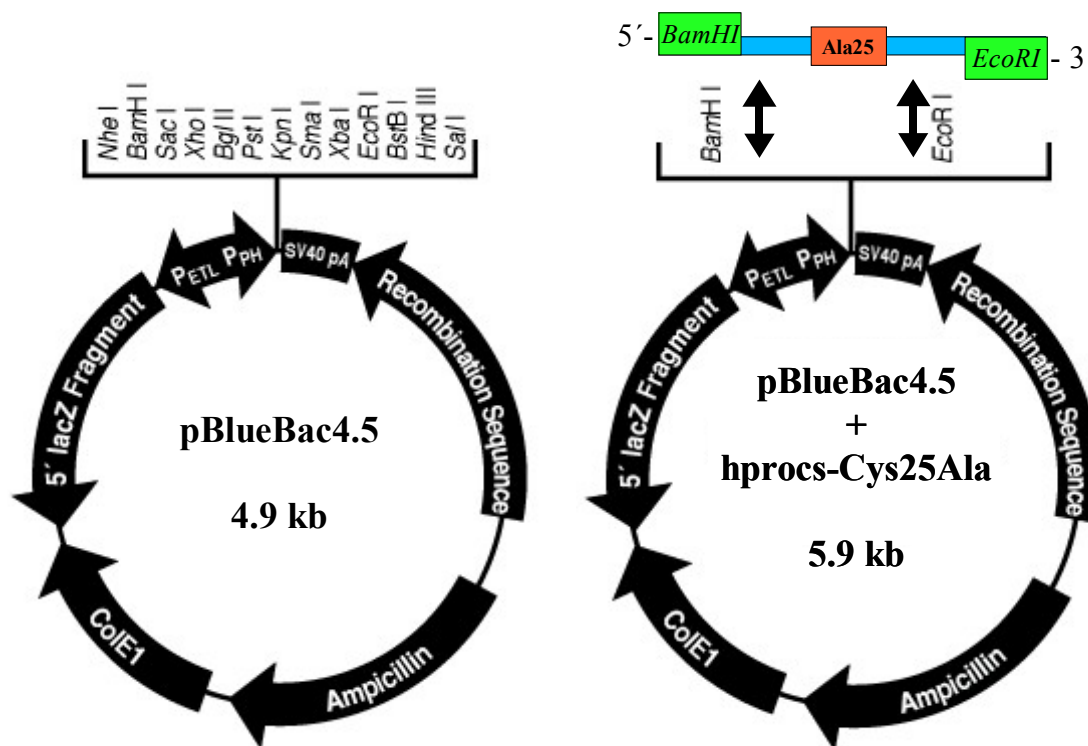
The ETL promoter directs the synthesis of β-galactosidase (Crawford & Miller, 1988) while the polyhedrin promoter controls the synthesis of foreign gene products. The vector recombines with modified wildtype AcMNPV-DNA (Bac-N-Blue™ DNA, Invitrogen) to yield recombinants, which produce foreign gene products and form blue plaques when 5-bromo-4-chloro-3-indolyl-β-D-galactosidase (X-gal) or a derivative is present in the agarose overlay of the plaque assay. Figure 4-1 shows the cloning strategy for the procathepsin S gene

into the transfer vector and Figure 4-2 the recombination strategy with the modified wildtype AcMNPV-DNA. The last step before proceeding to the virus stock preparation consists of the selection of recombinant virus using the plaque assay in order to isolate pure clones. The recombinant plaques (blue coloured, occ⁻) were analyzed using PCR (see Figure 4-3).

Figure 4-2

Recombination Strategy

pBacBlue4.5

ORF1629 Procathepsin S gene P_{PH} P_{ETL} 5´*lacZ*

5´ORF1629 3´*lacZ*
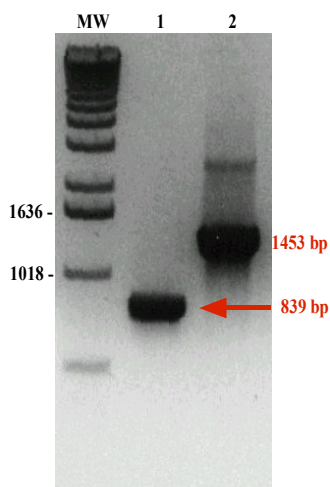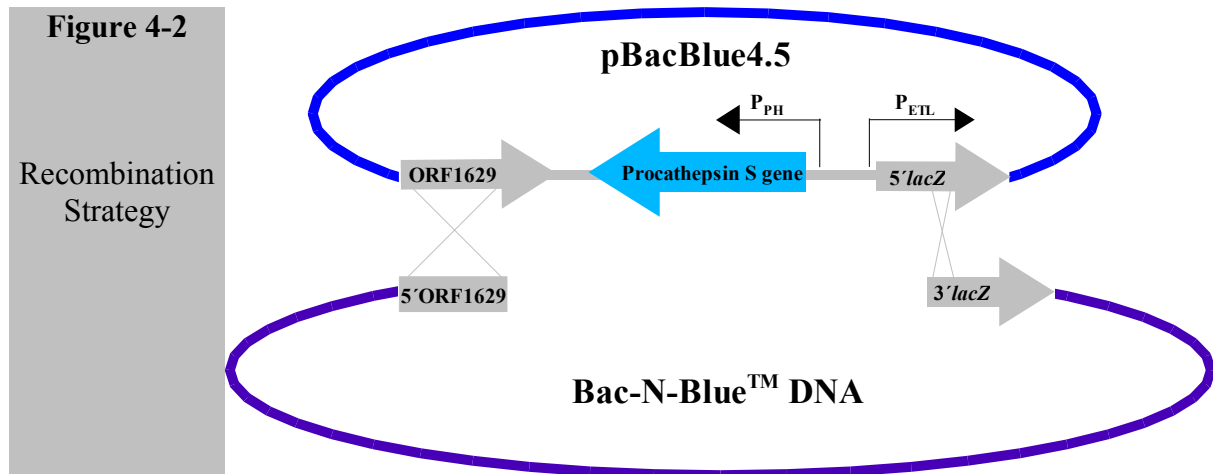
Bac-N-Blue™ DNA

Figure 4-3   PCR analysis of recombinant viral clones

Samples were analyzed for the recombinant PCR fragment (1453 bp) encoding human procathepsin S and its mutants (Cys25→Ala/Ser), and the wild-type Bac-N-Blue™ DNA PCR fragment (839 bp). DNA was isolated from the P-1 viral stocks and used as template in PCR reactions with the recombinant baculovirus forward and reverse primers. 20 µl from each sample were analyzed on a 1% agarose gel. DNA was also isolated from a wild-type plaque and used as a negative control. Lane 1: wildtype control. Lane 2: PCR sample using viral DNA from a pure recombinant plaque as template.

The identified pure recombinant low-scale P-1 virus stocks were amplified to the middle-scale P-2 and then subsequently amplified to the large-scale P-3 virus stocks. The first infection of HighFive™ insect cells (Invitrogen) with the virus stocks gave a successful expression of recombinant human procathepsin S and of the procathepsin Cys25→Ala/Ser mutants. The large scale expression was performed by Dr. G. Schmid from Hoffmann LaRoche in Basel (Switzerland). Initially, all virus stocks were tested for their expression rate in different insect cells (see Figure 4-4).

Figure 4-4 Western blot analysis of expressed human procathepsin S

The Western Blot shows the expression of human procathepsin S and mutants in different insect cells (Sf9, [*Spodoptera frugiperda*], and HighFive™, [*Trichoplusia nigra*]). The protein fractions were separated by SDS-PAGE under non-reducing conditions and blotted onto a nitrocellulose membrane. For the immunodetection of the recombinant procathepsin S, a rabbit polyclonal antibody against cathepsin S and its prodomain was used. All lanes show the precursor form (ProCS) and the mature enzyme form (CS). Protein fractions of the wildtype procathepsin S show additional intermediate products in form of double protein bands (red arrow).
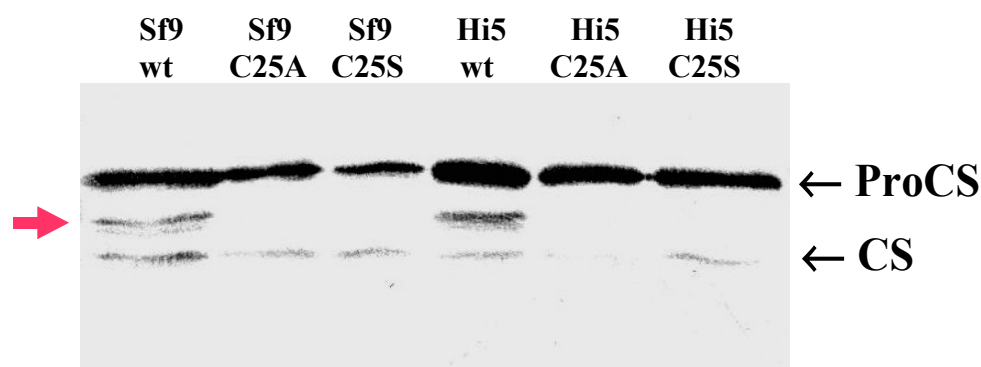
The expression of recombinant procathepsin S and the mutants showed no significant difference in Sf9 cells and in HighFive™ cells. In all samples, low amounts of processed mature enzyme were found due to proteolytic activity of endogenous proteinases in the insect cells. Additional intermediate products in form of double protein bands were observed in the lane of the wild-type proenzymes migrating at ~ 28 kDa, pointing to a cleavage in the propeptide region by processed cathepsin S. The maturation of procathepsin S is interpreted as an intramolecular or intermolecular cleavage of the propeptide (Quraishi and Storer, 2001) but not understood in detail.

4.1.1.3 Large-scale expression of the human procathepsin S Cys25→Ala mutant

The recombinant virus stock containing the gene encoding the procathepsin S Cys25→Ala mutant was used for the large-scale expression (see ***Methods***) in order to eliminate autocatalytic processing of the protein. The concentration of the virus stock was determined using plaque assays, which revealed a MOI (multiplicity of infection) of 1-1.5 x 10^8 pfu/ml (plaque forming units/ml) (G. Schmid, personal communication). The substitution of the active-site cysteine with alanine resulted in a catalytically inactive protein that was unable to process itself to the mature enzyme. This way, the loss of proenzyme during expression, purification and later crystallization by unimolecular degradation was avoided. HighFive insect cells were used for infection and expression since the growth of these cells under

serum-free conditions facilitate the purification of recombinant protein. 60 hours after infection the released recombinant human procathepsin S mutant was detected in the supernatant using SDS-PAGE and Western blot.
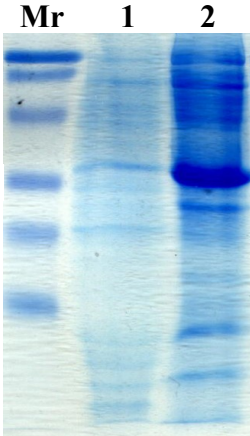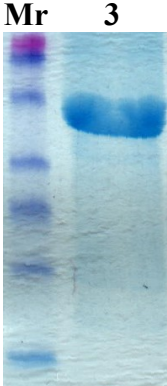
	Purification steps	Yield of protein
	Supernatant of culture broth (10 x concentrated) (*Lane 1*)	345 mg in 500 ml
	Affinity chromatography (ConA-Sepharose)	17.04 mg in 192 ml
	Concentration using Amicon MWL 10.000 (*Lane 2*)	14.79 mg in 9 ml
	Gelfiltration chromatography	
	Concentration using Centricon MWL 5000 (*Lane 3*)	$C_{[Protein]}: \approx 7.0$ mg/ml

Figure 4-5 SDS-PAGE analysis of the purified human procathepsin S Cys25→Ala mutant.

SDS-PAGE was performed under denaturing conditions.
The SDS-gels were stained with Coomassie Blue.
Mr: standard. Red arrows mark the procathepsin S band.
The protein concentrations were determined using the Bradford assay.

4.1.2 Purification of the recombinant human procathepsin S Cys25→Ala mutant

Procathepsin S is glycosylated at the position Asn104p in the propeptide, enabling the purification using affinity gel chromatography. Final purification of the recombinant human procathepsin S mutant to electrophoretic homogeneity was achieved using gelfiltration chromatography (see *Methods*). The purification of 500ml 10 times concentrated culture supernatant revealed ~ 1 % of the recombinant human procathepsin S Cys25→Ala mutant (see Figure 4-5). The protein was concentrated to 7.0 mg/ml in 0.1 M Tris · HCl, pH 7.4, 0.5 M NaCl and stored in aliquots at -20°C.

4.1.3 *Bacterial expression and purification of the human cathepsin S propeptide*

A glycerol stock of *E.coli* BL21 (DE3) cells (Novagen) transformed with the pASK75-hppCS expression plasmid was provided by G. Maubach. The plasmid DNA encoding the cathepsin S propeptide was isolated and checked by restriction analysis for size (see Figure 4-6), and sequenced for correct amplification and ligation. The recombinant human cathepsin S propeptide was expressed in *E. coli* BL21 (DE3) and recovered from inclusion bodies (10 mg/l bacterial culture) (see *Methods*). The purification (see *Methods*) using one gelfiltration step achieved a highly purified protein (see Figure 4-7), which was concentrated to 7.5 mg/ml in an acidic buffer containing 50mM sodium acetate (pH 4.5), 15 % acetonitrile.

Figure 4-6 (left)
Restriction analysis of pASK75-hppCS. After digestion with *Xba*I and *Hin*dIII, 2 fragments were separated on a 1% agarose gel: the 3054 bp pASK75 vector fragment and the 300 bp insert coding for the cathepsin S propeptide (red arrow).

Figure 4-7 (right)
SDS-PAGE analysis of purified protein samples under denaturing conditions.
The SDS gel was silver-stained. The red arrow (right) points towards the recombinant cathepsin S propeptide.

4.2 *Crystallization and structure determination*

4.2.1 *The crystallization of the cathepsin S propeptide*

4.2.1.1 Strategy

With a length of 99 amino acid residues (without N-terminal signal peptide) and a molecular weight of 11932 Da (Maubach *et al*. 1997) the propeptide of human cathepsin S is a small and hydrophobic protein. In sodium phosphate buffer (pH 6.5), the propeptide is in a folded state and only poorly soluble (~ 36μg/ml) (Maubach *et al*. 1997, Schilling *et al*. 2001, Pietschmann *et al*. 2002). The low solubility makes the cathepsin S propeptide unsuitable for structure determination using NMR as well as for crystallization. To circumvent this problem the protein was dissolved in an acidic buffer containing 50mM sodium acetate at pH 4.5 with 15 % acetonitrile. At these conditions, the protein is unfolded or only partially folded due to a breakdown of the tertiary structure (Maubach *et al*., 1997, Schilling *et al*., 2001, Pietschmann *et al*., 2002). The dissolution in an acidic buffer supplemented with acetonitrile, allows the rise of the protein concentration. The protein was then subjected to crystallization screens, where the crystallization buffer should fulfill two important conditions:

1. A fast pH shift from acidic (pH 4.5) to nearly neutral conditions after mixing of the protein buffer and the crystallization buffer, rendering a switch from **denaturing** to **folding** conditions. The pH shift is a precondition for the folding process of the cathepsin S propeptide (Maubach *et al*. 1997).

2. The crystallization of the **folded** cathepsin S propeptide. In the crystallization assay, the pH shift allows the use of unsoluble protein in a high concentration, dissolved in a denaturing buffer without the formation of aggregates, precipitates, or other phases before equilibration with the crystallization buffer.

4.2.1.2 Crystallization using the hanging-drop vapor diffusion method

Using Crystal Screens I and II (Hampton Research), crystallization buffers were selected which allow such a pH shift. Only MES (2-morpholinoethanesulfonic acid) was found as a suitable buffer, and isopropanol as precipitant. A 2 μl hanging drop containing the protein solution was equilibrated with the same amount of crystallization solution containing 1.14 M MES (2-morpholinoethanesulfonic acid) at pH 6.75 and 10 % isopropanol (see Figure 4-8).

The mixture of crystallization solution with protein solution (50mM sodium acetate, 15% acetonitrile, pH 4.5) in a ratio of 1:1 gave a pH of ~ 6.7, indicating a probable fast pH shift within in the droplet after mixing the two different solutions. The hanging drops were incubated at 4°C, 20°C and 30°C. Furthermore, protein, buffer, and precipitant were used in various concentrations in the crystallization assays.
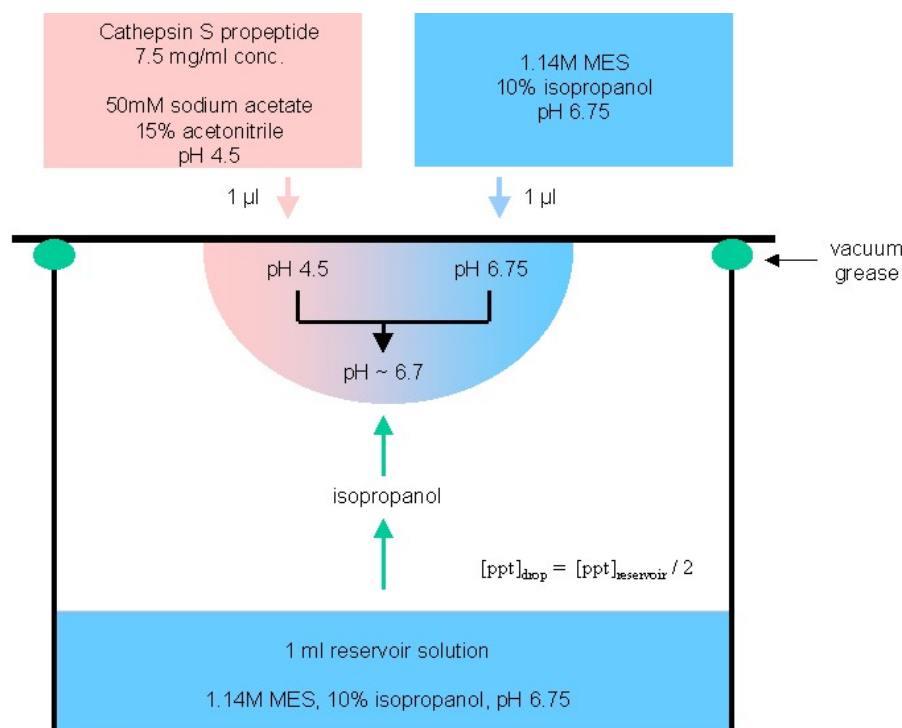
Figure 4-8 Crystallization strategy using the vapor diffusion method (hanging drop).

After mixing the crystallization buffer with the buffer containing the protein, a fast pH shift from 4.5 to nearly 6.7 in the hanging drop takes place. The concentration of the precipitant [ppt] in the reservoir is initially higher, usually about twice of that in the droplet, and the two reach a balance when the osmolarity of the drop becomes equal to that of the reservoir. This occurs by water exchange from the droplet to the reservoir. Here, a gradual transfer of isopropanol from the reservoir to the droplet occurs due to a higher vapor pressure of isopropanol compared to water.

The variation of buffer concentration revealed a significant influence on the number of crystals grown. Above 1.0 M MES, the appearance of crystals and their growth was observed, whereas below 1.0 M, many small crystals were observed in only a few droplets. The largest crystals were obtained at a MES-concentration of 1.14 M (see Table 4-1).

Temperature	conc. of MES (pH 6.75) [M/l]	conc. of isopropanol [%]	conc. of propeptide [mg/ml]	Observations
4 °C	0.5 - 1.5	1 – 30	7.5	no crystals only precipitate
20°C	1.0 - 1.5	5 – 20	5 - 7.5	**after 2 - 3 months:** • many small crystals, max. size of 20 x 20 x 10 μm • no further growth • formation of precipitate
30°C	1.0 - 1.5	10	5 - 7.5	**after 4-5 days:** • many small crystals, max. size of 50 x 20 x 20 μm • formation of precipitate **after 4 weeks:** • crystal growth to maximum size of 80 x 60 x 40 μm **after 3 - 6 months:** • crystal growth to maximum size of 100 x 70 x 50 μm • decreasing mass of precipitate, in some drops precipitates disappeared
30° C	1.14	10	7.5	**after 4-5 days:** • in some drops 10 - 15 crystals at maximum size of 50 x 20 x 20 μm • formation of precipitate **after 4 weeks:** • crystal growth to maximum size of 120 x 80 x 40 μm **after 3 - 6 months:** • crystal growth to maximum size of 250 x 150 x 100 μm • decreasing mass of precipitate, in some drops precipitates disappeared

Table 4-1

Crystal growth statistics

The temperature seems to be very important for the growth of cathepsin S propeptide crystals. A crystallization at a temperature of 30°C resulted in an acceleration of crystal growth, and only a few crystals appeared in some droplets. These crystals grew faster than those observed at 20°C and reached a larger size. Furthermore, the appearance of precipitate was observed 1-

2 days after setting up the droplets at all temperatures (see Table 4-1). In several droplets, the precipitate disappeared after several weeks or months. In that time, the crystals reached their maximal size. A variation of the concentration of the precipitant (+/−10%) yielded no improvement concerning number of crystals within the droplets, velocity of crystal growth, and crystal size. In droplets set up without isopropanol, no crystals were observed, independently of temperature, or concentration of protein or buffer. It was also possible to grow crystals using 10 % polyethylene glycole (PEG 8000) instead of isopropanol, but the crystals reached only a very small size. The pH of MES plays an crucial role in the crystallization of the propeptide. As expected, no crystals were observed below pH 6.75 of MES, indicating the importance of the pH shift from pH 4.5 to at least 6.5 for the folding of the propeptide molecules. Increasing of the pH to a maximum of 8.0 gave no improvement of the crystals. Alternatively, unbuffered ammonium sulfate was used as reservoir solution, which will develop NH_3 thus leading to a pH shift in the droplets. But growth of crystals could never be observed. Figures 4-9 and 4-10 give an impression of the crystals grown at 30°C:
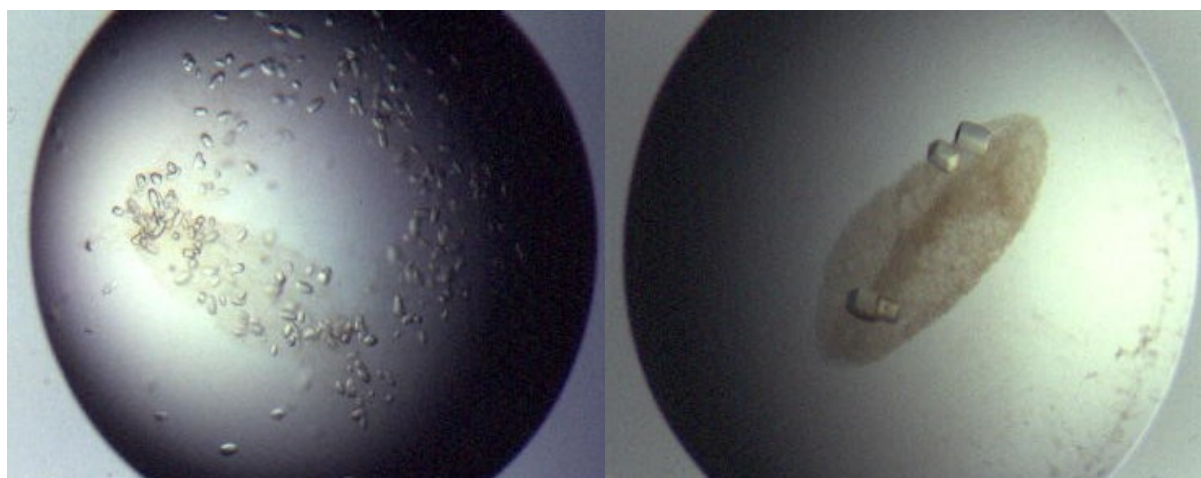
Figure 4-9: Drops with cathepsin S propeptide crystals in different numbers.

Two selected drops containing cathepsin S propeptide crystals grown at the same conditions (30°C). Picture A represents a drop containing a high number of crystals grown within 2 - 6 months. Picture B represents a drop containing only 4 crystals, which reached a bigger size than the crystals in the drop shown in picture A.

After 4 - 5 days:
crystals with a maximum
size of 50 x 20 x 20 µm

after 4 weeks:
crystals with a maximum
size of 100 x 80 x 40 µm

After 3 – 6 months:
crystals with a maximum
size of 250 x 150 x 100 µm
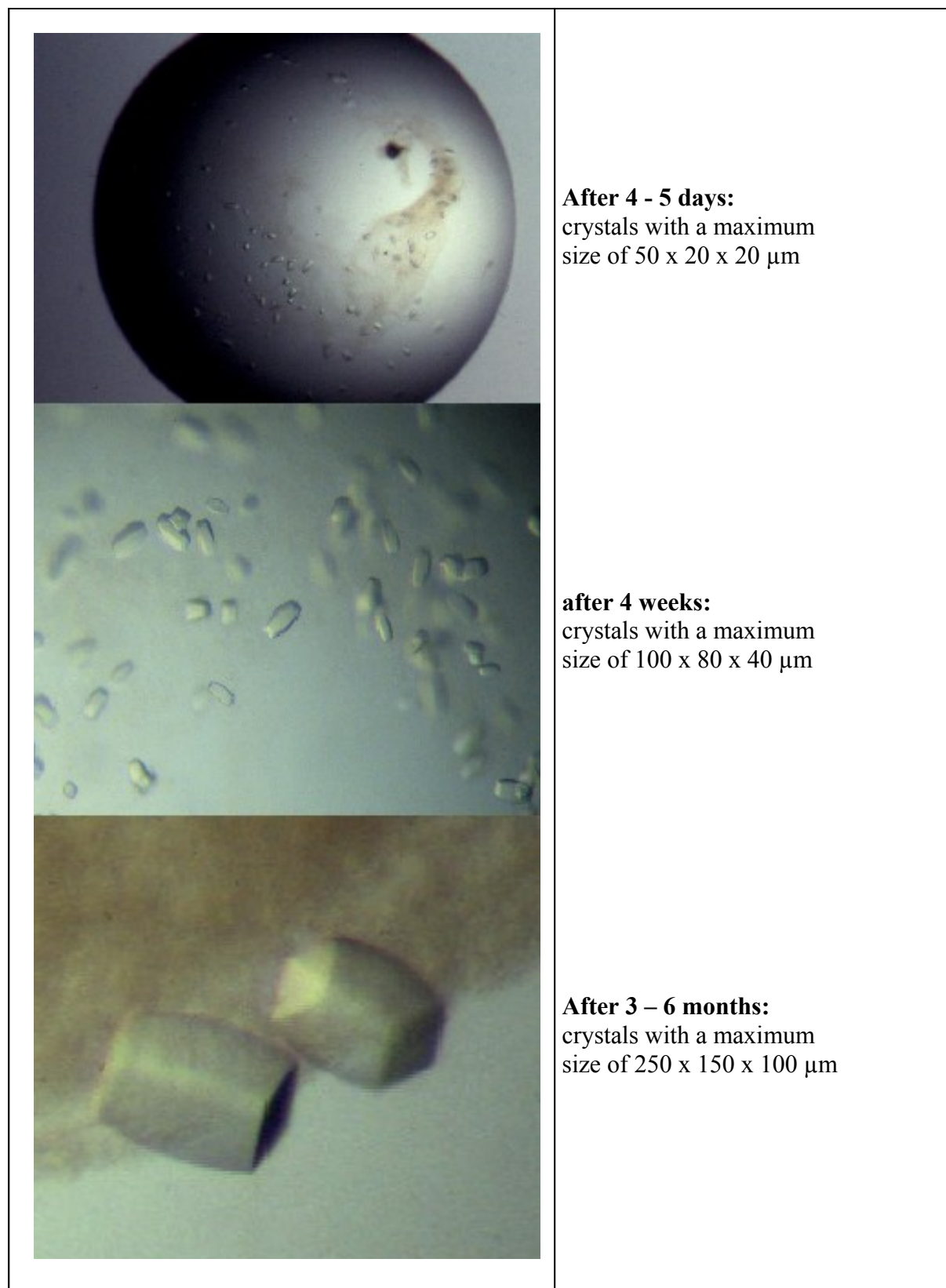
Figure 4-10 Cathepsin S propeptide crystals.

Apparently of tetragonal morphology, these crystals were grown in a buffer containing 1.14 M MES (pH 6.75) + 10 % isopropanol at 30°C.

4.2.1.3 Crystallization of the cathepsin S propeptide using the counter diffusion technique

In addition to the vapor diffusion method, crystallization was also undertaken in capillaries in order to improve cathepsin S propeptide crystals. J. M. Garcia-Ruiz has pioneered this method (Garcia-Ruiz & Moreno, 1994). Capillaries containing a protein-solution were stuck into a gel-precipitant reservoir and the crystal growth in the capillaries was observed. The crystals grew all along the capillary. In the capillary, a near continuum of growth conditions developed, caused by a diffusion of gelled crystallization buffer and precipitant into the capillary. J. M. Garcia-Ruiz called this method *gel acupuncture* approach. A variant of this method is the *counter diffusion* technique, which has been developed for the crystallization in space. Coupling between crystallization and diffusion is achieved by using a box, named Granada Crystallization Box (GCB) (see Figure 4-11).
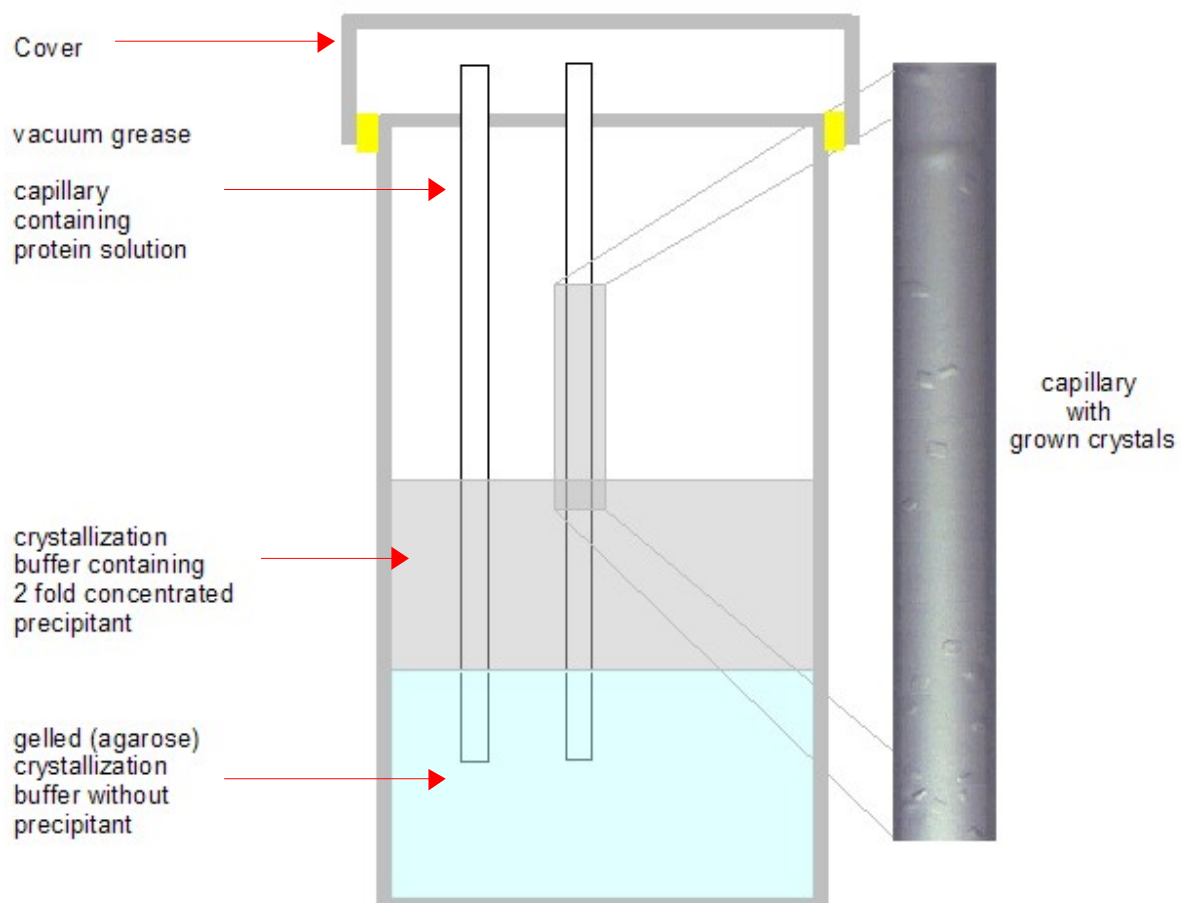
Figure 4-11   Counter diffusion technique using the Granada Crystallization Box (GCB)

The box was designed for the growth of protein crystals inside capillaries with un-gelled precipitating agent by the counter diffusion technique (Sauter *et al*. 2001). The counter-diffusion method requires the preparation of an agarose gel crystallization buffer mixture, which is filled into the bottom of the box. After insertion of the capillaries with the protein solution the gel is overlayed with the crystallization buffer containing 2 fold concentrated precipitant (see Figure 4-11). Firstly, the precipitant diffuses into the gel. When the concentration of the precipitant in the gel and the overlaid buffer has reached an equilibrium, the precipitant will further diffuse into the capillary. The method allows a delay of the crystallization process. The delay enables the transfer of crystallization assays into space without premature crystallization events in the capillaries, a precondition for the study of protein crystallization under microgravity conditions. The crystallization experiments using the GCBs were carried out in the laboratory of Prof. JuanMa Garcia-Ruiz in Granada (Spain).

- An agarose-crystallization buffer (1.14 MES, pH 6.75) mixture containing 1.5 % (w/v) agarose was heated to 80°C under continuous stirring. The mixture was boiled in order to break cross-links of the agarose fibers. When the solution had become transparent, boiling was continued for about two minutes.

- The agarose-buffer solution was cooled down to 50°C. 5ml of the solution were transferred into each GCB and the gel was allowed to polymerize at room temperature.

- Four Capillaries (0.3 mm diameter) with a length of 7 cm were filled with the protein solution (7.5mg/ml purified cathepsin S propeptide in 50mM sodium acetate + 15 % acetonitrile at pH 4.5). Then the upper end of the capillary was sealed with vacuum grease.

- The capillaries were pushed into the gel by about 2-3 mm. Finally, 5ml of the crystallization buffer containing 20% isopropanol were poured onto the gel layer. A cover was placed on the GCB and sealed with vacuum grease.

- he GCBs (each of them contained 2 capillaries) were incubated at 20 and 30°C, respectively.

After four weeks many tiny crystals have grown all along these capillaries, which had been incubated at 30°C. Precipitant has been formed in the capillary part inserted into the gel. After 6 months, the crystals reached a size of nearly 50 μm in length. After 1 year the largest crystals reached a size of nearly 150μm in length and 30 μm in width (see Figure 4-12). Pictures A, B, and C (Figure 4-12) show crystals, which were grown within the capillary. In the same GCB the crystals grown in a second capillary appeared of the same morphology but with additional edges (see pictures D, E, and F in Figure 4-13). No crystals were observed in these capillaries, which were incubated at 20°C.
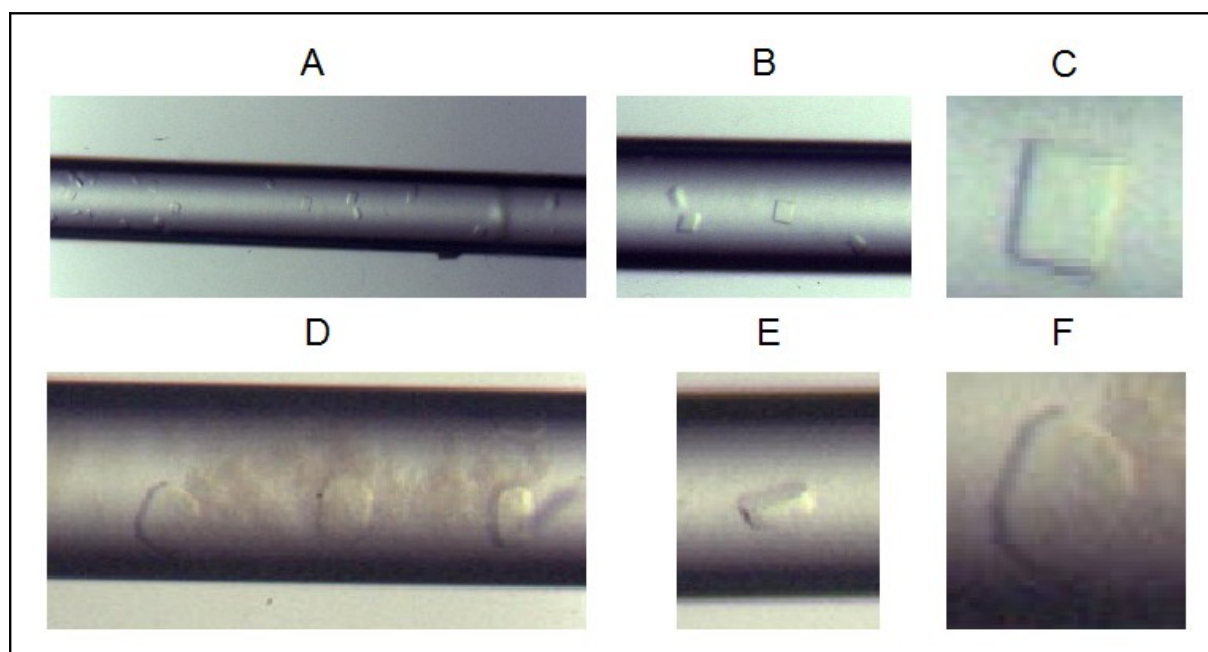
Figure 4-12 Cathepsin S propeptide crystals grown using the counter diffusion technique.

Crystals grown in capillaries using the counter diffusion technique, one year after incubation at 30°C. A, B, C show crystals with a maximum size of 100 x 100 x 30 μm. D, E, F show crystals in another capillary of the same GCB, with a maximum size of 150 x 120 x 30 μm. With respect to morphology, two slightly different crystal types were obtained.

4.2.1.4 Collection of diffraction data and processing (cathepsin S propeptide)

All crystals grown in hanging drops at a temperature of 30°C, had a minimal length of 100 μm and diffracted to a maximum resolution of 7 - 6 Å at the in-house X-ray source (IMB Jena). At the synchrotron beamline, the crystals diffracted to a maximal resolution of 3.5 Å. The crystals were flash-cooled in a 100 K nitrogen stream. The use of cryoprotectants (different oils) did not improve the diffraction of the crystals. Several trials with crystals mounted in capillaries revealed no improvement of the diffraction either. The crystals are tetragonal, with unit-cell dimensions $a = b = 151.1$ Å and $c = 75.8$ Å. The unsatisfactory quality and low resolution of the collected data prevented the exact determination of the space group, but the presence of two twofold axes in addition to the four-fold axis is evident ($P4_x2_y2$). Two datasets were collected and processed, the data statistics of the best dataset is shown in Table 4-2.

Cell parameters	$a = b = 151.1$ Å, $c = 75.8$ Å
Resolution range	30.0 – 3.50 Å
Number of observations	74 063
Number of unique reflections[a]	11482
R_{merge}[a]	8.8 % (44.5 %)
R_{rim}[b]	9.7 % (49.4 %)
R_{pim}[c]	3.8 % (19.0 %)
I/σ	11.8 (4.8)
Completeness of data	96.8 % (92.9 %)

Table 4-2   Data collection statistics for the cathepsin S propeptide

Statistics of data (cathepsin S propeptide crystal) collected at Synchrotron beamline X11 (EMBL/DESY Hamburg), at a wavelength of 0.9101 Å. Values for the outermost resolution shell (3.63 - 3.50 Å) are shown in parantheses.

The crystals, which were grown by the counter diffusion technique diffracted to a maximum resolution of 4.5 - 5 Å at the Joint IMB Jena-University of Hamburg-EMBL synchrotron beamline X13 at Deutsches Elektronen-Synchrotron, Hamburg, at a wavelength of 0.803 Å. They were also tetragonal with the same cell parameters. No data sets were collected due to the low diffraction resolution.

4.2.1.5 Attempts to determine the structure of the cathepsin S propeptide

Several trials with molecular replacement using AMoRe (Navaza, 1994) were carried out using an poly-alanine model of the propeptide from procathepsin K (PDB entry 1BY8, LaLonde *et al.*, 1999). After determination of the human procathepsin S Cys25→Ala mutant (this work) the propeptide domain of this proenzyme was also used for the rotation and translation functions. Assuming 8 molecules per asymmetric unit the packing density was 2.3 $Å^3$/Da, which corresponds to approximately 45.7 % solvent in the crystal (see Table 4-3). This value is within the range found by Matthews (1968) for most protein crystals.

Monomers/A.U.	V_m ($Å^3$/Da)	% solvent fraction
1	18.3	93.2
2	9.1	86.7
3	6.1	79.6
4	4.6	72.9
5	3.6	66.1
6	3.0	59.1
7	2.6	52.5
8	**2.3**	**45.7**
9	2.0	38.9
10	1.8	32.1

Table 4-3 Matthews coefficients for the cathepsin S propeptide

Matthews coefficients (V_m) and solvent fraction parameters (cathepsin S propeptide), dependent on the number of monomers per asymmetric unit. The values in bold are assumed to be correct. Values correspond to the estimated copy number per a.u.. Calculations based on an asymmetric unit volume of 1.73 x 10^6 $Å^3$ and an approximate mass for the monomer of 11932 Da.

The molecular replacement approach yielded very poor solutions. The rotations and translations gave correlation factors of ~ 0.20 and R factors of more than 60 %, independently of model and number of copies used in the calculations. After rigid body refinement using CNS (Brünger *et al.*, 1998), electron-density maps were calculated but were not interpretable. The low resolution of the data and the high number of monomers in the asymmetric unit has not yet allowed phasing by molecular replacement.

4.2.2 Structure determination of the human procathepsin S Cys25→Ala mutant

4.2.2.1 Crystallization

The human procathepsin S Cys25→Ala mutant was crystallized by using the hanging drop vapor diffusion method. Crystals were obtained by equilibrating 2µl 7.0 mg/ml protein solution with the same volume of reservoir solution (100mM Tris · HCl, 200mM magnesium acetate, 20% polyethylene glycol 8000, final pH 7.5–7.8). Within 7 days at 20°C, platelet-shaped crystals appeared and grew to a maximum size of 500 x 250 x 30 µm within the next 4 weeks (see Figure 4-13).

Figure 4-13 Procathepsin S crystals.

Platelet-shaped crystals (maximum size of 500 x 250 x 30 µm) of the human procathepsin S Cys25→Ala mutant grown in a hanging drop (vapor-diffusion method).

Collection of diffraction data and processing

A full dataset to 2.1 Å resolution was collected using the Joint IMB Jena-University of Hamburg-EMBL synchrotron beamline X13 at Deutsches Elektronen-Synchrotron, Hamburg, at a wavelength of 0.803 Å (X-ray diffraction pattern see Figure 4-14) (see **Methods**). The space group was determined to be C222$_1$, with unit cell dimensions a = 59.8 Å, b = 140.5 Å and c = 78.6 Å. The asymmetric unit contained one molecule, corresponding to a packing density of 2.31 Å^3/Da and a solvent content of ~ 47 % (Matthews, 1968).
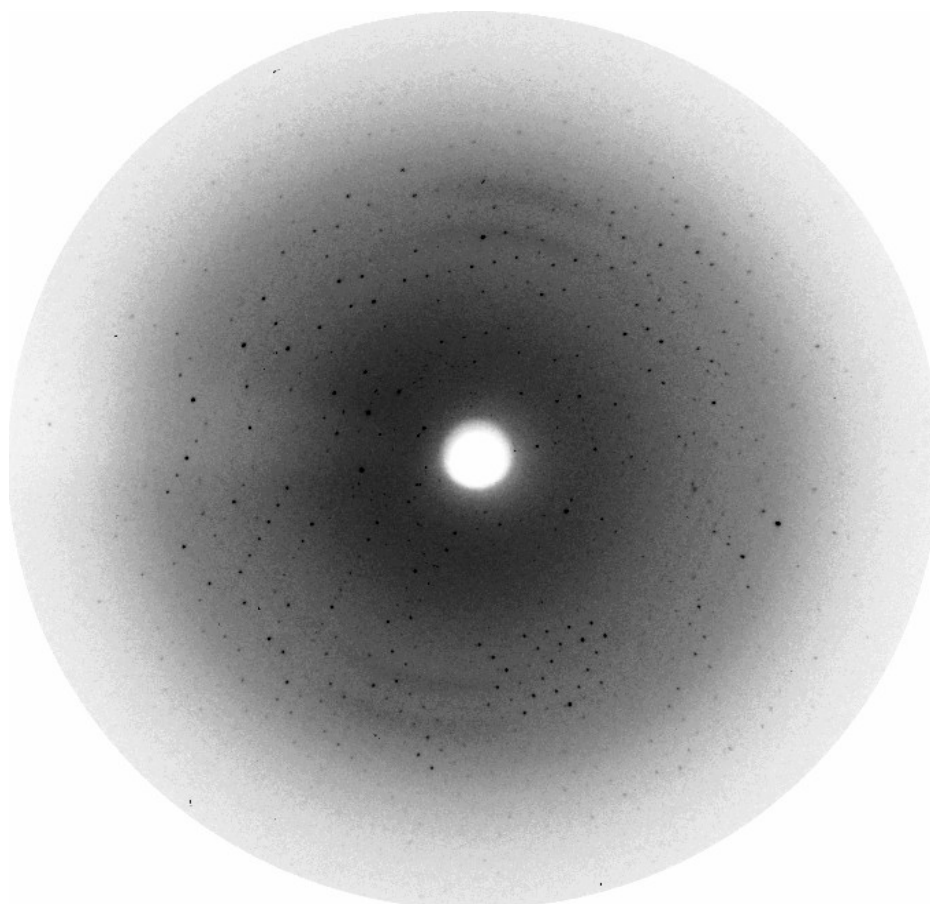
Figure 4-14   X-ray diffraction pattern of human procathepsin S crystal (Cys25→Ala)
at a resolution of 2.1 Å.

Space group	C 2 2 2$_1$
Cell parameters	a = 59.8, b = 140.5, c = 78.6 Å
Resolution range	30.0 – 2.10 Å
Number of observations	283.005
Number of unique reflections[a]	19.350
R_{merge}[a]	8.6 % (47.1 %)
R_{rim}[b]	9.5 % (53.6 %)
R_{pim}[c]	3.6 % (24.1 %)
I/σ	14.4 (3.0)
Completeness of data	97.9 % (93 %)

Table 4-4 Data collection statistics for the human procathepsin S Cys25→Ala mutantStatistics of data (cathepsin S propeptide crystal) collected at synchrotron beamline X13 (Joint IMB Jena-University of Hamburg-EMBL), at a wavelength of 0.803 Å. Values for the outermost resolution shell (2.14-2.10 Å) are shown in parentheses.

4.2.2.3 Structure determination and refinement

Molecular replacement was carried out using the program AMoRe (Navaza, 1994), with procathepsin K (LaLonde *et al.*, 1999) as a search model (PDB entry code 1BY8) (see **Methods**). The rotation function was calculated with reflections between 10 and 4 Å and yielded one solution with R = 52.5 % and a correlation coefficient of 16.7 %. The translation function also calculated with data between 10 and 4 Å gave a solution with R = 46.1 % and a correlation coefficient of 41.8 %. After rigid body refinement the R-value could be reduced to 44.7 % with a correlation coefficient of 46.6 % (see Table 4-5).

Table 4-5 *Molecular replacement statistics for the human procathepsin S Cys25→Ala mutant*			
	Rotation	*Translation*	*Rigid body fitting*
α	79.12	79.12	78.41
β	43.96	43.96	42.95
γ	212.21	212.21	213.55
x	0.00	0.1794	0.1809
y	0.00	0.0987	0.0829
z	0.00	0.1809	0.0857
Correlation coefficient	5 %	41.8 %	46.6 %
R ^{d}	52.5 %	46.1 %	44.7 %

The model was initially refined using CNS (Brünger *et al.*, 1998). After rigid body refinement, simulated annealing, and energy minimization in torsion angle space, the *R*-value at 3.0 Å could be reduced to 0.35. The refinement was continued against the synchrotron data that had become available in the meantime. 5% of the unique reflections were set aside for the calculations of R_{free} (Brünger, 1992). The *R* and R_{free} for 30-2.1 Å resolution data was 0.227 and 0.277, respectively. Several cycles of refinement, which at this stage consisted only of minimization and *B*-factor optimization, followed by refitting on a graphical workstation using the program O (Jones *et al.*, 1991), resulted in a model (see Figure 4-17) that contained

nearly all residues of the protein: residues 18p-107p, 114p of the prosegment (marked with the suffix p) and residues 1-217 of the mature cathepsin S. Residues 16p, 17p and 108p-113p are missing due to the poorly defined electron density. After introduction of 242 water molecules the R and R_{free} could be reduced to 19.8 % and 24.4 %, respectively. Refinement statistics are given in in Table 4-6.

Refinement

Resolution range	30.0 – 2.10 Å
Number of protein atoms	2459
Number of water molecules	242
Crystallographic *R*-factor	19.8 %
R_{free} [e]	24.4 %
Rmsd (bonds)	0.009 Å
Rmsd (angles)	1.35°
Average *B*-factor	27.6 Å^2

Table 4-6

Refinement statistics for the human procathepsin S Cys25→Ala mutant

4.2.2.4 Quality of the structure model

The asymmetric unit contains one molecule. The r.m.s. deviation of bond lengths and angles of the structure are 0.009 Å and 1.35° respectively, suggesting good chemical geometry. A Ramachandran plot (Ramachandran and Sasisekharan, 1968) (see ***Methods***) gave 86.1% of residues in the most favoured region and 13.5% of the residues in the additional allowed regions. No residues were found in the disallowed region, and only one residue (Thr106p) was found in the generously allowed region (see Figure 4-15). The residues of the published sequence (Shi *et* al., 1992, Wiederanders *et al.*, 1992) fit well into the electron density, with the exception of a few residues in the solvent exposed C-terminal region of the prodomain (Lys108p – Arg113p) as well as the residues Ala 16p and Gln 17p at the N-terminus (see Figure 4-16). Residues Lys108p - Arg113p were not built into the poorly defined electron density. Continous density was observed in F_o - F_c maps at 0.6 sigma, but this is far below noise level for an F_o - F_c map. A cleavage within this region can be excluded. In the samples of the purified procathepsin S protein checked by SDS-PAGE under denaturing conditions,

additional protein bands, which might be due to a cleavage of the propeptide domains of cathepsin S have never been observed.

Figure 4-15

Ramachandran plot for the 2.1 Å structure of the human procathepsin S Cys25→Ala mutant.
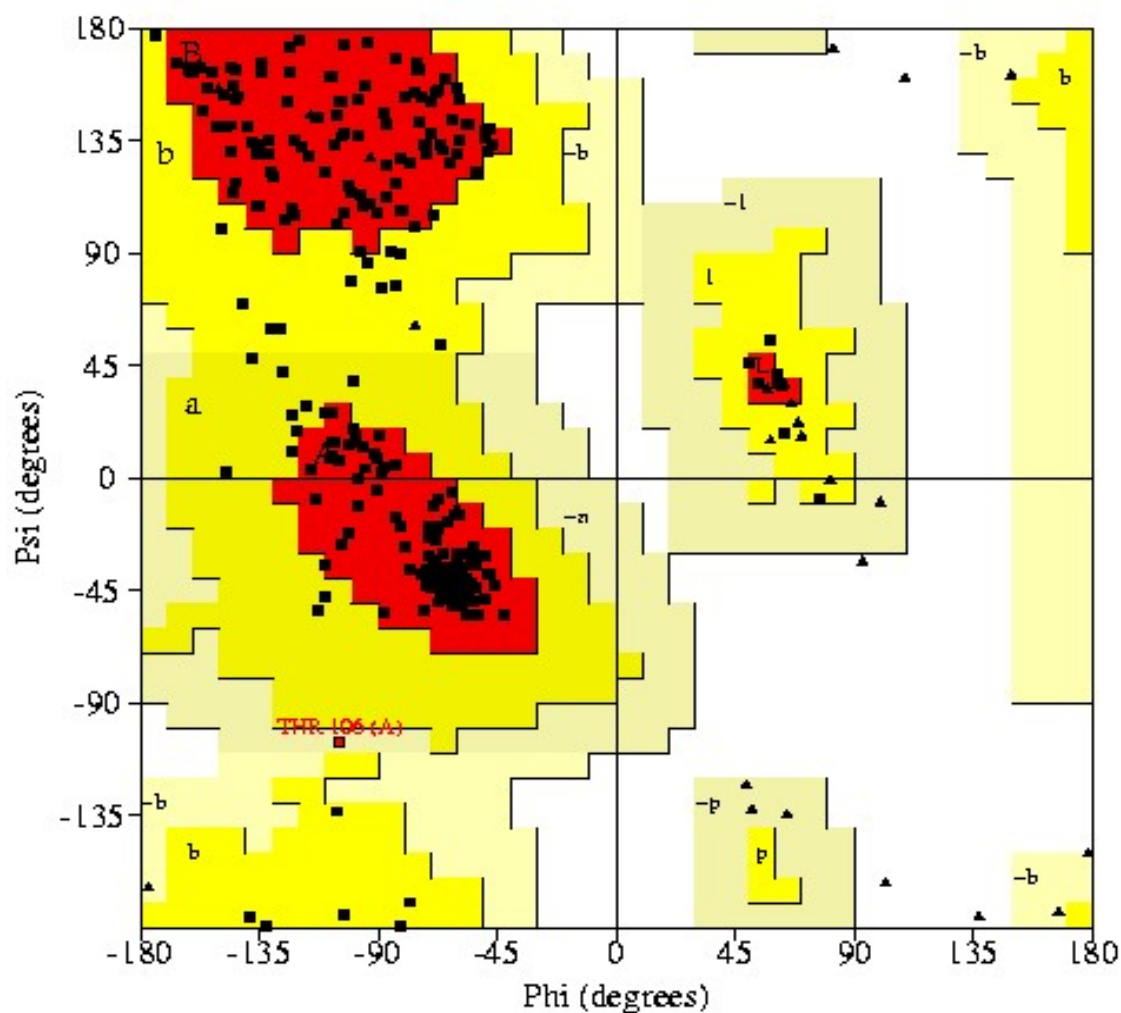
Plot statistics

Residues in most favoured regions [A, B, L]	230	86.1 %
Residues in additional allowed regions [a, b, l, p]	36	13.5 %
Residues in generously allowed regions [~a, ~b, ~l, ~p]	1	0.4 %
Residues in disallowed regions	0	0,0 %
Number of non-glycine and non-proline residues	267	100 %
Number of end-residues (excl. Gly and Pro)	4	
Number of glycine residues (shown as triangles)	27	
Number of proline residues	10	
Total number of residues	308	
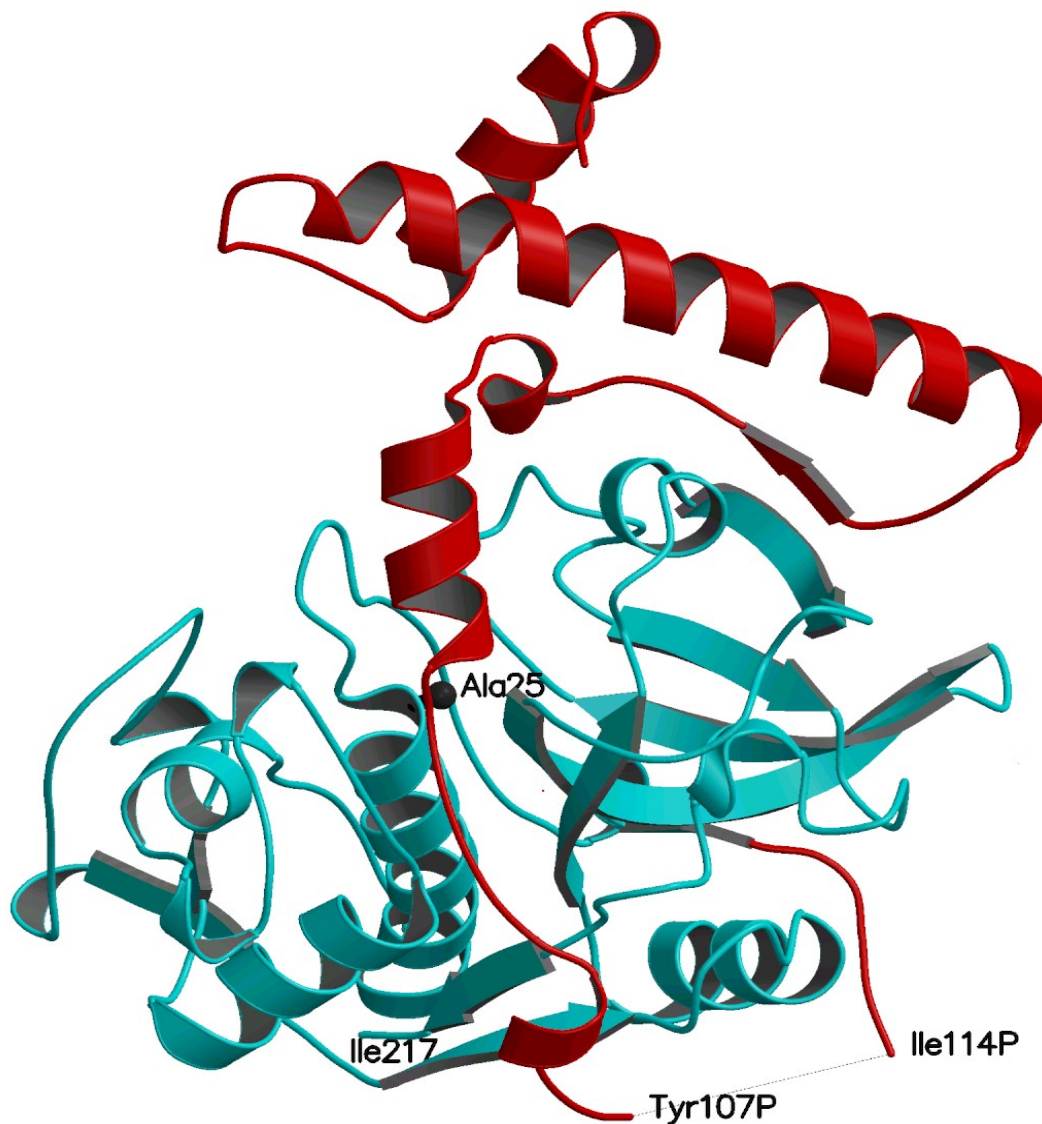
Figure 4-16 Overall structure of the human procathepsin S Cys25→Ala mutant.

The molecule consists of 97 residues of the propeptide (red) and 217 residues of the mature enzyme (turquois). The dashed line marks the 6 residues at the end of the C-terminal part of the propeptide missing due to poorly defined electron density. The C-terminus of the mature enzyme (Ile217) and the Ala25 in the active site are labeled. Picture drawn with Molscript (Kraulis, 1991).

4.2.3 The structure of the human procathepsin S Cys25→Ala mutant

4.2.3.1 Overall description

The molecule of procathepsin S consists of 114 residues of the proregion (marked with "p" in Figure 4-17) and 217 residues of the mature enzyme (Figure 4-17). The signal peptide (1p – 15p) has been cleaved off, so that the numbering of the propeptide starts with the first N-terminal residue Ala 16p of the proenzyme. The propeptide contains an N-terminal globular domain, which consists of three α-helices (α1p-α3p) and a β-strand (β1p). These interact with the prosegment-binding loop (PBL), which is an omega-loop of the enzyme domain. A 3_{10}-helix (3_{10}1p) is formed between helix α3p and strand β1p.
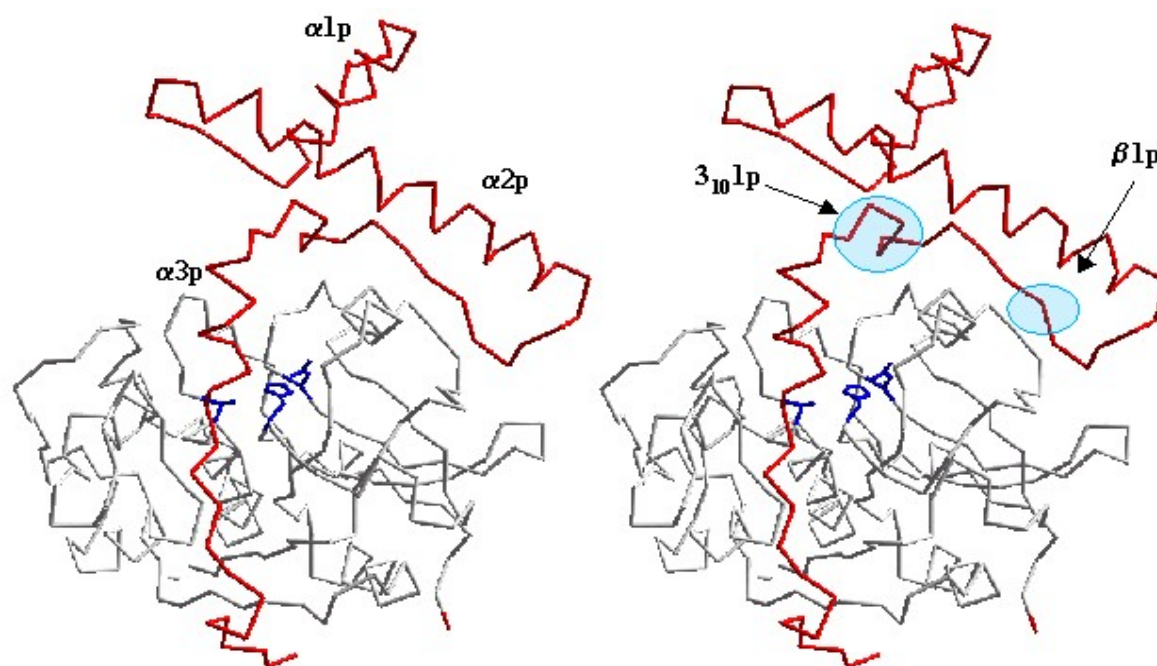
Figure 4-17 Stereo picture of the human procathepsin S Cys25→Ala mutant structure

The propeptide domain is coloured in red and the mature enzyme in grey. Residues of active site (Cys25→Ala25, His164, Asn184) are highlighted in blue. The regions including the 3_{10}-helix (3_{10}1p) and the β-strand (β1p) underlaid in light blue. The picture was prepared with the program Swiss-PdbViewer (Guex & Peitsch, 1997).

Being homologous to all members of the papain-like proteinases, the mature enzyme consists of two domains, referred to as L (left) and R (right) (see Figure 4-19), which between them form the V-shaped active-site cleft. The overall fold of human procathepsin S is similar to the structures determined for other members of the L-like cathepsins such as human procathepsin K and L (LaLonde *et al.*, 1999, Coulombe *et al.*, 1996) as well as to procaricain (Groves *et*

al., 1996). The superposition of procathepsin S with procathepsin K, L and B (PDB entries 1BY8, 1CS8 and 3PBH) gave an RMSD of 0.79, 0.88 and 1.62 Å for 278, 277 and 246 Cα atoms as target pairs (see Figure 4-19). The high RMSD between procathepsin S and B can be explained by structural differences in the globular domain of the propeptide (see Figure 4-18). Furthermore, the additional *occluding-loop* of the mature cathepsin B structure, a characteristic structural element of this enzyme, influences the RMSD of the superposition with both L-type cathepsins.
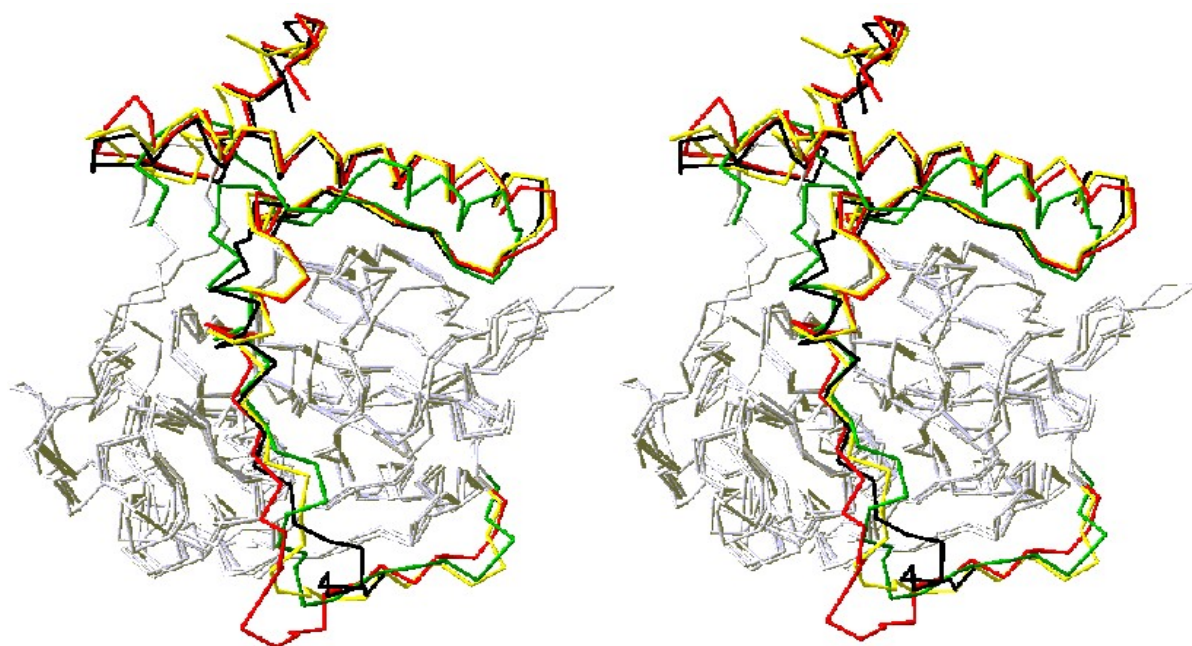
Figure 4-18 Superimposed procathepsin structures.

Stereo picture of superimposed structures of procathepsin S (black), K (red), L (yellow), and B (green) (PDB entries 1BY8, 1CS8, 3PBH). The picture was prepared with the program Swiss-PdbViewer (Guex & Peitsch, 1997).

The human procathepsin S Cys25→Ala mutant was crystallized as a glycosylated protein. At Asn104p in the propeptide region, the electron density is characteristic for N-glycosylation but its quality does not allow to model-build the carbohydrate. The overall B-factor for the structure is 27.6 Å^2. In general, the B-factors for propeptides of all cathepsins are larger than those for their cognate enzyme domains due to the highly flexible C-terminal segment of the propeptide. The average B-factor for the propeptide of human procathepsin S mutant is 29.6 Å^2, only slightly higher than for mature cathepsin S, 27.0 Å^2. This can be additionally explained by the lack of information for the residues 108p – 113p in the C-terminus of the

propeptide where larger B-factors can be assumed. The human procathepsin S Cys25→Ala mutant will be subsequently referred to as procathepsin S for the reason of simplificity.

4.2.3.2 *The enzyme domains of procathepsin S*

The structure shows that the overall fold of the enzyme domains (mature cathepsin S) is similar to that of other cathepsins but reveals significant differences in the active site cavity. The greatest structural similarity between cathepsin S and other homologous cysteine proteinases with known three dimensional structures is seen with the enzyme domains of procathepsin K and L (PDB entries 1BY8 and 1CS8).
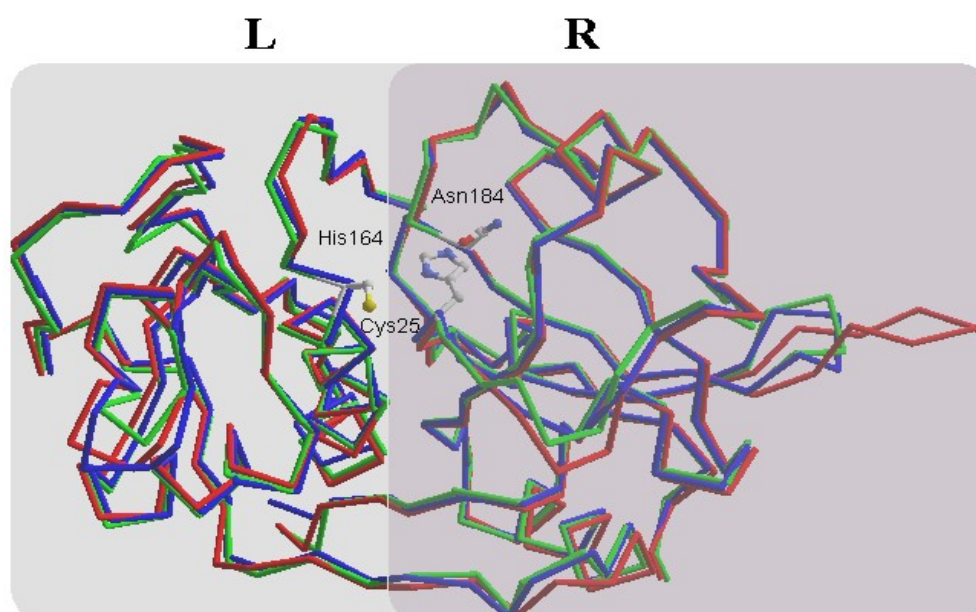
Figure 4-19 View of superimposed mature cathepsin structures.

The C$_\alpha$-trace of the enzyme domains of procathepsin S (this work) is coloured in blue, of procathepsin K (PDB entry 1BY8) in green, and of procathepsin L in red (PDB entry 1CS8). The amino acid residues of the catalytic triad of cathepsin S are shown. The R-domain and the L-domain are marked with R and L and underlaid by a transparent background. The picture was prepared with the program Swiss-PdbViewer (Guex & Peitsch, 1997).

Superposition of the structures of mature cathepsin S, K and L gives an RMSD of 0.61 Å with 211 and 205 Cα atoms as target pairs for both molecules (see Figure 4-19). The most evident backbone differences between the enzyme domains are confined to surface loops, and do not involve residues of the active-site cavity. Three disulfide bridges (Cys22 – Cys66, Cys56 – Cys99, Cys158 – Cys206) are formed in procathepsin S, which are equivalent to those

observed in in cathepsins L, K, porcine H (Guncar *et al*., 1998), B (Podobnik *et al*., 1997), V (Somoza *et al*., 2000) and papain (Kamphuis *et al*., 1984). The active-site cavity contains a catalytic triad consisting of Cys25 (substituted by an alanine in the procathepsin S mutant), His164 and Asn184, which are in nearly the same position as the active site residues of all known cysteine proteinase structures belonging to the cathepsins. The crystal structure of a cathepsin S Cys25→Ser mutant (Turkenburg *et al*., 2002) is virtually identical to the enzyme domains, mentioned above. The description of the active site, including the substrate-binding sites for substrate residues P1, P2, and P3 (Schechter & Berger, 1967) of the cathepsin S Cys25→Ser mutant is similar to what is found in the procathepsin S Cys25→Ala mutant. The comparison of both active sites gave no significant difference with regard to the orientation of the active-site residues (see Figure 4-20).
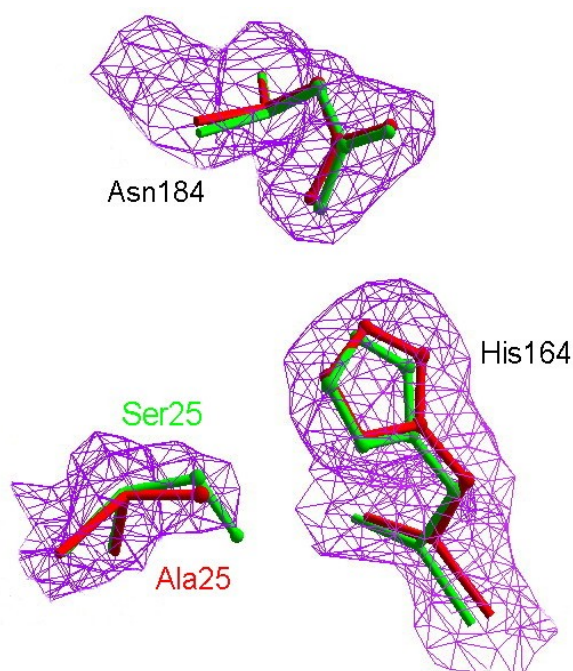
Figure 4-20

Active site residues of cathepsin S.

Superimposed amino-acid residues of the cathepsin S active-site. The residues of the procathepsin S Cys25→Ala mutant (this work) are coloured in red, the residues of the mature cathepsin S Cys25→Ser mutant (PDB entry 1GLO) in green. The $2F_o$-F_c electron density map for procathepsin S is contoured at the level of 1σ (magenta). The picture was prepared using the program Swiss-PdbViewer (Guex & Peitsch, 1997).

Only the orientation of some residues close to the substrate binding site differ due to interactions between the enzyme domains and the propeptide, which is missing in the crystal structure of the mature cathepsin S Cys25→Ser mutant. Structural details about the active site cavity and the substrate-binding sites are described under *4.2.3.6 - 8*.

4.2.3.3 Overall description of the propeptide

The N-terminal segment of the propeptide (Figure 4-18) forms a globular domain comprising three α-helices (α1p = 25p-35p, α2p = 44p-68p, α3p = 85p-92p) and a β-strand (β1p = 72p-76p). A 3_{10}-helix (3_{10}1p = 79p-82p) is found between β1p and α3p instead of a loop in the propeptide region of procathepsin K, L, B and procaricain. A different orientation of helix α 3p in comparison with the corresponding helices in the propeptides of K and L was found. The superimposed propeptide structures (see Figure 4-21) reveal that the α3p-helix in the propeptide of procathepsin S is positioned in nearly the same orientation as the corresponding helix in the propeptide of procathepsin B.

Figure 4-21

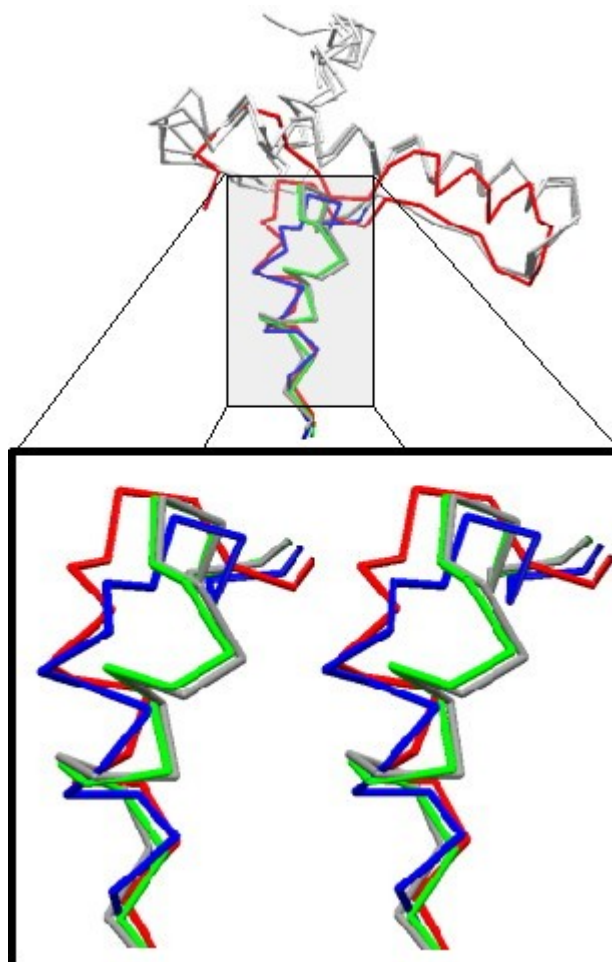Superimposed cathepsin propeptides.

Stereo picture of superimposed α3p-helices of cathepsin propeptides. S (blue) (this work), K (grey), L (green), and B (red) (PDB entries 1BY8, 1CS8 and 3PBH). The picture was prepared with the program Swiss-PdbViewer (Guex & Peitsch, 1997).

4.2.3.4 The globular domain of the propeptide

The helices α1p and α2p are connected by a loop of seven residues. They intersect at an angle of ~ 90° and are tightly packed against one another, thereby forming a hydrophobic core, which is comprised of the side chains of Leu24p, Trp28p, Trp31p, Ile51p, Trp52p, Leu56p, Met77p and the aliphatic part of Glu53p. The three aromatic rings of Trp28p, Trp31p and Trp52p are positioned perpendicular to each other and form the center of the hydrophobic core (Figure 4-24). Additional interactions between the two helices such as hydrogen bonds and salt bridges contribute to the stabilisation of the N-terminal globular domain (Table 4-7). A small number of these interactions are involved in the contact of the two helices to the α3p-helix. From the end of helix α2p, the chain turns four residues back, forms a hairpin structure and follows along this helix as an antiparallel β-strand (β1p = 72p-76p). The interface between helix α2p and strand β1p is formed by the hydrophobic portions of the side chains of Val59p, Asn63p, and Leu75p. Following the β-strand, the backbone forms a 3_{10}-helix (3_{10}1p = 79p-82p), which is capped by an N-terminal Asn-loop forming a hydrogen bond to the backbone of Leu80p (Asn79p^{OD1}···Leu 80p^{N} (2.79 Å)) (see Figure 4-22). This region seems to be composed of two type III ß-turns arranged in tandem. Such turns were first identified by Venkatachalam (1968) who found three types (I-III) each containing a hydrogen bond between the carbonyl oxygen of residue i and the amide nitrogen of i+3 as formed in the cathepsin S propeptide (Asn78p-Gly81p, Leu80p-Met83p, see Figure 4-23). Both type III ß-turns might be part of a 3_{10}-helix. By using the programs PROCHECK (Laskowski *et al*. 1993) and DSSP (Kabsch & Sander, 1983) the ϕ and ψ angles for the conserved residues His79p, Leu80p, Gly81p, and Asp82 were proven to be located in the sterically allowed regions for this helix. The 3_{10}-helix is not a common secondary structural element in proteins. Only 3.4% of all residues are involved in 3_{10}-helices in the Kabsch and Sander database (1983), and nearly all of these are located helical segments containing 1-3 hydrogen bonds. α-Helices sometimes begin or end with a single turn of a 3_{10}-helix (one hydrogen bond).
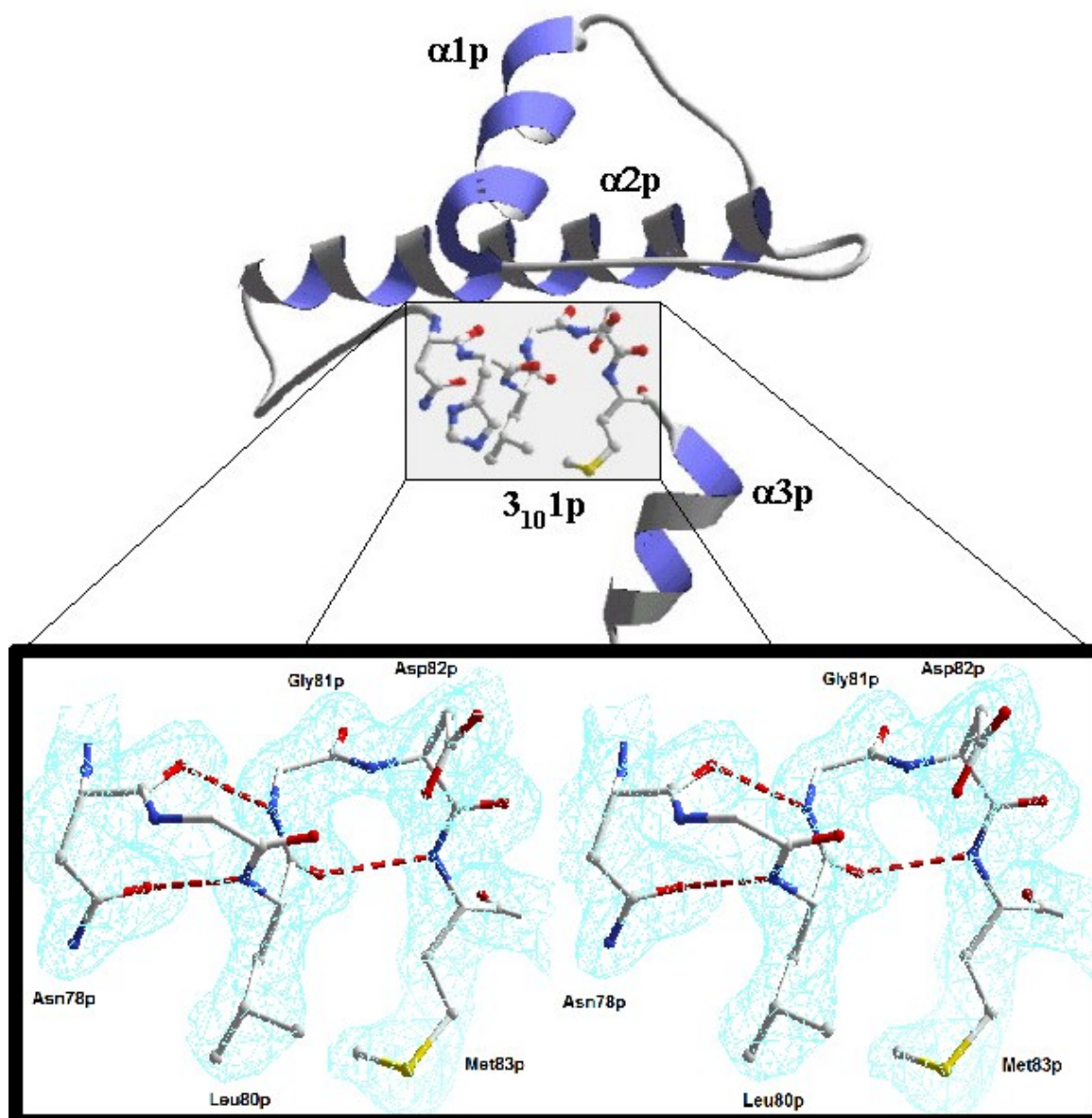
Figure 4-22 Stereo picture of the 3_{10}-helix ($3_{10}1p$) within the propeptide.

Residue His79p is not shown. The helix is stabilized by a N-terminal Asn-loop (hydrogen bond OE1^{Asn78p}··N^{Leu80p}). The $2F_o$-F_c electron density map (turqois) is contoured at a level of 1 σ. Hydrogen bonds are indicated as dashed lines.The atoms are coloured blue (N), red (O), grey (C), and yellow (S), respectively. The picture was prepared with the program Swiss-PdbViewer (Guex & Peitsch, 1997).

The protein sequence of the proregion between Asn78p at the N-terminus of helix 3_{10}1p and Val88p within helix α3p is similar to the corresponding proregions of cathepsin K and L. Helix α3p of the cathepsin S propeptide is rotated 20° counter-clockwise compared to the helices of K and L, and is rotated 10° clockwise compared to the corresponding helix in the cathepsin B proregion (Figure 4-23). In contrast to the corresponding propeptide helices of procathepsin K, L, and B, the differently oriented helix α3p seems to be extended by the residues Ser93p and Ser94p. The different orientation of helix α3p can be described by a rotation along the helix axis towards the L-domain. Beyond Ser94p, the propeptide emerges from the surface of the protein and follows along the groove through the two domains of cathepsin S.

Figure 4-23 Stereo picture of superimposed α3p helices of cathepsin S-, K-, and B-propeptides (PDB entries 1BY8, 3PBH). The residues of the cathepsin S α3p helix are coloured in grey, the helices of B (black) and K (red) are drawn as C_α backbones. The $2F_o$-F_c electron density map is contoured at a level of 1σ (tourquois). The picture was prepared with the program Swiss-PdbViewer (Guex & Peitsch, 1997).

4.2.3.5 Stabilization of the globular propeptide domain

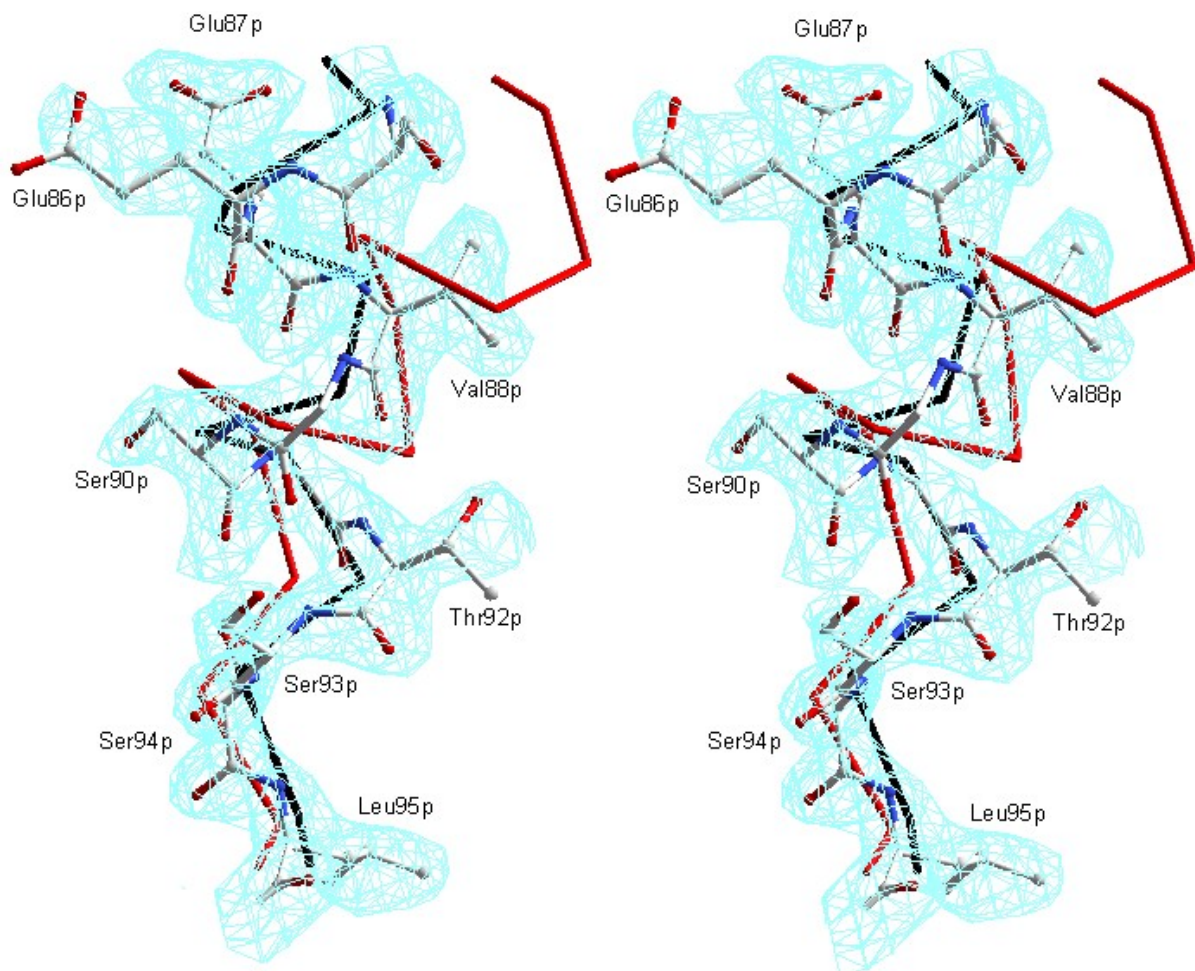The stabilization of the globular domain of L-cathepsin propeptides and the role of conserved residues contributing to the stability by hydrogen bonds and salt bridges have been discussed extensively in previous structural studies (Coulombe *et al*. 1996; Groves *et al*. 1996, Sivaraman *et al*. 1999). Especially electrostatic interactions (salt bridges) are likely to play a significant role in the stabilization of the globular structure (see Table 4-7). In procathepsin S, several clusters of salt bridges are formed within the helical parts of the propeptide. Glu44p and Arg48p, which contribute to the formation of such interactions are belonging to the conserved ER(F/W)N(I/V)N-motif (Karrer *et al*., 1993) and are located in the second propeptide α-helix (α2p) (see Figure 4-24).

Table 4-7 Stabilizing salt bridges and hydrogen bonds within the propeptide

Atom	Distance (Å)	Atom	Distance (Å)	Atom	Distance (Å)	Atom
		$Asp21p^N$	2.7	$Glu53p^{OE1}$		
		$Asp21p^O$	2.8	$Arg49p^{NH2}$		
		$Asp25p^{OD1}$	2.8	$Arg49p^{NH1}$		
		$Trp28p^{NE1}$	2.8	$Glu45p^{OE2}$		
		$Lys32p^{NZ}$	2.9	$Glu45p^{OE2}$		
$Tyr39p^{OH}$	2.6	$Asp82p^{OD1}$	3.4	$Lys37p^{NZ}$	3.2	$Asp82p^{OD2}$
		$Arg48p^{NH2}$	3.2	$Glu44p^{OE1}$	3.0	$Arg48p^{NE}$
$Arg49p^{NE}$	2.9	$Glu53p^{OE1}$	3.1	$Arg49p^{NH2}$		
		$Asn55p^{ND2}$	2.9	$Gly81p^O$		
		$Asn63p^{OD1}$	2.9	$Leu75p^N$		
		$Asn63p^{ND2}$	2.8	$Leu75p^O$		
		$Glu65p^{OE1}$	2.6	$His71p^{NE2}$		
		$His79p^O$	2.8	$Lys37p^{NZ}$		
		$Glu87p^{OE1}$	2.8	$Arg48p^{NH1}$		
		$Glu87p^{OE1}$	2.8	$Arg48p^{NH2}$		

Conserved residues belonging to the ER(F/W)N(I/V)N -motif are labeled in blue.
Salt bridges are underlaid in red.

Instead of Glu44p and Val59p, an aspartate (Asp27p) and an isoleucine (Ile42p) are located at the same position in the propeptides of cathepsins K and L. Apart from salt bridges and hydrogen bonds, hydrophobic interactions contribute greatly to the stabilization of the globular propeptide domain. The conserved amino-acid residues within the globular

propeptide domain were interesting targets for mutation studies in which some of these residues were replaced by amino acids with small side chains such as alanines (Kreusch *et al*., 2000, Schilling *et al*. 2001, Pietschmann *et al*., 2002). The consequences of the mutations on folding, transport, targeting and maturation of the cathepsin precursor are discussed below in view of the location and orientation of the mutated residues (see ***Discussion***).
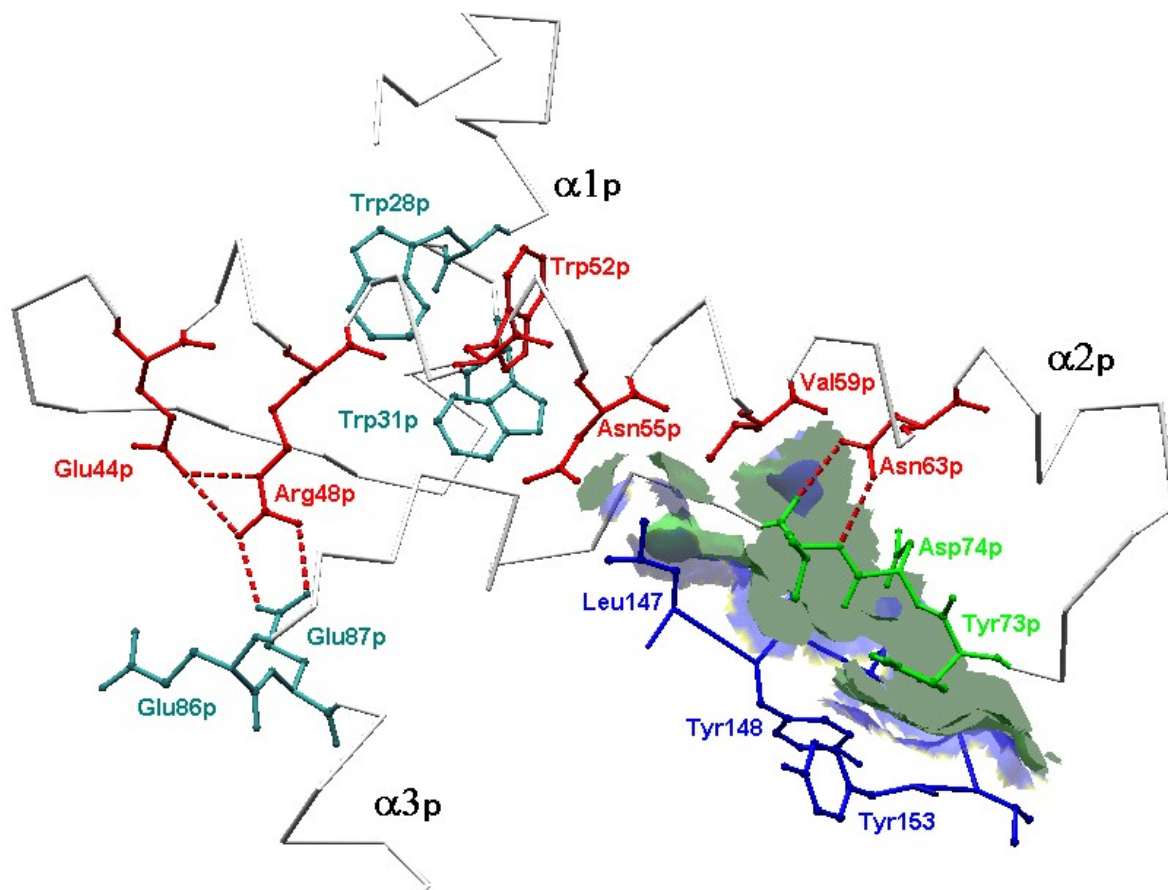
Figure 4-24 View of the conserved residues within the propeptide.

Overview of the conserved residues involved in stabilization of the globular domain of the cathepsin S propeptide. Residues belonging to the ER(F/W)N(I/V)N motif are labeled in red. Dashed lines mark hydrogen bonds. The residues of the PBL (propeptide binding loop) are coloured in blue, and residues of the propeptide interacting with the PBL are coloured in green. The calculation of the contact surface between the bottom side of the propeptide (green) and the upper side of the PBL (blue) as well as the picture preparation were carried out with the program Swiss-PdbViewer (Guex & Peitsch, 1997).

4.2.3.6 Anchoring of the propeptide onto the enzyme domains

The propeptide of cathepsin S contacts with the cognate enzyme via an extended area which can be subdivided into three adjacent areas: a large omega loop on the surface of the enzyme, referred to as the propeptide binding loop (PBL), a hydrophobic groove leading to the S´ subsites of the active site and the S1-S3 sites of the substrate-binding site. Along these contact areas the propeptide embraces the enzyme like a clamp and thereby covers the active site in a direction reverse to the substrate.

The PBL (Figure 4-24 and 4-25) of cathepsin S consists of the residues Arg141 - Tyr153, and the main contact between the propeptide and the PBL involves the residues of the β1p strand, forming a two stranded β-sheet with Arg149 - Val152. Additional hydrophobic and van der Waals contacts are formed between Phe58p, Val59p, Tyr73p and Leu75p and a cluster of hydrophobic amino acid residues within the PBL, including Leu147, Tyr148 and Tyr153. This cluster is extended by aromatic residues, such as Phe145, Phe190, Trp186 and His164, with the latter two being part of the active site. With the exception of some residues, a similar cluster arrangement can be found in procathepsins S, L and K.

Four side chains from the propeptide helix $3_{10}1p$ and helix α3p, namely Asn78p, Leu80p, Met83p and Leu91p (α3p), are sticking into the vicinity of the active site and extend the hydrophobic cluster of the S´-site groove.

The side chain of Met83p points towards the side chain of Phe146. A C-H···π interaction between Met83p^{CE} as H-donor and the aromatic ring (π-system) of Phe146 as acceptor (Met83p^{CE} - Phe146^{π} distance = 3.88 Å and ρMet83p^{CEH} - Phe146^{πx} = 125.92°) (see Figure 4-25) could be identified by using programs written by M. Brandl (Brandl *et al*., 2001) as well as by D. Pal (see ***Methods***). Such interactions, which were first discovered in proteins in the mid 1990s, (Derewenda *et al*., 1995) can be described as weakly polar interactions (Burley & Petsko, 1988), but they can also be classified as hydrogen bonds, although they are considerably weaker than `classical H-bonds´ (Weiss *et al*., 2001). Their functional role in folding and stabilization of proteins remains unclear, but investigations in this direction seem to be useful because of the frequent occurence of these interactions in proteins (Weiss *et al*., 2001). The interaction between Met83p and Phe146 found in procathepsin S is unique within the L-like cathepsin family. A superposition of the propeptides of cathepsins S, K and L revealed that the Met83p side chain gets close to the position of the side chains of Phe71p

(helix α3p of the cathepsin L propeptide) and Val71p (helix α3p of the cathepsin K propeptide), which also extend into the S´-groove (Figure 4-26). The equivalent methionines in the propeptides of cathepsin K and L are therefore directed rather to the outer surface of the propeptides. The position of Phe146 is occupied by a leucine in cathepsin L and a glutamine in cathepsin K. The superposition of the structures of the procathepsin S Cys25→Ala mutant (this work) and of the mature cathepsin S Cys25→Ser mutant (PDB entry 1GLO, Turkenburg *et al*. 2002) reveales a different orientation of Phe146. In the structure of the cathepsin S Cys25→Ser mutant, Phe146 is directed towards the solvent. In the structure of the procathepsin S Cys25→Ala mutant, the aromatic ring of Phe146 points towards the active-site cavity interacting with Met83p of the propeptide (see Figure 4-25).
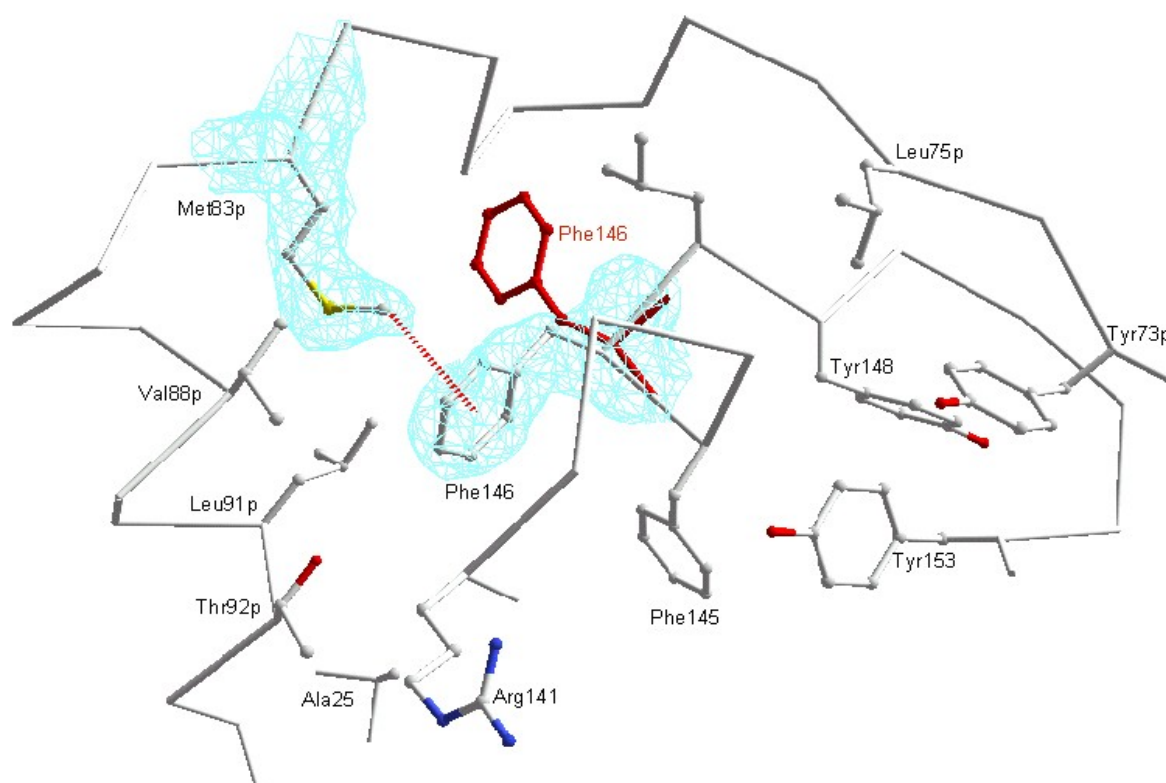
Figure 4-25 View of the interactions between cathepsin S propeptide and PBL (propeptide binding loop). PBL: Arg141-Tyr153. Labels of residues belonging to the propeptide are marked with "p". Phe146 of the superimposed mature cathepsin S Cys25→Ser mutant (PDB entry 1GLO, Turkenburg *et al*., 2002) is coloured and labeled in red. Atoms are shown in grey (C), red (O), blue (N), and yellow (S). The dashed line marks a C-H···π interaction. The $2F_o$-F_c electron density for Met83p and Phe146 is contoured at a level of 1σ (coloured in turquois). The picture was prepared with the program Swiss-PdbViewer (Guex & Peitsch, 1997).
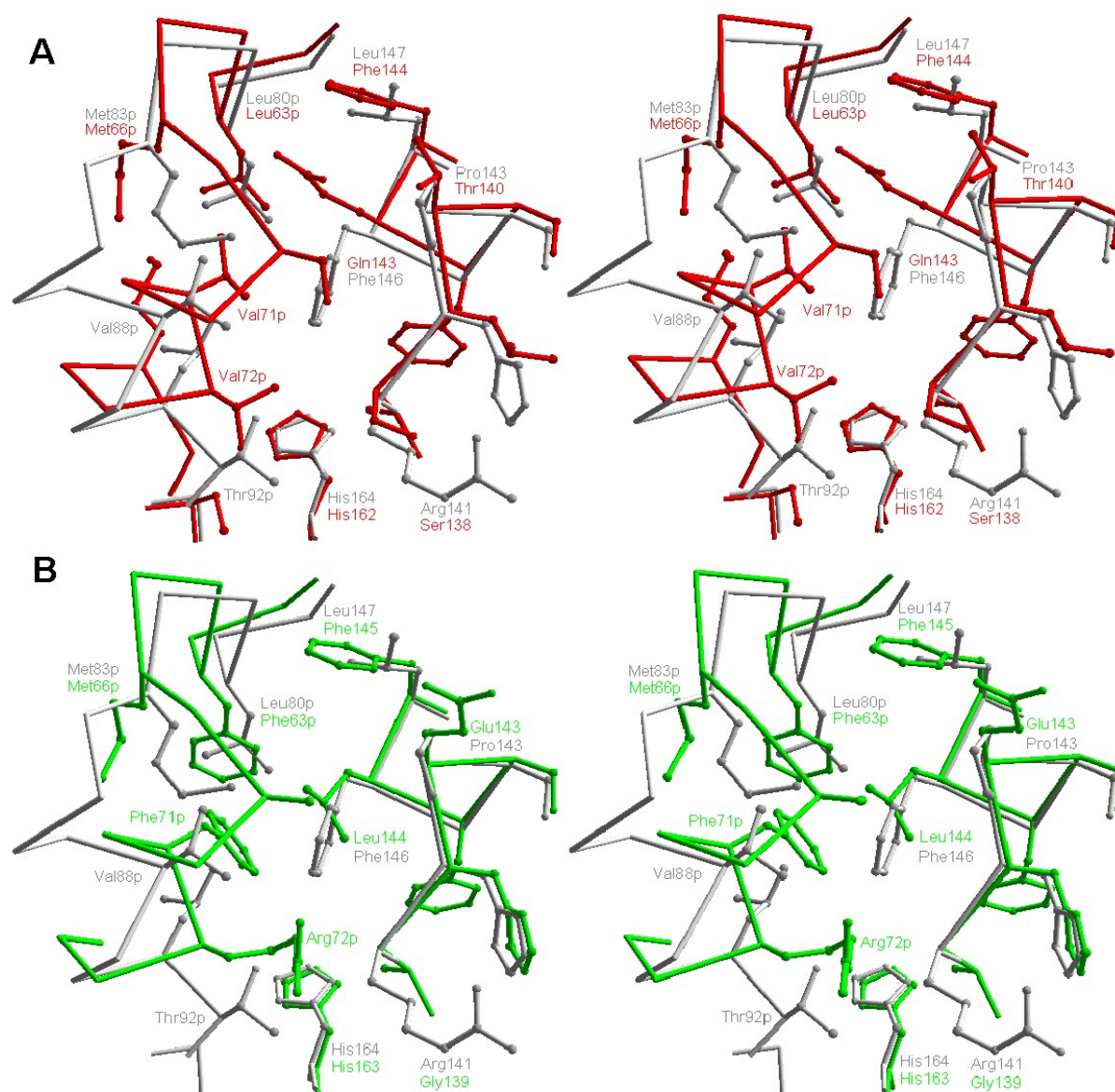
Figure 4-26 Superimposed cathepsin propeptides around the active-site.

View of the different oriented propeptides of cathepsins S (grey; this work), K (red; PDB entry 1BY8), and L (green; PDB entry 1CS8) around the active-site and, partially, the PBL. PBL: cathepsin S: Arg141-Leu147, cathepsin K: Ser138-Phe144, cathepsin L: Gly139-Phe145.
Stereo picture **A** shows the superimposed α3p helices of the cathepsin S- and K propeptides, PBL and the active site histidine.
Stereo picture **B** shows the same parts of the procathepsin S structure superimposed with the corresponding residues of procathepsin L.
The pictures were prepared with the program Swiss-PdbViewer (Guex & Peitsch, 1997).

4.2.3.7 Interactions between helix α3p and the S´- subsites

Regarding the α3p helices of the superimposed procathepsins, the helix of procathepsin S adopts a position intermediate between the corresponding helices of procathepsins B, L and K. In procathepsin S, this leads to larger distances between residues of the hydrophobic face of helix α3p and the S´subsites of cathepsin S than found in procathepsins L and K. Helix α3p (Figure 4-26 and 4-27) interacts with the hydrophobic surface of the hydrophobic S´-groove via the residues Val88p, Ser90p, Leu91p and Thr92p, the side chains of which point to the PBL (Arg141 - Phe146p) as well as to the region between Gln19 and Ala24 of the L-domain. In particular, Leu91p is extending into the centre of the hydrophobic groove and forms van der Waals contacts to Trp186 and Met83p. The side chains of Val88p and Thr92p line the right side of the helix, which is directed towards the PBL (see Figure 4-28). They extend nearly parallel to each other towards the residues Arg141 - Phe146 of the PBL and form a van der Waals contact with the hydrophobic part of Arg141 and Phe146, respectively. On the left side of the helix, the side chain of Ser90p points towards the L-domain, forming a van der Waals contact with Gly23. A cavity was detected between the PBL and the right side of the last helix turn, below Leu91p and Thr92p. This cavity is positioned in the S´1 pocket, which comprised by Ala140, Arg141, Trp186 and the active site residue His164, and is filled with one water molecule (Figure 4-27). This water molecule is positioned between Thr92p (α3p), Arg141 (S´1) and Asn163, interacting with these residues by hydrogen bonds. In the structures of procathepsins K and L no water molecules were detected between the equivalent propeptide helix α3p and the PBL (LaLonde *et al.*, 1999, Coulombe *et al.*, 1996). In the center of the active site, between Gln19, His164 and Ser94p (α3p), a second water molecule was detected, forming hydrogen bonds between active-site residues and the propeptide. In the C-terminal region of the propeptide, some residues interact with aromatic residues of the enzyme domains leading to a close contact of this flexible region with the surface of the mature enzyme molecule. These interactions are described below.

4.2.3.8 Interactions within the S subsites

The propeptides of cathepsin K and L follow the S subsites in an extended conformation, whereby Gly77p is located closest to the active site cysteine. The propeptide of cathepsin S forms an additional loop in this region, elongating helix α3p (Figure 4-27). At the end of this loop, Ser94p is oriented toward the mutated Cys25→Ala, forming a weak hydrogen bond (Ser94p^{OG}EEE Ala25^{N} (3.4 Å)). In comparison with Gly77p of the cathepsin K- and L

propeptides the distance of Ser94p to the active site residues is somewhat larger. Glycine allows for a close approach to the active-site nucleophile, and thus permittes these propeptide regions to reach the bottom of the cleft. The water molecule between the C-terminal part of the α3p helix and the active-site residues of procathepsin S (see above) may additionally influence the distance between the helix and the active site. Furthermore, the side chain of Ser94p is in contact with the oxyanion hole, forming a very weak hydrogen bond (3.7 Å) to N^{E1} of Gln19. The previously described S subsites (McGrath *et al.*, 1998; Turkenburg *et al.*, 2002) are occupied by residues C-terminal to helix α3p. Located in the C-terminal loop below the helix, Ser94p occupies the S1 pocket, which is comprised of the main-chain atoms of Asn64, Gly65, Cys66, Asn67, Gly68 and Gly23. These residues form van der Waals contacts to the side- and main chain atoms of Ser93p and Ser94p. Leu95p is extending into the S2 pocket, which comprises Met71, Gly137, Val138, Gly165, Val167 and Phe211.
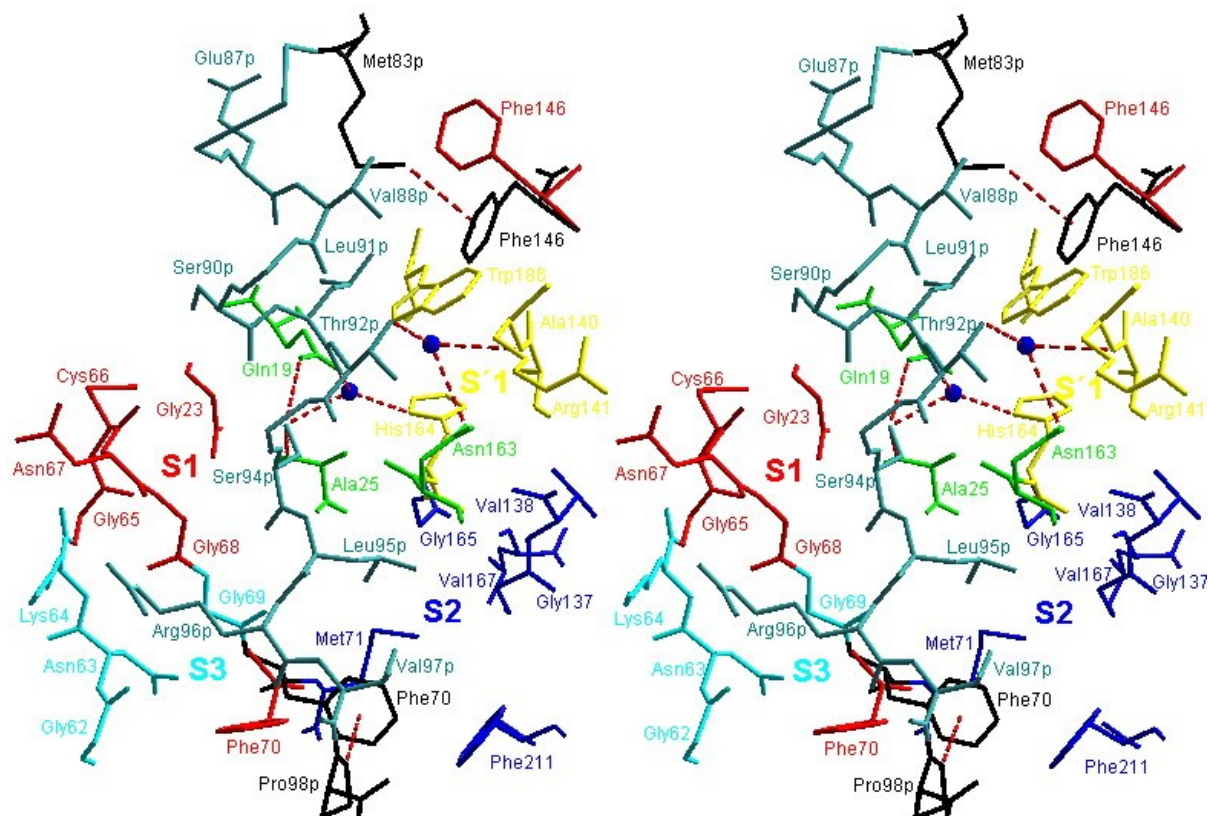
Figure 4-27   Propeptide of procathepsin S in the vicinity of the active site.

Propeptide of procathepsin S in the vicinity of the active site. Residues belonging to S1′ are shown in yellow, to S1 in red, to S2 in blue, and to S3 in green. Atoms of the propeptide are displayed in grey. Two water molecules in the active site cavity are shown in blue. Hydrogen bonds are marked as dashed lines (red). Phe146 and Phe70 of the superimposed mature cathepsin S Cys25→Ser mutant (PDB entry 1GLO, Turkenburg *et al.*, 2002) are colored and labeled in black. Residues involved in C-H⋯π interactions are colored and labeled in black. Labels of residues belonging to the propeptide are marked with "p". Hydrogen bonds and C-H⋯π interactions are marked as dashed lines (red). The stereo picture was prepared with the program Swiss-PdbViewer (Guex & Peitsch, 1997).

A leucine is also preferred in the S2 subsite of procathepsin K, and in procathepsin L a phenylalanine occupies the same subsite. Mainly large side chains, such as those of glutamine (procathepsin L), lysine (procathepsin K) or, especially in procathepsin S, Arg96p extend into the S3 pocket, which in cathepsin S comprises Gly62, Asn63, Lys64, Gly68, Gly69, and Phe70. A cathepsin S - specific loop containing the residues 61 - 64 shapes the back of the S3 pocket and limit the subsite to a smaller size than formed in cathepsin K. Below the S3 pocket, Pro98p of the propeptide forms a C-H⋯π bond with Phe70 (Pro98p^{CD} as H-donator and the aromatic ring of Phe70 as acceptor; Pro98p^{CD} – Phe70$^{\pi}$ distance = 3.64 Å and ρ Pro98p^{CDH} - Phe70$^{\pi x}$ = 122.93°); identified with programs written by M. Brandl and D. Pal. A superposition of the mature cathepsin S Cys25→Ser mutant (PDB entry 1GLO, Turkenburg *et al*. 2002) with the procathepsin S Cys25→Ala mutant revealed an orientation of the corresponding Phe70 away from the domain towards the solvent (see Figure 4-28).
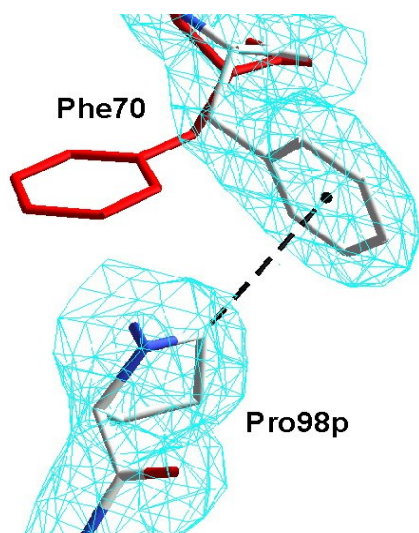
Figure 4-28 C-H⋯π interaction between Pro98p and Phe70.

View of the propeptide residue Pro98p interacting with Phe70 of the enzyme domain. Labels of residues belonging to the propeptide are marked with "p".
Phe70 of the superimposed mature cathepsin S Cys25→Ser mutant (PDB entry 1GLO, Turkenburg *et al*., 2002) is coloured and labeled in red. Atoms are shown in grey (C), red (O), and blue (N). The dashed line marks a C-H⋯π interaction. The 2F$_o$-F$_c$ electron density is contoured at a level of 1σ (turquois). The picture was prepared with the program Swiss-PdbViewer (Guex & Peitsch, 1997).

4.2.3.9 Electrostatic potential

The pI for the human procathepsin S Cys25→Ala mutant is 8.43. The mature cathepsin S molecule is nearly neutrally charged with a pI of 7.64, and the pI of the cathepsin S propeptide is 9.26, calculated for the protein sequence without the signal peptide (see **Methods**). In order to compare the interactions between enzyme domains and propeptides or substrates, the focus was on a comparison of the electrostatic potentials of the cathepsin enzyme domains and the propeptides. Based on the protein sequences of known cathepsin structures including cathepsin S and papain, the pIs of their propeptides were calculated: they all range between 9 and 9.8 (see Table 4-8). Only the calculated pIs for the propeptides of papain (4.62) and cathepsin K (6.32) differ strikingly from the values found for the other cathepsin propeptides. An assessment of the electrostatic potential of all cathepsin propeptides

with regard to function is difficult, because the propeptides differ in length and structure. Furthermore, the structures of several precursors, such as of human cathepsin H, F, and V, are not known.

Enzyme	pI $_{mature}$	pI $_{propeptide}$
Papain	8.88	4.62
Cathepsin S	7.64	9.26
Cathepsin K	8.29	6.32
Cathepsin L	4.43	9.00
Cathepsin V	8.59	9.56
Cathepsin B	5.23	9.56
Cathepsin H	5.93	9.76
Cathepsin F	5.82	9.43
Cathepsin X	5.48	9.77

Table 4-8 Isoelectric points of cathepsins.

The isoelectric points (pI) were calculated independently for each protein sequence of the mature enzyme and for the propeptide without the N-terminal signal peptide sequence. The calculations were carried out with COMPUTE pI/MW (Wilkins *et al.*, 1998). The L-like cathepsins are coloured in red.

The structural similarity between the propeptides of the L-like cathepsins L, K, and S allows a correlation of the structure and electrostatic potential on their surfaces. This might elucidate aspects such as pH-dependent folding of the propeptides and their selectivity towards the cognate enzymes. Figure 4-29 shows the different electrostatic potentials for the molecular surfaces of the cathepsin K, L, and S propeptides. In comparison with the propeptides of cathepsin S and L, the cathepsin K propeptide shows more negatively charged areas on the surface, especially on the top of helix α1p, furthermore along the hairpin structure at the end of helix α2p, and of helix α3p facing the PBL. On the surfaces of the cathepsin S and L propeptides more positively charged regions are detected. Figure 4-29 shows the different electrostatic potentials for the molecular surfaces of the cathepsin K, L, and S propeptides. In comparison with the propeptides of cathepsin S and L, the cathepsin K propeptide shows more negatively charged areas on the surface, especially on the top of helix α1p, furthermore along the hairpin structure at the end of helix α2p, and of helix α3p facing the PBL. On the surface of the cathepsin L propeptide the positive charges are found mainly on the surface of helix α3p and the region following C-terminal. The inner surface of these propeptides faces the active site cavity and the PBL is more or less neutraly charged due to the hydrophobic residues located there. On the surface of the cathepsin L propeptide the positive charges are found mainly on the surface of helix α3p and the region following C-terminal. The inner

surface of these propeptides faces the active site cavity and the PBL is more or less neutraly charged due to the hydrophobic residues located there. The comparison of the calculated pI´s for the mature cathepsin molecules and papain gave a broad range between 4.4 and 8.9 (see Table 4-8). Different pIs are even found within the L-type cathepsin subfamily.

Figure 4-29

Electrostatic surface potentials of cathepsin propeptides.

The electrostatic potentials on the surface of the cathepsin propeptides S (this work), K (PDB entry 1BY8) and L (PDB entry 1CS8) were calculated with Swiss-PdbViewer(Guex & Peitsch, 1997) using the atomic partial charges for each propeptide. The colour of the electrostatic potential maps ranges from red (-8 kT/mol) via white (neutral) to blue (+8 kT/mol). The α-helices are labeled in yellow (1-3p).

Figure 4-30 (Page 75) The electrostatic surface potentials of cathepsins.

The propeptides of cathepsins S, K, L, B, and X are shown as Cα traces. All shown molecules (PDB entries in brackets) are in the same orientation with view on the active-site cavity and the contact area of the mature enzyme and the propeptide. The electrostatic surface potentials of the cathepsins were calculated with Swiss-PdbViewer (Guex & Peitsch, 1997) using the atomic partial charges for each mature enzyme molecule. The colour of the electrostatic potential maps ranges from red (-8 kT/mol) via white (neutral) to blue (+8 kT/mol).
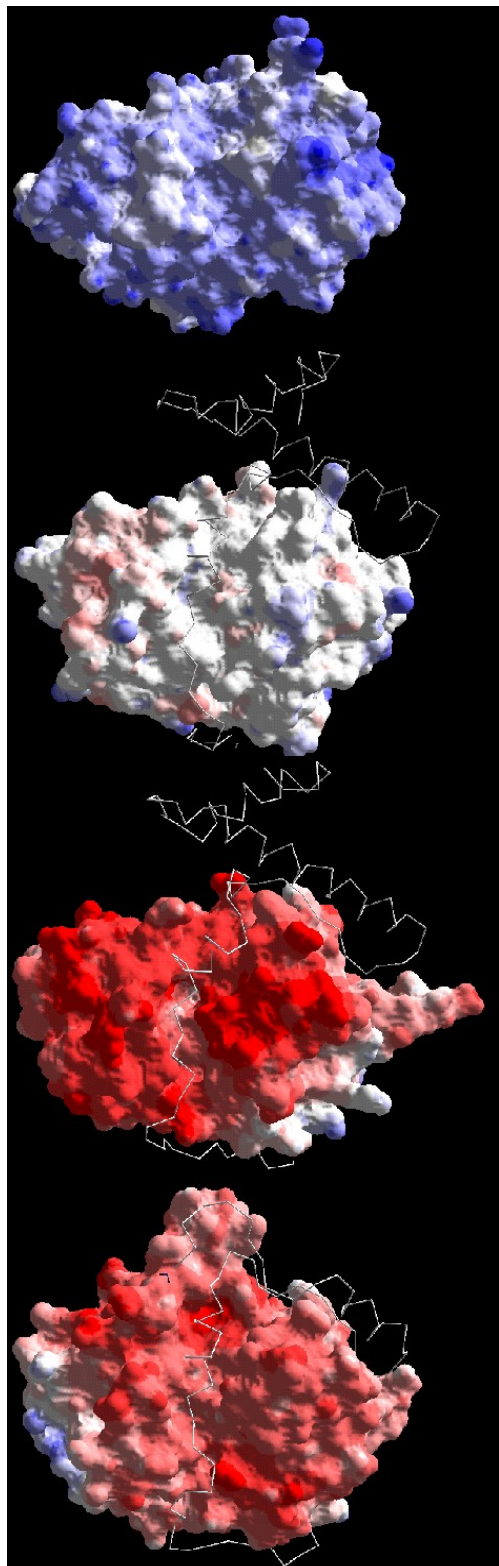
Papain
← (9PAP)

Figure 4-30

Legend: page 74.

Procathepsin S
←(this work)

Procathepsin K
(1BY8) →

Procathepsin L
← (1CS8)

Cathepsin V
(1HFO) →

Procathepsin B
← (3PBH)

Procathepsin X
(1DEU) →
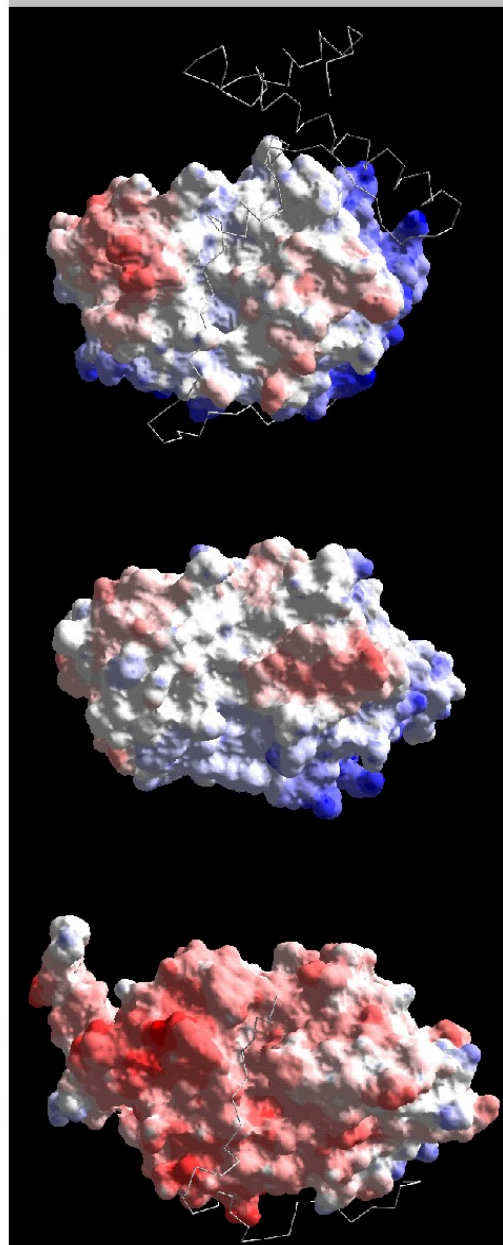
The enzyme domains of cathepsin K, V, S, and papain show a significantly higher pI than those of cathepsins L, B, H, F and X, which predicts larger or more regions of negative charges on the molecule surfaces of the latter cathepsins (see Figure 4-30). In contrast to cathepsin S, the potentials of cathepsin L, B, and X are negative over extended regions (see Figure 4-30). Especially within the active-site cavity and on the surface around this cavity the negative charges are concentrated, whereas the electrostatic potential map of papain is positive, which is also expressed by the pI of 8.88. The molecules of cathepsin K and V show an interesting subdivision of the potentials. Around the active-site cavity patches of negative potentials are observed. However, the backsides of these molecules show large positively charged areas, which explain the calculated pI of 8.92 for K and a pI of 8.59 for V. With the exception of cathepsin L, B, and X the active-site cavities and the propeptide-enzyme contact areas of the cathepins are more neutrally charged, indicating the hydrophobic nature of the propeptide-enzyme domain interactions. With regard to their functional role, the electrostatic potentials are discussed below (see ***Discussion***).

5.   Discussion

5.1   Crystallization as implication for an independently folded structure of the cathepsin S propeptide

5.1.1   A novel crystallization approach

For the first time, a poorly soluble propeptide of a lysosomal cysteine proteinase has been crystallized. Only one other example of a crystallized propeptide had been found earlier, the so-called pro-region of a bacterial serine proteinase called α-lytic protease (Sauter *et al.*, 1998). The strategy used for the crystallization of the cathepsin S propeptide consists of the combination of a conventional method such as vapor diffusion with the use of a crystallization buffer that allows a pH shift from acidic to nearly neutral conditions and leads to folding of the cathepsin S propeptides and to the growth of crystals. Figure 5-1 shows the strategy. The growth of crystals can be explained in the following way: After equlibration of both buffer compounds within the droplet, formation of precipitate can be observed due to the aggregation of partially or completely folded propeptide molecules. At this stage, only a limited number of folded molecules are solubilized in a folded state.
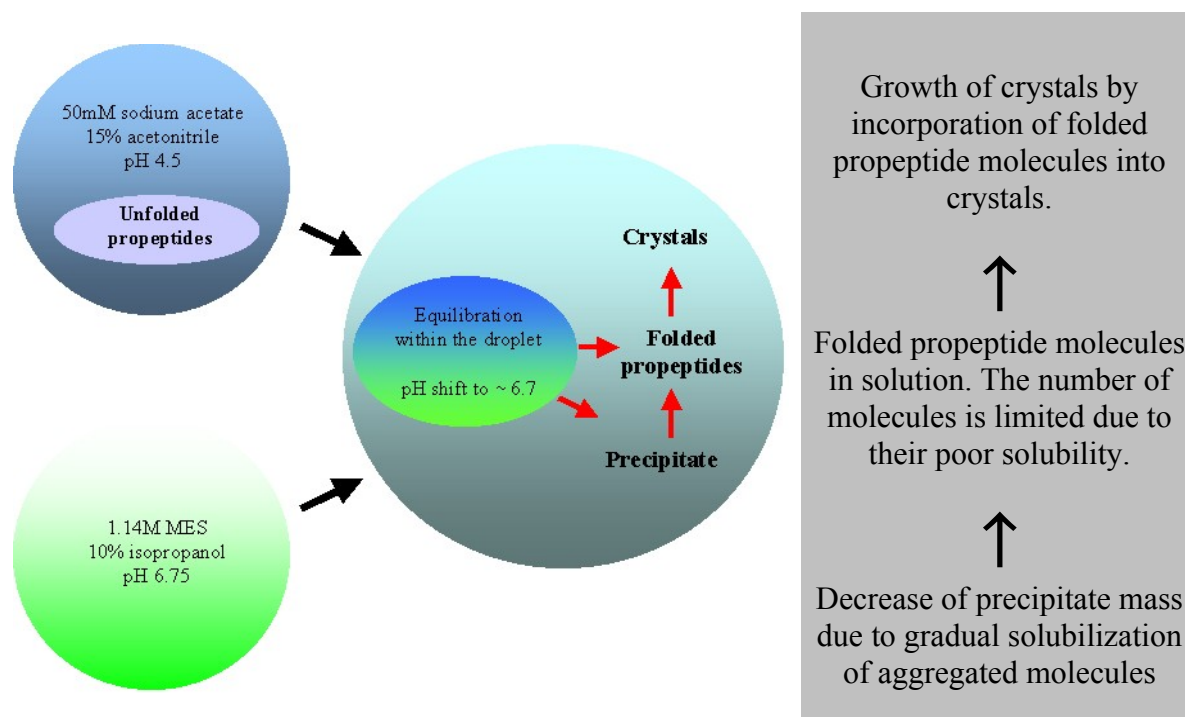

Figure 5-1   Crystallization strategy for the cathepsin S propeptide

The concentration of the droplet components including the protein, may increases with time because water diffuses out of the droplet. A possible transfer of isopropanol from the reservoir to the droplet may occur as well. The folded propeptide molecules begin to aggregate and to order, leading to nucleation of crystals. The growth of crystals depends on a pool of folded and solubilized molecules in the droplet. This pool seems to be maintained by molecules dissolved from the precipitate, which decreases in mass with time. A balance between solubilized molecules, aggregated molecules in the form of amorphous precipitate, and the crystals within the droplet can be assumed. The crystals incorporate folded molecules, and the molecules removed from the solution are substituted by molecules from the precipitate. This raises the question of how different factors like protein properties, buffer compounds, pH, and temperature influence the crystal growth:

- **The protein:** The cathepsin S propeptide is largely of hydrophobic character. The elevation of salt concentration or ionic strength resulted in no improvement of the solubility at neutral pH, at which the propeptide is folded. The solubility of the propeptide in an acidic buffer is enhanced in the presence of the solubilizing agent acetonitrile, but the protein is not or only partially folded. This way, a high protein concentration could be achieved, which is necessary for a successful crystallization in the hanging drop or in the capillary (counter diffusion technique). The cathepsin S propeptide folds spontaneously via pH shift from acidic to neutral (Maubach *et al.*, 1997). The folding process is complete within one second, as determined by stopped-flow measurements (Schilling *et al.*, 2001).

- **The crystallization buffer:** The only buffer to allow the crystallization is 2-(N-morpholino) ethanesulfonic acid (known as MES), which has a pK_a of 6.1. These buffer molecules are zwitterions at low pH. For the growth of diffracting crystals a high concentration of MES is necessary. At concentrations below 1 M MES, a high number of microcrystals was observed in the droplets. The chance to grow only a few crystals of larger size increases with MES concentrations higher than 1 M, and the best results were obtained at 1.14 M. The amphoteric character of MES may support the ordering of the propeptide molecules in the droplet with time.

- **The organic solvent:** Without any organic solvent such as isopropanol, no crystals were observed in the droplets. The best diffracting crystals were obtained in the presence of 10% isopropanol. A screen below 10% yielded only microcrystals within the droplets. The

elevation of the isopropanol concentration gave no improvement concerning shape and size of the grown crystals. It was also possible to grow crystals in the presence of polyethylene glycol, but no improvement of crystal growth could be achieved. More important seems to be the concentration of the added organic solvent, which influences the nucleation of the crystals.

- **pH:** The folding of the propeptide is a spontaneous process, which takes place directly after the equlibration of the protein solution with the crystallization buffer. The pH of 6.7, which was measured after equlibration, is necessary for such a folding process. The use of a MES buffer with a pH below 6.75 gave no crystal growth.

- **The temperature:** The velocity of crystal growth is determined by the temperature. Crystal growth was faster at 30°C than at 20°C. However, even when the temperature was elevated to 30°C the crystal growth was a very slow process. Generally, several months passed before the crystals reached a suitable size for diffraction measurements. It seems that the higher temperature facilitates the folding process, and increases the solubility of the protein, which are the limiting factors in the crystallization of the cathepsin S propeptide.

5.1.2 Ostwald ripening

Many small crystals form in a system initially but slowly disappear except for a few that grow larger, at the expense of the small crystals. The smaller crystals act as "nutrients" for the bigger crystals. As the larger crystals grow, the area around them is depleted of smaller crystals. This process is termed "Ostwald ripening" (Ostwald, 1896, 1897). This is a spontaneous process that occurs because larger crystals are more energetically favoured than smaller crystals and is observed in some macromolecular systems (Giegé *et al*., 1996). While the formation of many small crystals is *kinetically* favoured, (i.e. they nucleate more easily) large crystals are *thermodynamically* favoured. Thus, from a standpoint of kinetics, it is easier to nucleate many small crystals. However, small crystals have a larger surface area to volume ratio than large crystals. Molecules on the surface are energetically less stable than the ones already well ordered and packed in the interior. Large crystals, with their greater volume to surface area ratio, represent a lower energy state. Thus, many small crystals will attain a lower energy state if transformed into large crystals and this is what can be observed in Ostwald

ripening. This does not happen all the time, and one reason is that the nucleation of many small crystals reduces the amount of supersaturation and thus, the thermodynamically favored large crystals never get a chance to appear.

- In the droplets containing the cathepsin S propeptide crystals, a disappearance of small crystals followed by growth of larger crystals was never observed. But in all probability, the cathepsin S propeptide crystals were grown at the expense of the precipitate observed within the droplets. Within the droplets containing cathepsin S propeptide crystals, precipitate was observed, which appeared as crystalline in some cases. In this case, Ostwald ripening cannot be excluded because the requirement of crystals grown within the precipitate, is given for the growth of crystals at the expense of small crystals, as described by Giegé *et al*. (1996) for other macromolecules. Further experiments are required.

5.1.3 *Counter diffusion technique*

It is also possible to grow crystals using the counter diffusion approach (Garcia-Ruiz and Moreno, 1994), but no improvement concerning crystal size and growth rate was achieved. Very little is known about crystallization, the behavior of such poorly soluble proteins and about the influence of the chemical and physical parameters on the solubility of the protein. Phase diagrams (Miers & Isaac, 1907, Feigelson, 1988) may help to reveal the relationship between temperature, precipitant concentration and solubility. Furthermore, the solubility of the cathepsin S propeptide and its folding process might be tested in MES buffer using spectroscopic methods. The renaturation experiments of denatured cathepsin S in presence of its propeptide (Schilling *et al*. 2001, Pietschmann *et al*. 2002) could be carried out using MES as storing buffer for the propeptide or as renaturation buffer. These kind of experiments may help to understand the processes during the crystallization of the cathepsin S propeptide and the role of MES. The crystallization method described in this work, might be useful for the crystallization of other proteins and peptides of similar solubility properties.

5.1.4 How is the isolated cathepsin S propeptide structured ?

The successful crystallization of this small protein is a strong argument for an independent folding process to ordered structure. The results of the refolding experiments with the wild-type cathepsin S propeptide (Maubach *et al.*, 1997) and with its mutants (Kreusch *et al.*, 2000, Schilling *et al.*, 2001, Pietschmann *et al.*, 2002) point to a tertiary structure, which should be similar to the propeptide part of the determined procathepsin S structure. With high probability, the crystallized propeptide consists of the globular domain, which contains two crossed helices and a third helix behind the hairpin structure. The assumed high number of molecules in the asymmetric unit is an indicator for dimer or higher propeptide aggregates in the crystal. Further work is required to determine the structure. An optimization of the crystals is necessary. Another way to grow crystals of better quality may consists of truncation of the propeptide. Especially, the C-terminal part below helix $\alpha 3p$ might be an ideal region for a truncation (see Figure 4-16). This part is very flexible and may prevent the improvement of the crystals. Despite of the low diffraction resolution and the poor quality of the data collected, the successful crystallization is the first step to a prospective structure determination of the cathepsin S propeptide.

5.2   Procathepsin S structure and the functional role of the propeptide

5.2.1 Cathepsin S-specific interactions between the propeptide and the enzyme

The structure of human procathepsin S is quite similar to the structures of the previously solved L-like proenzymes. Nevertheless, some striking features are found. The cathepsin S-specific orientation of helix $\alpha 3p$ in the propeptide is the result of an interaction between Met83p of the loop region between helix $3_{10}1p$ and helix $\alpha 3p$ and Phe146 of the PBL. The cathepsin S-specific residues Phe146 and Arg141 are mentioned in the previously determined structures of mature cathepsin S (McGrath *et al.* 1998; Turkenburg *et al.* 2002), but their functional role could not be clarified because the enzymes were crystallized without any peptides as possible interacting partner. In the structure of human procathepsin S cathepsin S-specific interactions in form of a C-H$\cdots\pi$ bond between Phe146 and a residue of the propeptide can be observed. Furthermore, an interesting C-H$\cdots\pi$ bond can be seen in the C-terminal part below the S3 pocket between Phe70 and Pro98p (see Figures 4-27 and 4-28). As found for Phe146, the aromatic ring of Phe70 is flipped towards the propeptide, whereas the

same residues in the determined mature cathepsin S structure are directed away from the enzyme towards the solvent. Only in procathepsin K, a similar interaction was found. The superposition of all determined cathepsin K-inhibitor complexes with cathepsin S and their proenzymes revealed that the corresponding Tyr67 shows the same orientation as Phe70 in mature cathepsin S. In contrast to the mature enzyme, in procathepsin K the aromatic ring of Tyr67 is also flipped towards a proline (Pro81p) of the C-terminal part of the propeptide (see Figure 5-2). Between Pro81p and Tyr67, a similar C-H···π bond could be identified. The consequences of such interactions, which result in a different propeptide orientation with respect to the enzyme surface are discussed below in context to the functional of role of the propeptide.

5.2.2 *Inhibition by the propeptide and selectivity*

Many investigations were focused on the design of highly potent and selective proteinase inhibitors but up to now the selective inhibition of a proteinase is an unsolved challenge. In recent years, attention was directed to the inhibitory properties of propeptides of papain-like proteinases, especially of the L-type cathepsins. The high structural homology and the same mode of binding of the propeptides on the enzyme surface make this subfamily interesting for seeking selective differences. The precondition for a successful inhibition consists of a complex formation of the propeptide and the cognate enzyme, which is strongly pH-dependent, correlating with significant pH-dependent changes of the propeptide structure (Maubach *et al*. 1997, Jerala *et al*., 1998). The N-terminal globular domain of the propeptides is folded at physiological pH conditions. A pH shift to acidic conditions may lead to a break of stabilizing salt bridges and hydrophobic interactions within the propeptides, which is concomitant with a breakdown of the tertiary structure and the loss of the inhibitory capacity. The majority of structured determinants for recognition of the cognate enzyme surface are located in the globular domain of the propeptide. Previous *in vitro* experiments on the inhibitory specificity of the propeptides towards different members of the cathepsins revealed that the cathepsin S propeptide inhibits cathepsin L as much as its cognate enzyme cathepsin S. However, the inhibition of cathepsin S by the cathepsin L propeptide was only very weak (Maubach *et al*. 1997; Guay *et al*. 2000; Guo *et al*. 2000), which could recently be confirmed by Anne Schlabrakowski (personal communication by K. Schilling). In an initial study the cathepsin L propeptide gave a significant higher K_i for Cathepsin S (22.04 ± 0.43 nM) than for cathepsin L (0.082 ± 0.004 nM) and cathepsin K (1.74 ± 1.32 nM). However, no

significant differences of the K_i values were found for the inhibition of cathepsin K by the cathepsin S propeptide and the inhibition of cathepsin S by the cathepsin K propeptide.

5.2.2.1 The functional role of Arg141 and Phe146 in cathepsin S

The comparison of the L, K and S propeptide structures shows, that the cathepsin L propeptide contains two phenylalanines (Phe63p and Phe71p). Phe63p of the loop between strand β1p and helix α3p, and Phe71p of helix α3p extend into the vicinity of the active site (Figure 4-26). These two phenylalanines make hydrophobic contacts to Leu144 of the PBL. In cathepsin S, Leu144 is replaced by Phe146. A complex between cathepsin S and the cathepsin L propeptide may lead to sterical clashes between the three phenylalanines. The side chains of Leu80p and Val91p of the cathepsin S propeptide point in the same direction as the equivalent residues Phe63p and Phe71p of the cathepsin L propeptide. On the other hand, the short residues of the helix α3p of the cathepsin S propeptide may allow an effective binding onto the surface of the cathepsin L active site without sterical clashes. In the propeptide of cathepsin K, Phe63p is also replaced by a leucine, namely Leu63p, and helix α 3p contains short residues extending into the S′-groove, as well. This may guarantee a good inhibition of cathepsin L and S, resulting in the similar inhibition constants. The orientation of Phe146 in the procathepsin S structure indicates a potential involvement in the formation of the S′1 substrate-binding site. Crystallization of the enzyme in complex with different substrates may elucidate this.

Another cathepsin S - specific residue in the PBL is Arg141, which may play a key role in the different inhibitory behavior of the propeptides. In the previously determined cathepsin S structures the side chain of Arg141 is directed away from the active site to the solvent. The high averaged B-factor of 58.0 Å^2 for the side chain and the poorly defined electron density indicate a flexible side chain, which may adopt different conformations and interact with possible substrates (McGrath *et al*. 1998; Turkenburg *et al*. 2002). From helix α3p of cathepsin S Thr92p points towards Arg141 forming a possible van der Waals contact with the hydrophobic part of the arginine side chain (Figure 4-26 and 4-27). In the corresponding helix of the cathepsin L propeptide, Arg72p occupies the position of Thr92p (Figure 4-27). In the complex between the cathepsin L propeptide and cathepsin S, there may be a repulsion between the long and positively charged residues. Between Thr92p of the cathepsin S propeptide helix and the PBL residue Gly139 in cathepsin L, likewise repulsion is not

expected. Consequently, Phe146 and Arg141 may play a key role in the specificity of inhibitory propeptides and designed inhibitors, which could be proven by mutation of Phe146 and Arg141. The different K_i values determined for the cathepsin L propeptide with respect to cathepsin L and K could be explained by a possible sterical hindrance between the phenylalanines 63p and 71p and Gln143 of the PBL of cathepsin K, which may lead to a weaker inhibition of cathepsin K in comparison with L.

Furthermore, Arg141 is discussed as a critical residue, which prevents the inhibition of cathepsin S by the MHC class II-associated p41 li fragment (Guncar *et al*., 1999). After cleavage of p41 by cathepsin L or S, the fragment inhibits cathepsin L. In this way an enhancement of ther antigen presentation by inhibiton of the destructive cathepsin L activity is assumed (Rodriguez & Diment, 1995). The p41 fragment does not inhibit the closely related cathepsin S, which is explained in a modeled cathepsin S structure by sterical clashes between Arg141 with p41 residues and unfavourable electrostatic interactions between positively charged residues of p41 and the positively charged Arg141 (Guncar *et al*., 1999). In cathepsin L, Arg141 is replaced by Gly139, wich allows, in connection with neighbouring hydrophobic as well as negatively charged residues, the occupation of the active site by p41.

5.2.2.2 *Inhibitory activity by the C-terminal part of the propeptide*

Inhibition studies with truncated cathepsin L propeptides towards cathepsin L revealed that the C-terminal region contributes little to the overall inhibition (Carmona *et al*., 1996). However, similar investigations on the inhibitory activity of cathepsin B propeptide with mutated amino-acid residues revealed that the C-terminal part containing the residues occupying the S pockets is essential for the inhibitory activity of the cathepsin B propeptide (Chen *et al*.,1996). But concerning selectivity no exact data are presently available about the complete C-terminal part of the cathepsin propeptides. The C-terminal part between the residues occupying the S- subsites and the connection to the enzyme domain is highly flexible, and up to now no implications were found for a role in the specificity of inhibition.
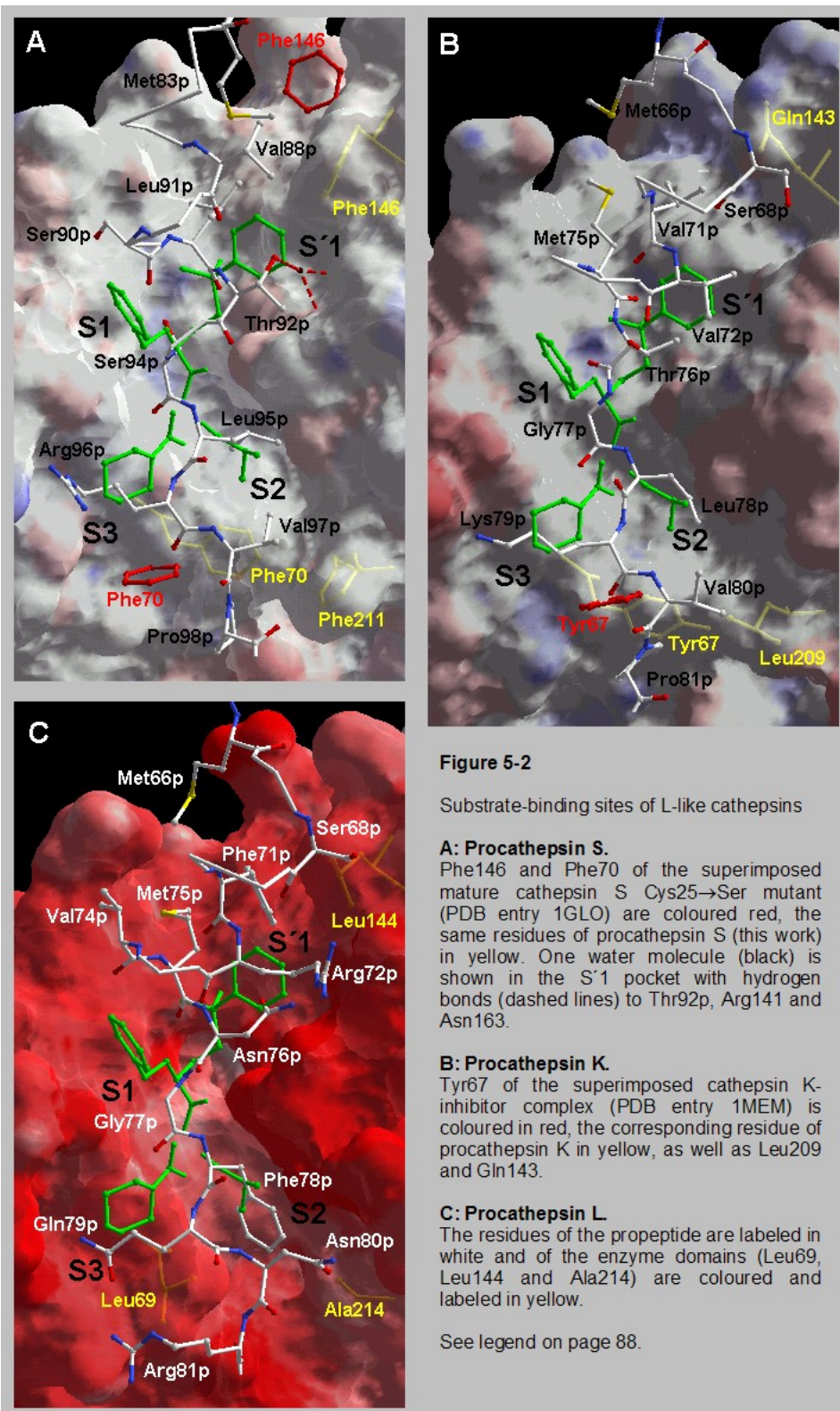
The C-terminal part of the cathepsin propeptides can be subdivided into two parts:

1. Amino acid residues, that occupy the S pockets:

The S1 and S2 pockets of the L-like cathepsins are occupied by short residues, whereas large residues such as arginine, lysine or glutamine are sticking into the S3 pocket. The pictures of Figure 5-2 display the residues of the C-terminal propeptide part, which are located in the substrate-binding sites. In order to visualize the substrate-binding sites, the inhibitor APC3328 (vinyl-sulfone-inhibitor, PDB entry 1MEM) was modelled into the active-site cavities of the cathepsins S and L using the crystal structure of cathepsin K with this inhibitor as a template (McGrath *et al*., 1997). The modelling of this inhibitor was based on the assumption that its orientation would be nearly the same in the active-site cavities of all L-like cathepsins. In cathepsin S, Phe70 seems to determine the anchoring of the C-terminal propeptide part via a C-H⋯π bond to Pro98p (Figure 4-28 and 5-2), and the same interaction between an aromatic residue (Tyr67) and a proline (Pro81p) can be observed in procathepsin K (Pro81p^{CD}EEETyr73^π distance = 3.34 Å and ρ Pro81p^{CDH} – Tyr73^{πx} = 131.69°). Lecaille *et al.* (2002) replaced Tyr67 and the C-terminal residue Leu209, which border the S2 pocket, by leucine and alanine, respectively. The introduction of the double mutation into the S2 subsite of cathepsin K changed the unique S2 binding preference of the protease for proline and leucine into a cathepsin L-like preference for phenylalanine. The cathepsin L-like S2 specificity of the mutated proteinase and the integrity of its catalytic site were confirmed by kinetic analysis of synthetic di- and tripeptide substrates as well as pH stability and pH activity profile studies. The loss of the ability to accept proline in the S2 binding pocket by the mutated proteinase completely abolished collagenolytic activity of cathepsin K. These results indicate that Tyr67 and Leu209 play a key role in the binding of proline residues in the S2 pocket of cathepsin K and are required for its collagenase activity. In the procathepsin S structure, Phe70 and Phe211 are arranged similarly as Tyr67 and Leu209 in cathepsin K and aromatic residues of Tyr67 and Phe70 interact in a similar way with prolines coming from the equivalent location of the propeptide region (see Figure 5-2). As described for cathepsin K, cathepsin S shows collagenolytic activity as well (Kirschke *et al*., 1989). The S2 pocket of cathepsin L is occupied by a phenylalanine (Phe78p). The preference of less bulky residues in the P2 position for cathepsins S and K may explain the weaker inhibition of the two enzymes by the cathepsin L propeptide.

2. Residues forming a loop surrounding the bottom side of the enzyme surface and connecting covalently the propeptide with the enzyme domain:

In the C-terminal part, below the residues occupying the S pockets of the cathepsin L propeptide, large residues such as Asn80p, Arg81p, Lys82p, and Arg84p are found instead of the anchoring prolines of the cathepsin S and K propeptides (see Table 5-1 and Figure 5-2). These residues may prevent a closeness of the C-terminal propeptide part to the enzyme surface due to possible sterical clashes (see Figures 4-17 and 5-2). This region has a positively charged electrostatic potential (see Figure 4-29), and a repulsion with positively charged residues of the cathepsin S and K enzymes could lead to an inefficient inhibition by the cathepsin L propeptide. But it is difficult to identify residues of the cathepsin K and S enzymes that may interact with residues from the C-terminal part of the cathepsin L propeptide. A superposition of all cathepsin L-like propeptides on different mature cathepsin enzymes gave differences between the compared C-terminal parts of the propeptides. Beyond the S3 subsite, the propeptide of cathepsin K forms a loop or, in the cathepsin S propeptide, a type III β-turn (see also Table 5-1). The positively charged residues in this region of procathepsin L interact with the residues of the enzyme domains through hydrogen bonds and salt bridges. These aspects raise the question in how the second part of the C-terminal propeptide region is involved in the inhibition of cathepsin-like proenzymes. The results of the investigations in recent years exclude an inhibitory capacity of the flexible C-terminal propeptide parts connecting the propeptide with the enzyme domains. But all these studies were carried out with expressed or chemically synthesized propeptides. The C-terminal parts of these structural elements used in the experiments are not covalently bound to the enzyme domains. Further investigations are required to elucidate the influence of these propeptide parts on inhibition. One step into this direction are studies with synthetic peptides (Chowdhury *et al*., 2002) and endogenous cysteine proteinase inhibitors of different species (Yamamoto *et al*., 2002), which show a high homology to the cathepsin propeptides.

Figure 5-2

Substrate-binding sites of L-like cathepsins

A: Procathepsin S.
Phe146 and Phe70 of the superimposed mature cathepsin S Cys25→Ser mutant (PDB entry 1GLO) are coloured red, the same residues of procathepsin S (this work) in yellow. One water molecule (black) is shown in the S´1 pocket with hydrogen bonds (dashed lines) to Thr92p, Arg141 and Asn163.

B: Procathepsin K.
Tyr67 of the superimposed cathepsin K-inhibitor complex (PDB entry 1MEM) is coloured in red, the corresponding residue of procathepsin K in yellow, as well as Leu209 and Gln143.

C: Procathepsin L.
The residues of the propeptide are labeled in white and of the enzyme domains (Leu69, Leu144 and Ala214) are coloured and labeled in yellow.

See legend on page 88.

Legend to Figure 5-2 (page 87)

The propeptide regions of cathepsin S (Image A), K (Image B), and L (Image C) are shown as ball-and-sticks-models in the substrate-binding sites of their cognate enzymes. The atoms are coloured in blue (N), red (O), yellow (S), and white (C). All molecules (PDB entry 1BY8 for procathepsin K, 1CS8 for procathepsin L) are in the same orientation with view on the active-site cavity. The inhibitor APC3328 (green) was modelled into the active sites by superposition of the cathepsin K-inhibitor complex (PDB entry 1MEM) with procathepsin S, K, L. The electrostatic potentials on the surface (transparent) of the enzyme domains of the procathepsins were calculated with the Swiss-PdbViewer (Guex & Peitsch, 1997) using the atomic partial charges for each mature enzyme molecule. The colour of the electrostatic potential maps range from red (-8 kT/mol) via white (neutral) to blue (+8 kT/mol).

S	K	L	
Ser94p	Gly77p	Gly77p	**S 1**
Leu95p	Leu78p	Phe78p	**S 2**
Arg96p	**Lys79p**	Gln79p	**S 3**
Val97p	Val80p	Asn80p	
Pro98p	Pro81p	Arg81p	
Ser99p	Leu82p	Lys81p	
Gln100p	Ser83p	Pro83p	
Trp101p	His84p	Arg84p	
Gln102p	Ser85p	Lys85p	
Arg103p	**Arg86p**	Gly86p	

Table 5-1 Propeptide residues of the C-terminal part

S1-3: Substrate binding sites.
Positively charged residues are highlighted in blue, residues involved in a loop (cathepsin K) or in a type III β-turn (cathepsin S) in green.

5.2.3 Consequences for the cathepsin S propeptide structure after mutation of conserved residues of the ER(F/W)N(I/V)N-motif

The procathepsin S structure allows the localization of conserved residues in the globular domain of the propeptide. Some of these residues were exchanged by alanines (Kreusch *et al*., 2000), and the mutated propeptides analyzed with respect to folding competence, inhibitory capacity and their ability to fold independently to a tertiary structure (Schilling *et al*., 2001, Pietschmann *et al*., 2002). These investigations have been described in the introduction. Considering the positions of the mutated tryptophanes 28, 31, and 52 in the globular domain, it can be assumed that lack of one of these residues may prevent an interaction between the crossed helices α1p and α2 (see Figure 4-24). The loss of inhibitory and folding properties of the tryptophane mutants is coupled with a loss of the tertiary structure, which cannot be

completely compensated by salt bridges and hydrogen bonds formed by conserved amino-acids within the propeptide.

Apart from the tryptophanes, only a substitution of the conserved Val59p and Asn63p by alanines led to a significant decrease of folding competence and inhibitory capacity towards cathepsin S (Val59p: Ki 1.27 nM; Asn65p: Ki 1.85 nM; wildtype: Ki 0.06 nM; Schilling *et al.* 2001, Pietschmann *et al.*, 2002). In comparison to the wild-type propeptide, only a minor red-shift of the fluorescence emission spectra was found for the *in-vitro* folded mutants Val59pAla and Asn63pAla, indicating a folded structure similar to that of the wildtype propeptide. The same mutations in the zymogen did not affected transport through the transfected cells as well as maturation and secretion of the proenzyme (Kreusch *et al.*, 2000). The hydrogen bonds between Asn63p and the main-chain atoms of Leu75p contribute to the stabilization of the helix-loop-sheet structure near the end of helix α2p, where a hydrophobic interface to the PBL is formed. This hydrophobic interface displays the upper side of a contact area with the prosegment-binding loop (PBL) (Figure 4-24) and represents an important propeptide-enzyme interaction site. Alterations in this region may influence the orientation of the propeptide to the substrate-binding site and may prevent an optimal folding process and an inhibition of cathepsin S by its propeptide. The replacement of Asn63p by alanine may lead to a loosening of the loop-hairpin-loop-helix α2p interaction. Val59p forms a van der Waals contact to the hydrophobic side chain of Leu147 of the PBL (Figure 4-24 and 4-25). The lack of the Val59 side chain may lead to a possible reduction of the hydrophobic contact area to the PBL resulting in an altered position of the propeptide towards the surface of cathepsin S. The exchange of Glu87p by an alanine had no significant consequences. The conserved salt bridge (Figure 4-24) between Glu87p (α3p) and Arg48p (α2p) anchors the third propeptide α-helix onto the second one, but it does not seem to be necessary for the propeptide to cover effectively the active site.

The functional role of the conserved ER(F/W)N(I/V)N-motif is unclear. With respect to all mutated residues of the propeptide region, only four residues (Trp28p, Trp31p, Trp52p, Asn63p) are critically involved in the formation of the tertiary propeptide structure, and Val59p is involved in the anchoring of the propeptide onto the PBL. Only three of these residues, namely Trp52p, Val59p and Asn63p belong to the conserved motif. Several other residues are involved in the stabilization of the N-terminal globular domain of the propeptide, and are also involved in the interaction between propeptides and their cognate enzymes (see

also Table 4-7). It cannot be excluded that the mutation of other non-conserved residues may have more dramatical effects on the propeptide structure than observed so far. The most important aspect is the formation of the hydrophobic core by three interdigitating tryptophanes coming from the two crossed helices in the N-terminal region of the cathepsin L-like cathepsins. This, and the formation of the helix-loop-sheet structural motif mentioned above, are the most important parts of the propeptide structure, which determine the folding and the inhibition of the cathepsin L-like proteinases.

5.2.4 Electrostatic potentials and their influence on the activity of cathepsins

Cathepsin S is proteolytically active at a broad pH range (pH 4.5-7.5) against substrates such as collagen and elastin (Kirschke *et al*., 1989, Brömme *et al*., 1993). At pH 7.5 cathepsin S shows a stable catalytic activity against synthetic substrates for at maximum 15 hours, whereas similar studies on cathepsin K (Brömme *et al*., 2000), L, V (Turk *et al*., 1993, Brömme *et al*., 1999) and B (Turk *et al*., 1994) gave a complete inactivation after 6 minutes (cathepsin K, substrate: collagen), 1 hour (cathepsin V), 15-20 minutes (cathepsin L), and 11 minutes (cathepsin B) at pH 7.0-7.5. The unusual stability of cathepsin S at neutral pH could be explained by the electrostatic potential on the surface of cathepsin S, which is, in contrast to the other cathepsins, more neutral (see Figure 4-30). Especially in the region of the active site cavity and nearby, the molecule is more or less uncharged (pI see Table 4-8), whereas the same region in cathepsin L and B has a strong negative electrostatic surface potential. However, in the same region cathepsin K and V show some negative patches. Papain exhibits the greatest stability with 40 hours at pH 8.5 (Skelton, 1968, Groeger *et al*., 1989) and the highest pI (see Table 4-8) in comparison with the cathepsins mentioned above. On the surface of papain an equal distribution of neutral and positive patches is observed (see Figure 4-30). The neutral and positive electrostatic potentials in the area around the active site cavity may contribute to the stability of papain and cathepsin S, because at neutral/alkaline pH there are no repulsion effects between hydrophobic and basic amino acid residues in the region of the active site cavity which may lead to destructive alterations of the enzyme domains. In the same region of cathepsin L, B, K, and V, alkaline pH conditions may lead to repulsion effects between the negatively charged residues resulting in the observed inactivation of the enzymes but data about possible alterations in their structures are not published. However, the destructive influence of low pH on the tertiary structure combined with an irreversible inactivation of the enzyme was shown for cathepsin L by circular dichroism spectroscopy

(Turk *et al*., 1999). The measurements resulted in a decrease of the helical content at pH 3.0 and 4.0, which is in agreement with an acid-denatured protein described by Tanford *et al*. (1968).

Another aspect consists of the observation that the collagenolytic activity of cathepsin K increases in the presence of chondroitin-4-sulfate (Li *et al*., 2000). Based on the fact, that cathepsin K plays a predominant role in the osteoclastic matrix degradation the influence of glycosaminoglycans on the activity of cathepsin K was studied. In the presence of chondroitin-4-sulfate the residual activity of cathepsin K increased about 5-10-fold over a pH range of 5.5 to 7.5. Furthermore, its half-life increased at pH 7.0 from 6 to 40 minutes, which was interpreted as an increase of the intrinsic stability of cathepsin K. The same experiments carried out with cathepsin L gave no increasing of its half-life, the enzyme was completely inactivated above pH 6.5. The increased stability of cathepsin K in the presence of chondroitin-4-sulfate is interpretated as an interaction between the negatively charged glycosaminoglycans and the positively charged surface area on the backside of cathepsin K, which results in a structure stabilization. Such interactions can be excluded for cathepsin L due to the strong negative electrostatic potential on the surface of the enzyme. Further investigations concerning the interactions of substrates and parts of cellular components with cathepsins via their surface area potential are necessary for a better understanding of the influence of such compounds on the activity of lysosomal proteinases.

6.   *References*

Anfinsen, C. B. & Scheraga, H. A. 1975. Experimental and theoretical aspects of protein folding. *Adv. Protein Chem*. 29, 205-300

Baoguang, Z., Janson, C. A., Amegadzie, B. Y., D'Alessio, K., Griffin, C., Hanning, C. R., Jones, C., Kurdyla, J., McQueney, M., Xiayang, Q., Smith, W. W. & Abdel-Meguid, S. S. 1997. Crystal structure of human osteoclast cathepsin K complex with E-64. *Nat. Struct. Biol*. 4, 109-111

Billington, C. J., Mason, P., Magny, M. C. & Mort, J. S. 2000. The slow-binding inhibition of cathepsin K by its propeptide. *Biochem. Biophys. Res. Commun*. 276, 924-929

Bogdanov, M. & Dowhan, W. 1999. Lipid-assisted Protein Folding. *J. Biol. Chem*. 274, 36827-36830

Bossard, M. J., Tomaszek, T. A., Thompson, S. K., Amegadzie, B. Y., Hanning, C. R., Jones, C, Kurdyla, J. T., McNulty, D. E., Drake, F. H., Gowen, M. & Levy, M. A. 1996. Proteolytic activity of human osteoclast cathepsin K. Expression, purification, activation, and substrate identification. *J. Biol. Chem*. 24, 12517-12524

Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem*. 72, 248-254

Brandl, M., Weiss, M., Jabs, A., Sühnel, J. & Hilgenfeld, R. 2001. C-H$\cdots\pi$-Interactions in Proteins. *J. Mol. Biol.* 307, 357-377

Brömme, D., Bonneau, P. R., Lachance, P., Wiederanders, B., Kirschke, H., Peters, C., Thomas, D. Y, Storer, A. C. & Vernet, T. 1993. Functional expression of human cathepsin S in Saccharomyces cerevisiae. Purification and characterization of the recombinant enzyme. *J. Biol. Chem*. 268, 4832-4838

Brömme, D. & McGrath, M. E. 1996. High level expression and crystallization of recombinant human cathepsin S. *Protein Sci*. 5, 789-791

Brömme, D., Okamoto, K., Wang, B. B. & Biroc, S. 1996. Human cathepsin O2, a matrix protein-degrading cysteine protease expressed in osteoclasts. Functional expression of human cathepsin O2 in *Spodoptera frugiperda* and characterization of the enzyme. *J. Biol. Chem.* 271, 2126-2132.

Brömme, D., Li, Z., Barnes, M. & Mehler, E. 1999. Human Cathepsin V Functional Expression, Tissue Distribution, Electrostatic Surface Potential, Enzymatic Characterization, and Chromosomal Localization. *Biochemistry* 38, 2377-2385

Brown, J. A. & Swank, R. T. 1983. Subcellular redistribution of newly synthesized macrophage lysosomal enzymes. Correlation between delivery to the lysosomes and maturation. *J. Biol. Chem*. 258, 15323-15328

Brünger, A. T. 1992. The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472-475

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

Burley, S. K. & Petsko, G. A. 1988. Weakly polar interactions in proteins. *Adv. Prot. Chem.* 39, 125-189

Cantor, A. B., Baranski, T. J. & Kornfeld, S. 1992. Lysosomal enzyme phosphorylation. II. Protein recognition determinants in either lobe of procathepsin D are sufficient for phosphorylation of both the amino and carboxyl lobe oligosaccharides. *J. Biol. Chem.* 267, 23349-23356

Carmona, E., Dufour, E., Plouffe, C., Takebe, S., Mason, P., Mort, J. S. & Menard, R. 1996. Potency and selectivity of the cathepsin L propeptide as an inhibitor of cysteine proteases. *Biochemistry* 35, 8149-8157

CCP4 (1994). Collaborative Computional Project Number 4, The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sec. D* 50, 760-763

Chen, Y., Plouffe, C., Menard, R., Storer A. C. 1996. Delineating functionally important regions and residues in the cathepsin B propeptide for inhibitory activity. *Febs Letter*, 939, 24-26

Chowdhury, S. F., Sivaraman, J., Wang, J., Devanathan, G., Lachance, P., Qi, H., Menard, R., Lefebvre, J., Konishi, Y., Cygler, M., Sulea, T. & Purisima, E. O. 2002. Design of noncovalent inhibitors of human cathepsin L. From the 96-residue proregion to optimized tripeptides. *J. Med. Chem.* 45, 5321-5329

Coulombe, R., Grochulski, P., Sivaraman, J., Ménard, R., Mort, J.S. & Cygler, M. 1996. Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J* 15, 5492-5503

Delbruck, R., Desel, C., von Figura, K., Hille-Rehfeld, A. 1994. Proteolytic processing of cathepsin D in prelysosomal organelles. *Eur. J. Cell Biol.* 64, 7-14

Derewenda, Z. S., Lee, L., & Derewenda, U. 1995. The occurence of C-H··O hydrogen bonds in proteins. *J. Mol. Biol.* 252, 248-262

Diederichs, K. & Karplus, P. A. 1997. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* 4, 269–275

Dolinar, M., Maganja, D. B., & Turk, V. 1995. Expression of full-length human procathepsin L cDNA in *Escherichia coli* and refolding of the expression product. *Biol. Chem.* 376, 385-388

Feigelson, R. S. 1988. The relevance of small molecule crystal growth theories and techniques to the growth of biological macromolecules. *J. Cryst. Growth.* 90, 1-13

Gafvelin, G., Sakaguchi, M., Andersson, H. & von Heijne, G. 1997. Topological Rules for Membrane Protein Assembly in Eukaryotic Cells. *J. Biol. Chem.* 272, 6119-6127

Garcia-Ruiz, J. M. & Moreno, A. 1994. Investigations on protein crystal growth by the gel acupuncture method. *Acta Crystallogr*. D 50, 484 – 452

Giegé, R, Ng, J. D., Lorber, B., Théobald-Dietrich, A., Kern, D. & Witz, J. 1996. The crystallization of biological macromolecules from precipitates: Evidence for Ostwald ripening. *J. Cryst. Growth* 168, 50-62

Griffiths, G., Hoflack, B., Simons, K., Mellman, I. & Kornfeld, S. 1988. The mannose-6-phosphate receptor and the biogenesis of lysosomes. *Cell* 52, 329-341

Groeger, U., Stehle, P., Fürst, P., Leuchtenberger, W. & Drauz, K. 1989. Papain-catalyzed synthesis of dipeptides. *Food Biotechnol.* 2, 187-198

Groves, M.R., Taylor, M.A.J., Scott, M., Cummings, N.J., Pickersgill, R.W. & Jenkins, J.A. 1996. The prosequence of procaricain forms an α-helical domain that prevents access to the substrate-binding cleft. *Structure* 4, 1193-1203

Groves, M.R., Coulombe, R., Jenkins, J. & Cygler, M. 1998. Structural basis for specificity of papain like cysteine protease proregions toward their cognate enzymes. *Proteins Struct. Funct. Genet*. 32, 504-514

Guay, J., Falguereyret, J. P., Ducret, A., Percival, M. D. & Mancini, J. A. 2000. Potency and selectivity of inhibition of cathepsin K, L and S by their respective propeptides. *Eur. J. Biochem*. 267, 6311-6318

Guex, N & Peitsch, M. C. 1997. SWISS-MODEL and Swiss-PdbViewer: an environment for comparative protein modelling. *Electrophoresis*, 18, 2714-2723

Guncar, G., Podobnik, M., Pungercar, J., Strukelj, B., Turk, V. & Turk, D. 1998. Crystal Structure of porcine Cathepsin H determined at 2.1 Å Resolution: Location of the Mini-chain C-terminal Carboxyl Group defines Cathepsin H Aminopeptidase Function. *Structure* 6, 51-61

Guncar, G., Pungercic, G., Klemencic, I. & Turk, D. 1999. Crystal structure of MHC class II-associated p41 li fragment bound to cathepsin L reveals the structural basis for differentiation between cathepsins L and S. *EMBO J.* 18, 793-803

Guo, Y. L., Kurz, U., Schultz, J. E., Lim, C. C., Wiederanders, B. & Schilling, K. 2000. The α1/2 helical backbone of the prodomaines defines the intrinsic inhibitory specificity in the cathepsin L-like cysteine protease subfamily. *FEBS Letters* 469, 203-207

Jerala, R., Zerovnik, E., Kidric, J. & Turk, V. 1998. pH-induced Conformational Transitions of the Propeptide of Human Cathepsin L. *J. Biol. Chem*. 273, 11498-11504

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. Sec. A* 47, 110-119

Kabsch, W. & Sander, C. 1983. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* 22, 2577-2637.

Kamphuis, I.G., Kalk, K.H., Swarte, M.B.A. & Drenth, J. 1984. Structure of papain at 1.65 Å resolution. *J. Mol. Biol.* 179, 233-256

Karrer, K.M., Pfeiffer, S.L. & DiThomas, M.E. 1993. Two distinct gene subfamilies within the family of cysteine protease genes. *Proc. Natl. Acad. Sci.* 90, 3063-3067

Kirschke, H., Schmidt, I. & Wiederanders, B. 1986. Cathepsin S. The cysteine proteinase from bovine lymphoid tissue is distinct from cathepsin L (EC 3.4.22.15). *Biochem. J.* 240, 455-459

Kirschke, H., Wiederanders, B., Brömme, D. & Rinne, A. 1989. Cathepsin S from bovine spleen. Purification, distribution, intracellular localization and action on proteins. *Biochem. J.* 264, 467-473

Kirschke, H., Rawlings, N. D., Barett, A. J. 1995. Proteases 1: lysosomal cysteine proteinases. *Protein Profile* 2. *Oxford University* Press, 1587-1643

Kopitar, G., Dolinar, M., Strukelj, B., Pungercar, J. & Turk, V. 1996. Folding and activation of human procathepsin S from inclusion bodies produced in Escherichia coli. *Eur. J. Biochem.* 236, 558-562

Kornfeld, R. Kornfeld, S. 1985. Assembly of asparagine-linked oligosaccharides. *Ann. Rv. Biochem.* 54, 631-664

Kornfeld, S. & Mellman, I. 1989. The biogenesis of lysosomes. *Ann. Rev. Cell Biol.* 5, 483-525

Kos, J., Werle, B., Lah, T. & Brunner, N. 2000. Cysteine proteinases and their inhibitors in extracellular fluids: markers for diagnosis and prognosis in cancer. *Int. J. Biol. Markers* 15, 84-89.

Kraulis, P. J. 1991. MOLSCRIPT: A program to Produce both Detailed and Schematic Plots of Protein Structures. *J. Appl. Cryst.* 24, 946-950

Kreusch, S., Fehn, M., Maubach, G., Nissler, K., Rommerskirch, W., Schilling, K., Weber, E., Wenz, I. & Wiederanders, B. 2000. An evolutionary conserved tripartite tryptophane motif stabilizes the prodomains of cathepsin L-like cysteine proteases. *Eur. J. Biochem.* 267, 2965-2972

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

Lai, C. M., Shen, W. Y., Constable, I. & Rakoczy, P. E. 2000. The use of adenovirus-mediated gene transfer to develop a rat model for photoreceptor. *Invest. Ophthalmol. Vis. Sci.* 41, 580-584

LaLonde, J. M., Baoguang, Z., Janson, C. A., D´Alessio, K. J., McQueney, M. S., Orsini, M. J., Debouck, C. M. & Smith, W. W. 1999. The Crystal Structure of Human Procathepsin K. *Biochemistry* 38, 862-869

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. 1993. PROCHECK: a program to check the sterochemical quality of protein structures. *J. Appl. Cryst.* 26, 283-291

Lecaille, F., Choe, Y., Brandt, W., Li, Z., Craik, C. S. & Brömme, D. 2002. Selective inhibition of the collagenolytic activity of human cathepsin K by altering its S2 subsite specificity. *Biochemistry* 41, 8447-8854

Lecaille, F., Kaleta, J. & Brömme D. 2003. Human and Parasitic Papain-like Cysteine Proteases: Their Role in Physiology and Pathology and Recent Developments in Inhibitor Design. *Rev. Chem.* in press

Li, Z., Hou, W. S. & Brömme, D. 2000 Collagenolytic Activity of Cathepsin K Is Specifically Modulated by Cartilage-Resident Chondroitin Sulfates. *Biochemistry* 39, 529-536

Luckow, V. A. & Summers, M. D. 1988. Trends in development of baculovirus expression vectors. *Biotechnology* 6, 47-55

Mason, C. S., Lamers, M. B., Henderson, I. M., Monk, T. & Williams, D. H. 2001. Baculovirus expression and characterization of rodent cathepsin S. *Protein Expr. Purif.* 23, 45-54

Matsuura, Y., Possee, R. D., Overton, H. A. & Bishop, D. H. L. 1987. Baculovirus expression vectors: the requirements for high level expression of proteins, including glycoproteins. *J. Gen. Virol.* 68, 1233-1250

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

Maubach, G., Schilling, K., Rommerskirch, W., Wenz, I., Schultz, J. E., Weber, E. & Wiederanders, B. 1997. The inhibition of cathepsin S by its propeptide: specificity and mechanism of action. *Eur. J. Biochem.* 250, 745-750

McGrath, M. E., Klaus, J. L., Barnes, M. G., Brömme, D. 1997. Crystal structure of human cathepsin K complexed with a potent inhibitor. *Nat. Struct. Biol.* 4, 105-108

McGrath, M.E., Palmer, J.T., Brömme, D. & Somoza, J.R. 1998. Crystal structure of human cathepsin S. *Protein Sci.* 7, 1294-1302

McIntyre, G. F. & Erickson, A. H. 1991. Procathepsin L and D are membrane-bound in acidic microsomal vesicles. *J. Biol. Chem.* 266, 15438-15445

McIntyre, G. F. & Erickson, A. H. 1993. The lysosomal proenzyme receptor that binds procathepsin L to microsomal membranes at pH 5 is a 43-kDa integral membrane protein. *Proc. Nathl. Acad. Sci. USA* 90, 10588-10592

McIntyre, G. F., Godbold, G. D. & Erickson, A. H. The pH-dpendent Membrane Association of Procathepsin L Is Mediated by a 9-Residue Sequence within the Propeptide. *J. Biol. Chem.* 269, 567-572

McPherson, A.1982. *The preparation and analysis of protein crystals.* Wiley, New York

McPherson, A. 1999. *Crystallization of Biological Macromolecules*. Cold Spring Harbor Laboratory Press, New York

Miers, H. A. & Isaac, F. 1907. The spontaneous crystallization of binary mixtures: Experiments on salol and betol. *Proc. Roy. Soc. Lond. A*, 79, 322

Munger, J. S., Haass, C., Lemere, C. A., Shi, G. P., Wong, W. S., Teplow, D. B., Selkoe, D. J. & Chapman, H. A. 1995. Lysosomal processing of amyloid presursor protein to A beta peptides: a distinct role for cathepsin S. *Biochem. J.* 311, 299-305

Nakagawa, T. Y., Brissette, W. H., Lira, P. D., Griffiths, R. J., Petrushova, N., Stock, J., McNeish, J. D., Eastman, S. E., Howard, E. D., Clarcke, S. R., Rosloniec, E. F., Elliott, E. A. & Rudensky, A. Y. 1999. Impaired invariant chain degradation and antigen presentation and diminished collagen-induced arthritis in cathepsin S null mice. *Immunity* 10, 207-217

Navaza, J. 1994. AMoRe: An automated package for molecular replacement. *Acta Crystallogr. Sec. A* 50, 157-163

Nicholls, A., Sharp, K. & Honig, B. 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, 11, 281-296

Nissler, K., Kreusch, S., Rommerskirch, W., Strubel, W., Weber, E. & Wiederanders, B.1998. Sorting of non-glycosylated human procathepsin S in mammalian cells. *Biol. Chem.* 379, 219-224

Nissler, K., Strubel, W., Kreusch, S., Rommerskirch, W., Weber, E. & Wiederanders. 1999. The half-life of human procathepsin S. *Eur. J. Biochem*.263, 717-25

Nissler, K., Oehring, H., Krieg, R., Pierskalla, A., Weber, E., Wiederanders, B. & Halbhuber, K. J. 2002. Cytochemical demonstration of expression and distribution of non-glycosylated human lysosomal cathepsin S in HEK 293 cells. *Cell. and Mol. Biol.* 48, 297-308

Ogino, T., Kaji, T., Kawabata, M., Satoh, K., Tomoo, K., Ishida, T., Yamazaki, H., Ishidoh, K. & Kominami, E. 1999. Function of the propeptide region in recombinant expression of active procathepsin L in Escherichia coli. *J. Biochem. Tokyo*. 126, 78-83

Ostwald, W. 1896. *Lehrbuch der Allgemeinen Chemie,* vol. 2, part 1. Leipzig, Germany

Ostwald, W. 1897. Studien über die Bildung und Umwandlung fester Körper. *Z. Phys. Chem.* 22, 289-330

Otwinowski, Z. & Minor, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307-326

O´Reilly, D. R., Miller, L. K. & Luckow, V. A. 1992. *Baculovirus Expression Vectors: A Laboratory Manual*. W. H. Freeman and Company. New York, N.Y.

Petanceska, S. & Devi, L. 1992. Sequence analysis, tissue distribution, and expression of rat cathepsin S. *J. Biol. Chem.* 267, 26038-26043

Petancesca, S., Burke, S., Watson, S., J. & Devi, L. 1994. Differential distribution of messenger RNAs for cathepsin B, L, and S in adult rat brain: an *in situ* hybridisation study. *Neuroscience* 59, 729-738

Petancesca, S., Canoll, P. & Devi, L. A. 1996. Expression of cathepsin S in phagocytic cells. *J. Biol. Chem.* 271, 4403-4409

Pietschmann, S., Fehn, M., Kaulmann, G., Wiederanders, B. & Schilling, K. 2002. Cathepsin S proregion´s foldase function is strictly based upon its domain structure. *Biol. Chem.* 383, 1453-1458

Podobnik, M., Kuhelj, R., Turk, V. & Turk, D. 1997. Crystal Structure of the Wild-type Human Procathepsin B at 2.5 Å Resolution Reveals the Native Active Site of a Papain-like Cysteine Protease Zymogen. *J. Mol. Biol.* 271, 774-788

Quraishi, O., Nagler, D. K., Fox, T., Sivaraman, J., Cygler, M., Mort, J. S. & Storer, A. C. 1999. The occluding loop in cathepsin B defines the pH dependence of inhibition by its propeptide. *Biochemistry* 38, 5017-5023

Quarishi, O. & Storer, A. C. 2001. Identification of Internal Autoproteolytic Cleavage Sites within the Prosegments of Recombinant Procathepsin B and Procathepsin S. *J. Biol. Chem.* 276, 8118-8124

Qian, F., Chan, S. J., Cong, Q., Bajkowski, A. S., Steiner, D. F. & Frankfater, A. 1991. The expression of cathepsin B and other lysosomal cysteine proteinases in normal tissues and in tumors. *Biomed. Biochim. Acta* 50, 4-6

Ravelli, R. B. G., Sweet, R. M., Skinner, J. M., Duisenberg, A. J. M. & Kroon, J. 1997. STRATEGY: a program to optimize the starting spindle angle and scan range for X-ray data collection. *J. Appl. Crystallogr.* 30, 551-554

Rawlings, N. D. & Barrett, A. J. 1993. Evolutionary families of peptidases. *Biochem. J.* 290, 205-218

Rawlings, N. D. & Barrett, A. J. 1994. Families of cysteine peptidases. *Methods Enyzmol.* 244, 461-486

Riese, R. J., Wolf, P. R., Brömme, D., Natkin, L. R., Villadangos, J. A., Ploegh, H. L & Chapman, H. A. 1996. Essential role for cathepsin S in MHC class II-associated invariant chain processing and peptide loading. *Immunity* 4, 357-366

Riese, R. J., Mitchell, R. N., Villadangos, J. A., Shi, G. P., Palmer, J. T., Karp, E. D., De Sanctis, G. T., Ploegh, H. L. & Chapman, H. A. 1998. Cathepsin S activity regulates antigen presentation and immunity. *J. Clin. Invest.* 101, 2351-2363

Rodriguez, G. M. & Diment, S. 1995. Destructive proteolysis by cysteine proteases in antigen presentation of ovalbumin. *Eur. J. Immunol.* 25, 1823-1827

Saegusa, K., Ishimaru, N., Yanagi, K., Arakaki, R., Ogawa, K., Saito, I., Katunuma, N. & Hayashi, Y. 2002. Cathepsin S inhibitor prevents autoantigen presentation and autoimmunity. *J. Clin. Invest.* 110, 361-369

Samarel, A. M., Ferguson, A. G., Worobec, S. W. & Lesch, M. 1989. Effects of cysteine protease inhibitors on rabbit cathepsin D maturation. *Am. J. Physiol.* 257, 1069-1079

Sauter, C., Otala, F., Gavira, J. A., Vidal, O., Giege, R. & Garcia-Ruiz, J. M. 2001. Structure of tetragonal hen egg-white lysozyme at 0.94 Å from crystals grown by the counter-diffusion method. *Acta Crystallogr. D Biol. Crystallogr.* 57, 1119-1126

Sauter, N. K., Mau, T., Rader, S. D. & Agard, D. A. 1998. Structure of α-lytic protease complexed with its pro region. *Nat. Struct. Biol.* 5, 945-950

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

Schilling, K., Pietschmann, S., Fehn, M., Wenz, I. & Wiederanders, B. 2001. Folding Incompetence of Cathepsin L-Like Cysteine Proteases May Be Compensated by the Highly Conserved, Domain-Building N-Terminal Extension of the Proregion. *Biol. Chem.* 382, 859-865

Shi, G. P., Munger, J. S., Meara, J. P., Rich, D. H. & Chapman, H. A., 1992. Molecular cloning and expression of human alveolar macrophage cathepsin S, an elastolytic cysteine protease. *J. Biol. Chem.* 267, 7258-7262

Sivaraman, J., LaLumière, M., Ménard, R. & Cygler, M. 1999. Crystal structure of wild-type human procathepsin K. *Protein Sci.* 8, 283-290

Sivaraman, J., Nägler, D.K., Zhang, R., Ménard, R. & Cygler M. 2000. Crystal Structure of Human Procathepsin X: A Cysteine Protease with Proregion Covalently Linked to the Active Site Cysteine. *J. Mol. Biol.* 295, 939-951

Skelton, G.S. Papaya proteinases. I. Temperature-and pH-stability curves. 1968. *Enzymologia* 35, 270-274

Smith, S.M. & Gottesman, M. M. 1989. Activity and deletion analysis of recombinant human cathepsin L expressed in *Escherichia coli*. *J. Biol. Chem.* 264, 20487-20495

Smith, G. L., Machett, M. & Moss, B. 1983. Infectious vaccina virus recombinants that express hepatitis B virus surface antigen. *Nature* 302, 490-495

Smith, G. E., Ju, G., Ericson, B. L., Moschera, J., Lahm, H., Chizzonite, R. & Summers, M. D. 1985. Modification and secretion of human interleukin 2 produced in insect cells by baculovirus expression vector. *Proc. Nat. Acad. Sci. USA* 82, 8404-8408

Somoza, J.R., Zhan, H., Bowman, K.K., Yu, L., Mortara, K.D., Palmer, J.T., Clark, J.M. & McGrath, M.E. 2000. Crystal Structure of Human Cathepsin V. *Biochemistry* 39, 12543-12551

Steed, P. M., Lasala, D., Liebmann, J., Wigg, A., Clark, K. & Knap, A. K. 1998. Characterization of recombinant human cathepsin B expressed at high levels in baculovirus. *Protein Sci.* 7, 2033-2037

Stout, G. H. & Jensen, L. H. 1968. *X-ray Structure Determination: A Practical Guide*, p. 402, Macmillan, New York

Sukhova, G. K., Shi, G. P., Simon, D. I., Chapman, H. A. & Libby, P. 1998. Expression of the elastolytic cathepsins S and K in human atheroma and regulation of their production in smooth muscle cells. *J. Clin. Invest.* 102, 576-583.

Summers, M. D. & Smith, G. E. 1988. *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures*. Texas Agricultural Experiment Station Bulletin No. 1555

Tanford, C. 1968. Protein denaturation. *Adv. Prot. Chem*. 23, 121-282

Tao, K., Stearns, N. A., Dong, J., Wu, Q. L. & Sahagian, G. G. 1994. The proregion of cathepsin L is required for proper folding, stability, and ER exit. *Arch. Biochem. Biophys*. 311, 19-27

Taylor, M. A., Baker K. C., Briggs, G. S., Connerton, I. F., Cummings, N. J., Pratt, K. A., Revell, D. F., Freedman, R. B. & Goodenough, P. W. 1995. Recombinant pro-regions from papain and papaya proteinase IV are selective high affinity inhibitors of the mature papaya enzymes. *Protein Eng.* 8, 59-62

Tobbell, D. A., Middleton, B. J., Raines, S., Needham, M. R., Taylor, I. W., Beveridge, J. Y. & Abbott, W. M. 2002. Identification of in vitro folding conditions for procathepsin S and cathepsin S using fractional factorial screens. *Protein Expr. Purif.* 24, 242-254

Turk, B., Dolenc, I., Turk, V. & Bieth, J. G. 1993. Kinetics of the pH-induced inactivation of human cathepsin L. *Biochemistry* 32, 375-380

Turk, B.,.Dolenc, I., Zerovnik, E., Turk, D., Gubensek, F. & Turk, V. 1994. Human cathepsin B is a metastable enzyme stabilized by specific ionic interactions associated with the active site. 1994. *Biochemistry* 33, 14800-14806

Turk, B., Turk, D. & Turk, V. 2000. Lysosomal cysteine proteinases: more than scavengers. *Biochim. Biophys. Acta*. 1477, 98-111

Turkenburg, J. M., Lamers, M. B., Brzozowski, A. M., Wright, L. M., Hubbard, R. E., Sturt, S. L. & Williams, D. H. 2002. Structure of a Cys25→Ser mutant of human cathepsin S. *Acta Crystallogr. Sec. D* 58, 451-455

Turnsek, T., Kregar, I. & Lebez, D. 1975. Acid sulfhydryl protease from calf nodes. *Biochim. Biophys. Acta* 403, 514-520

van Klompenburg, W., Nilsson, I., von Heijne, G. & de Kruijff, B. 1997. Anionic phospholipids are determinants of membrane protein topology. *EMBO J.* 16, 4261-4266

Venkatachalam, C. M. 1968. Stereochemical Criteria for Polypeptides and Proteins. V Conformation of a System of Three Linked Peptide Units. *Biopolymers* 6, 1425-1436.

Vernet, T., Berti, P. J., de Montigny, C., Musil, R., Tessier, D. C., Menard, R., Magny, M. C., Storer, A. C. & Thomas, D. Y. 1995. Processing of the papain precursor. The ionization state

of a conserved amino acid motif within the proregion participates in the regulation of intramolecular processing. *J. Biol. Chem*. 270, 10838-10846

Völkel, H., Kurz, U., Linder, J., Klumpp, S., Gnau, V., Jung, G. & Schultz, J. E. 1996. Cathepsin L is an intracellular and extracellular protease in *Paramecium tetraurelia*. Purification, cloning, sequencing and specific inhibition by its expressed propeptide. *Eur. J. Biochem*. 238, 198-206

von Figura, K. & Hasilik, A. 1986. Lysosomal enzymes and their receptors. *Ann. Rev. Biochem.* 55, 167-193

von Heijne, G. 1997. Getting greasy: how transmembrane polypeptide segments integrate into the lipid bilayer. *Mol. Microbiol.* 24, 249-253

Webb, N. R. & Summers, M. D. 1990. Expression of Proteins Using Recombinant Baculoviruses. *Techinque* 2, 173-188

Weiss, M. S. & Hilgenfeld, R. 1997. On the use of the merging R-factor as a quality indicator for X-ray data. *J. Appl. Crystallogr.* 30, 203–205

Weiss, M. S., Brandl, M., Sühnel, J., Pal, D. & Hilgenfeld, R. 2001. More hydrogen bonds for the (structural) biologist. *Trends Biochem. Sci.* 26, 521-523

Wiederanders, B., Brömme, D., Kirschke, H., Figura, K., Schmidt, B. & Peters, C. 1992. Phylogenetic conservation of cysteine proteinases. Cloning and expression of a cDNA coding for human cathepsin S. *J. Biol. Chem.* 267, 13708-1371

Wiederanders, B. 2000. The function of propeptide domains of cysteine proteinases. *Adv. Exp. Med. Biol.* 477, 261-270

Wilkins M. R., Gasteiger E., Bairoch A., Sanchez J.-C., Williams K.L., Appel R.D. & Hochstrasser D.F. *Protein Identification and Analysis Tools in the ExPASy Server* in: 2-D Proteome Analysis Protocols, 1998. Editor A.J. Link. Humana Press, New Jersey

Yamamoto, Y., Watabe, S., Kageyama, T. & Takahashi, S. Y. 1999. Proregion of *Bombyx mori* cysteine proteinase functions as an intramolecular chaperone to promote proper folding of the mature enzyme. *J. Arch. Insect. Biochem. Physiol.* 42, 167-178

Yeyeodu, S., Kyujeong, A., Madden, V., Chapman, R., Song, S. & Erickson, A. H. 2000. Procathepsin L Self-Association as a Mechanismen for Selective Secretion. *Traffic* 1, 724-737

Zerovnik, E., Kopitar, G., Kos, J. & Turk, V. 1998. Refolding of recombinant sulphonated procathepsin S and of reduced chicken cystatin; implications for renaturation experiments. *Biochimica et Biophysica Acta* 1383, 211-218

Lebenslauf

Angaben zur Person

Name:	Guido Kaulmann
Geburtsdatum/-ort:	20.01.1966
Familienstand:	verheiratet mit Ursula (geb. Schell)

Schulbildung

1972 – 1976	Grundschule in Bielefeld
1976 – 1986	Gymnasium in Bielefeld (Bavink-Gymnasium, heute Gymnasium „Am Waldhof")
Juni 1986	Abitur

Wehrdienst

Juli 1986 – Juni 1988	Zeitsoldat (SaZ 2)

Studium

Oktober 1988	Beginn des Studiums der Biologie an der Universität Bielefeld
Oktober 1993 – Dezember 1994	Diplomarbeit am Institut für Neurobiologie in Magdeburg: „Methodische Aspekte zur funktionellen Untersuchung Calcium bindender Proteine in einem in vitro-Modell"
März 1996	Diplom
April 1996 – November 1997	Wissenschaftlicher Mitarbeiter am Institut für Anatomie I (Medizinische Fakultät) der Universität Köln
Dezember 1997 – Juni 2002	Wissenschaftlicher Mitarbeiter am Institut für Biochemie I (Medizinische Fakultät) der Friedrich-Schiller-Universität Jena. Experimentelle Arbeiten zur vorliegenden Arbeit unter der Anleitung von Prof. Dr. Bernd Wiederanders mit dem Ziel Promotion
Dezember 1999 – Juni 2002	Gastaufenthalt am Institut für Molekulare Biotechnologie in Jena (Abeilung Strukturbiologie und Kristallographie von Prof. Dr. Rolf Hilgenfeld). Kristallograpische Arbeiten bis zum Abschluss der Dissertation.
Juli 2002 – September 2002	Wissenschaftlicher Mitarbeiter am Institut für Molekulare Biotechnologie in Jena (Abteilung Strukturbiologie und Kristallographie).
Seit Juni 2003	Wissenschaftlicher Mitarbeiter am Birkbeck College, School of Crystallography, University of London, UK

Publikationen

The crystal structure of the Cys25→Ala mutant of human procathepsin S reveals unique propeptide-enzyme interactions. Kaulmann, G., Palm, G. J., Schilling, K., Hilgenfeld, R. & Wiederanders, B. (in Vorbereitung)

Functions of Propeptide Parts in Cysteine Proteases. Wiederanders, B., Kaulmann, G. & Schilling, K. 2003, *Curr. Prot. and Pept. Sci.* **4**, 309-326

An unfolding/refolding step helps crystallizing a poorly soluble protein. Kaulmann, G., Palm, G. J., Schilling, K., Hilgenfeld, R. & Wiederanders, B. 2003, *Acta Cryst. Sec. D* **59**, 124, 1243-1245

Cathepsin S proregion´s foldase function is strictly based upon its domain structure. Pietschmann, S., Fehn, M., Kaulmann, G., Wenz, I., Wiederanders, B. & Schilling, K. 2002, *Biol. Chem.* **383**, 1453-1458

Wissenschaftliche Vorträge

- **9th International Conference On The Crystallization Of Biological Macromolecules (ICCBM 9)**, *23. bis 28. März 2002 in Jena, Deutschland*

  Poster Presentation: 'Crystallization of Cathepsin S Propeptide Using Conventional and Novel Methods'.

- **6th Proteolyse Tag**, *23. November 2001 in Eisenberg, Deutschland,*

  Vortrag: 'Kristallisation und Strukturanalyse von menschlichem Procathepsin S'.

- **2nd General Meeting of the International Proteolysis Society (IPS),** *31.Oktober bis 4. November 2001 in Freising, Deutschland*

  Poster Presentation: 'Crystallization and preliminary X-ray Diffraction Studies of Human Procathepsin S'.

- **5th Proteolyse Tag**, *5. Dezmeber 2000 in Jena, Deutschland*

  Vortrag: 'Kristallisation und Strukturanalyse von Procathepsin S und seines Propeptids: Eine Untersuchung zur funktionellen Bedeutung von Propeptiden im Faltungsprozess von Cystein-Proteasen'.

- **18th Winter School on Proteinases and their Inhibitors**, *8. bis 11. März 2000 in Tiers, Italien*

  Vortrag: 'Propeptide Domains of Cysteine Proteases: Necessary Elements for the correct Folding of Mature Enzymes'.

Danksagung

Die Arbeit entstand am *Institut für Biochemie I des Klinikums der Friedrich-Schiller-Universität* in Kooperation mit der *Abteilung Strukturbiologie und Kristallographie des Instituts für Molekulare Biotechnologie* in Jena.

Herzlichst bedanken möchte ich mich bei

Ehrenwörtliche Erklärung:

Hiermit erkläre ich, dass mir die Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität bekannt ist,

ich die Dissertation selbst angefertigt habe und alle von mir benutzten Hilfsmittel, persönlichen Mitteilungen und Quellen in meiner Arbeit angegeben sind,

mich folgende Personen bei der Auswahl des Materials sowie bei der Herstellung des Manuskripts unterstützt haben:
Prof. Dr. B. Wiederanders und Prof. Dr. R. Hilgenfeld.

Die Hilfe eines Promotionsberaters nicht in Anspruch genommen wurde und dass Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen,

dass ich die Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht habe und

dass ich die gleiche, eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung nicht bei einer anderen Hochschule als Dissertation eingereicht habe.

Jena den 06.02.2003   Guido Kaulmann
Ort, Datum   Unterschrift des Verfassers