

# **Mechanism of action of mammalian cystatins**

**Studies of inhibition of cysteine endo- and  
exopeptidases by cystatins A and C**

**Alona Pavlova**

*Department of Molecular Biosciences  
Uppsala*

**Doctoral thesis  
Swedish University of Agricultural Sciences  
Uppsala 2003**

**Acta Universitatis Agriculturae Sueciae**  
Veterinaria 154

ISSN 1401-6257  
ISBN 91-576-6379-3  
© 2003 Alona Pavlova, Uppsala  
Tryck: SLU Service/Repro, Uppsala 2003

## Abstract

Pavlova, A., 2003. *Mechanism of action of mammalian cystatins. Studies of inhibition of cysteine endo- and exopeptidases by cystatins A and C*. Doctor's dissertation. ISSN 1401-6257, ISBN 91-576-6379-3.

Mammalian cystatins are tight-binding, reversible inhibitors of potentially destructive cysteine proteases, both endogenous and exogenous. The N-terminal region residues, Ile2 and especially Pro3, and the second binding loop residues, Leu73 and to some extent Pro74, of the family 1 cystatin, cystatin A, were shown to be essential for effective inhibition of papain and cathepsins L and B. Both regions were found to act in endopeptidase inhibition primarily by anchoring the enzymes in long-lived complexes with the inhibitor. However, in inhibition of the exopeptidase, cathepsin B, the N-terminal region also contributes to the high association rate.

As previously shown for cystatin C, stopped-flow kinetics demonstrated a two-step mechanism of cystatin A binding to cathepsin B, with a conformational change in the second, rate-limiting step. In contrast, binding of cystatins A and C to a cathepsin B variant with an increased mobility of the occluding loop that blocks the active-site cleft occurred in an apparent one-step reaction. The second step of cathepsin B binding by cystatins was therefore concluded to involve displacement of the occluding loop. Cystatin C was shown to be a much better displacer of this loop than cystatin A.

New effective and specific inhibitors of cysteine proteases were engineered by grafting individual binding regions of more potent cystatins into cystatin A. Introduction of motifs from the second binding loops of cystatins B or C into the corresponding loop of cystatin A increased the affinity for papain by ~10-fold. Moreover, replacement of the N-terminal region of cystatin A by that of cystatin C resulted in the affinity for cathepsin B being increased ~15-fold, to a level higher than that of any natural cystatin known. The latter chimeric inhibitor was as effective as cystatin C in displacing the occluding loop of cathepsin B, the rate of this step being increased by ~100-fold. The N-terminal region of a cystatin was thus concluded to promote occluding loop displacement by appropriately positioning the binding loops of the inhibitor entering the active-site cleft of cathepsin B.

*Keywords:* cysteine protease, papain, cathepsin, inhibition, cystatin, affinity, binding kinetics, two-step reaction mechanism, occluding loop, binding loop.

*Author's address:* Alona Pavlova, Department of Molecular Biosciences, SLU, BMC, Box 575, S-751 23 UPPSALA, Sweden.



# Contents

**Introduction, 9**

## **Previous investigations, 9**

Cysteine proteases of the papain family, 9

*Classification, 9*

*Catalytic activity and substrate binding, 10*

*Lysosomal cathepsins, 11*

Protein inhibitors of cysteine proteases of the papain family, 15

*Cystatins, 16*

*Other protein and peptide inhibitors of cysteine proteases of the papain family, 21*

*Physiological functions of cystatins, 22*

*Three-dimensional structure of cystatins and mode of inhibition of papain-like cysteine proteases by cystatins, 24*

*Contributions of individual binding regions of cystatins to binding of target proteases, 29*

*Kinetic mechanism of the interaction between cystatins and their target proteases, 35*

## **Present investigation, 37**

Contribution of cystatin A N-terminal residues to the inhibition of different cysteine proteases (Paper I), 37

Importance of the second binding loop of cystatin A for stabilization of the complexes with different target proteases (Paper II), 39

Evidence for dislocation of the occluding loop of cathepsin B in the second step of the binding by cystatins (Paper III), 41

Improvement of the inhibitory characteristics of cystatin A by grafting of the protease binding regions of more potent inhibitors (Paper IV), 44

**General conclusions, 46**

**References, 48**

**Acknowledgements, 64**

# Appendix

## Papers I-IV

The present thesis is based on the following papers, which will be referred to by their Roman numerals:

I. Estrada, S., Pavlova, A., and Björk, I. (1999) The contribution of N-terminal region residues of cystatin A (stefin A) to the affinity and kinetics of inhibition of papain, cathepsin B, and cathepsin L. *Biochemistry* 38, 7339–7345.

II. Pavlova, A., Estrada, S. and Björk, I. (2002) The role of the second binding loop of the cysteine protease inhibitor, cystatin A (stefin A), in stabilizing complexes with target proteases is exerted predominantly by Leu73. *Eur. J. Biochem.* 269, 5649–5658.

III. Pavlova, A., Krupa, J. C., Mort, J. S., Abrahamson, M., Björk, I. (2000) Cystatin inhibition of cathepsin B requires dislocation of the proteinase occluding loop. Demonstration by release of loop anchoring through mutation of His110. *FEBS Letters* 487, 156–160.

IV. Pavlova, A. and Björk, I. (2003) Grafting of features of cystatins C or B into the N-terminal region or second binding loop of cystatin A (stefin A) substantially enhances inhibition of cysteine proteinases. *Manuscript*.

Reprints are published with the permission of the journals concerned.

## Abbreviations

The following abbreviations are used in the text:

$k_{\text{ass}}$	Second-order association rate constant
$K_{\text{d}}$	Equilibrium dissociation constant
$k_{\text{diss}}$	Dissociation rate constant
$K_{\text{i}}$	Inhibition constant
$k_{\text{obs}}$	Observed pseudo-first-order rate constant

*To my teacher, O.B.*



## **Introduction**

Proteases, also termed peptidases, are crucial for normal functioning of all cells and organisms. They are enzymes that catalyze the hydrolysis of peptide bonds in proteins and peptides. Depending on the location of the cleavage site, proteases are divided into endopeptidases (or proteinases) and exopeptidases. Endopeptidases cleave internal peptide bonds in target polypeptide chains, whereas exopeptidases, comprising amino- and carboxypeptidases, hydrolyze bonds near either the N- or C-terminus of polypeptides, respectively. Based on the specific chemical groups responsible for catalysis, proteases are classified into five major types: the serine, cysteine, aspartic, threonine and metallo-peptidases (Rawlings & Barrett, 1999). There are also proteases for which the catalytic mechanism remains to be determined; these are classified as proteases of unknown type.

Proteases have a key role in general protein turnover and also have other vital functions, exerted via processing of proteins by limited cleavage. Their activity is strictly regulated at different levels, since uncontrolled proteolysis would damage the finely tuned biological machinery of cells and tissues and lead to pathophysiological states and eventual death of the organism. Inhibition of proteases by protein inhibitors represent one of the most important mechanisms of regulating protease activity.

Cysteine proteases have been found in all kind of organisms, from viruses, bacteria and protozoa to plants and vertebrates. This thesis briefly reviews different aspects concerning the classification, structures, functions and inhibition of cysteine proteases. It focuses on cystatins, protein inhibitors of papain-like cysteine proteases. Cystatins are believed to protect an organism from the potentially harmful activity of endogenous target proteases as well as from proteases released by infectious virus, bacteria and parasites. The mode of inhibition of papain-like cysteine proteases by mammalian cystatins is highlighted in both Previous and Present investigations.

## **Previous investigations**

### **Cysteine proteases of the papain family**

#### *Classification*

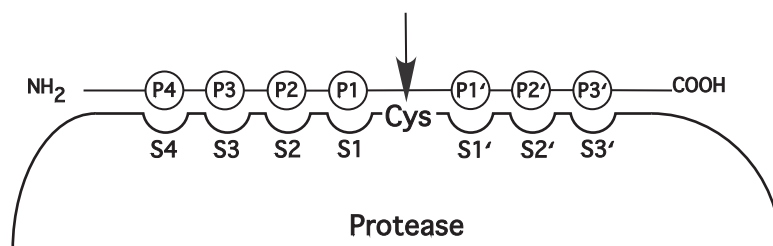
In the recently developed MEROPS database, proteases of the major types are classified into clans and families according to their evolutionary relationship, reflected by structural similarities in the "peptidase unit", *i.e.* in the part of the enzyme responsible for peptide bond hydrolysis (Barrett & Rawlings, 2001; Rawlings & Barrett, 1993, 1999, 2000; Rawlings *et al.*, 2002). Proteases with statistically significant similarities in amino-acid sequences are grouped into families. Clans are formed by related families, which are suggested to have evolved from a common ancestor in light of similar tertiary and/or secondary structures, as well as identical catalytic motifs and limited similarities around the

catalytic residues (Barrett & Rawlings, 2001). Cysteine proteases are subdivided into eight clans: CA, CD, CE, CF, CH, CK, PA and PB (Barrett & Rawlings, 2001). A common mode of action of members of the largest clan, CA, is reflected by the presence of the catalytic triad Cys/His/Asn(Asp) in these proteases. Papain-like cysteine proteases, inhibition of which is the main subject of this thesis, comprise the C1 family of this clan. This family includes plant proteases (such as papain, chymopapain, caricain, bromelain, actinidin, ficin and aleurain), cysteine proteases from *Trypanosoma*, *Leishmania*, *Schistosoma*, *Giardia*, *Fasciola* and other parasites, bleomycin hydrolase from various organisms and a number of mammalian lysosomal cathepsins. The origin of the C1 family may predate the eukaryote/prokaryote divergence or, at the latest, the ancestor of the papain family could be a digestive protease in vacuoles of an archeprotzoa, which then diverged into proteases in plant vacuoles and mammalian lysosomes (Barrett *et al.*, 1986b; Berti & Storer, 1995).

### *Catalytic activity and substrate binding*

In proteases of the C1 family, Cys25, His159, Asn175 and also Gln19 (in papain numbering) are highly conserved residues, directly involved in peptide bond hydrolysis. In general, a protease active site is comprised of such catalytically active residues as well as of residues contacting the bound substrate. The substrate binding area can be subdivided into a number of so called binding pockets or subsites, each contacting a single amino-acid residue of the bound substrate (Fig. 1) (Schechter & Berger, 1967). In papain, seven binding pockets were initially suggested to accommodate seven subsequent residues of a substrate, four S subsites binding substrate residues N-terminal to the scissile peptide bond (P residues), and three S' subsites binding substrate residues C-terminal to this bond (P' residues). This nomenclature for substrate binding subsites has long been used for description of the subsites in other papain-like cysteine proteases. However, a more recent analysis of available three-dimensional structures of papain-like proteases in complex with a substrate analogue and of kinetic data has prompted a revision of this nomenclature (Turk *et al.*, 1998). Only three binding sites in the different proteases, S2, S1 and S1', were shown to interact with the corresponding substrate residues via both main-chain and side-chain atoms, whereas the interactions between residues of the S3 or S2' subsites and the substrate involved only side-chain contacts. As no evidence for the existence of common S4 and S3' subsites in different papain-like proteases was found, the use of the less precise term "binding area" was suggested for the regions interacting with substrate residues beyond P2 and P2' (Turk *et al.*, 1998, 2000).

The substrate of a papain-like cysteine protease binds along the active-site cleft of the enzyme in a rather extended conformation and in such a manner that the peptide bond to be hydrolyzed comes in close proximity to the sulfhydryl group of the catalytic Cys25, which acts as a nucleophile attacking the carbonyl carbon of the scissile bond. The imidazolium ring of the neighboring His159 accepts a proton from the nucleophile and simultaneously acts as a proton donor for the nitrogen atom of the scissile bond. Asn175 contributes to a proper orientation of this imidazolium ring, whereas the side chain of Gln19 participates in stabilizing an intermediate acyl-enzyme product (Storer & Menard, 1994; Otto &



*Fig. 1.* Schematic representation of a peptide substrate bound in the active site of a papain-like cysteine protease, according to Schechter & Berger (1967). Seven successive amino-acid residues (P4 to P3') of the substrate are suggested to interact with seven distinct substrate-binding subsites (S4 to S3') of the protease. The prime notation denotes residues C-terminal of the scissile peptide bond (indicated by an arrow), which is in close proximity to the catalytic Cys.

Schirmeister, 1997). In the first, acylation, step of the hydrolysis, an acyl-enzyme is formed, and a peptide with the P1' residue at its N-terminus dissociates from the active site of the protease. The next, deacylation, step involves a nucleophilic attack on the acyl-enzyme by a water molecule, which results in dissociation of a second peptide product with the P1 residue as C-terminus and in the regenerated protease.

### *Lysosomal cathepsins*

#### General characterization

Mammalian proteases that are localized in lysosomes are in general termed cathepsins (Kirschke *et al.*, 1995; Turk *et al.*, 2002b). A major part of them belongs to the papain family of cysteine proteases, *i.e.* cathepsins B, C, H, L, S, K, O, F, X, V and W. Most papain-like cathepsins are endopeptidases, although both cathepsins B and H also exhibit endopeptidase activity. Cathepsin B is thus a carboxypeptidase and cathepsin H an aminopeptidase. Cathepsins C and X are true exopeptidases, cathepsin C being an aminopeptidase and cathepsin X a carboxy-mono or -dipeptidase (Turk *et al.*, 2001b). Lysosomal cathepsins are generally monomers with molecular masses of ~30 kDa. The only exception known is cathepsin C, which is a tetramer of ~200 kDa. Most cathepsins exhibit an optimum of their activity at a slightly acidic pH and in the presence of reducing compounds.

Lysosomal cathepsins are synthesized as inactive precursors with an N-terminal extension, comprising a signal sequence that directs the proteins to the endoplasmic reticulum and a propeptide (also called a proregion or prosegment). The length of the propeptide varies from 38 residues in cathepsin X to 251 residues in cathepsin F (Cygler & Mort, 1997; Groves *et al.*, 1998; Wiederanders, 2000). The propeptide binds into the S and S' binding subsites of the cathepsin in the reverse orientation relative to that of a substrate, thereby inactivating the enzyme. Activation of the proenzyme occurs by cleavage and removal of the propeptide from the active site with no global conformational rearrangements of the protein. Besides keeping the proenzyme inactive, the propeptide has been shown to be essential for correct folding of the cathepsin and for transport of the

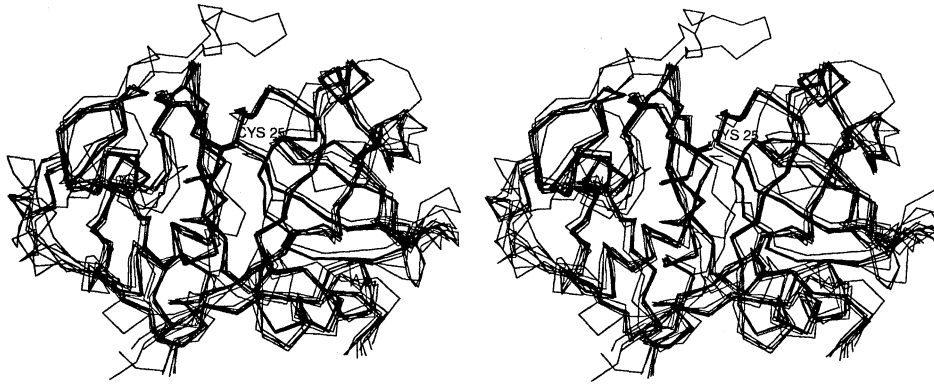
proenzyme to lysosomes or endosomes. It has also been suggested to contribute to increased stabilization of the proenzyme at neutral pH (Khan & James, 1998; Turk *et al.*, 2000). Endopeptidases can be autoactivated in an acidic environment, whereas the true exopeptidases, cathepsins C and X, are activated by other proteases (Turk *et al.*, 2001b). The autocatalytic activation is an intermolecular process, in which one of cathepsin molecules activates another one in a chain-like reaction (Turk *et al.*, 2000). Lysosomal glycosaminoglycans have been proposed to promote cathepsin processing *in vivo* by weakening the interaction between the propeptide and the active-site cleft (Turk *et al.*, 2000).

On the basis of sequence homology, papain-like cathepsins have been suggested to be divided into two main subgroups, the cathepsin L- and B-like proteases (Cygler & Mort, 1997; Turk *et al.*, 2000). The propeptides of proteases of these two types differ both in size and sequence. Cathepsins of the first subgroup, comprising cathepsins L, V, K, S, W, F, and H, have propeptides that are usually ~100 residues long with two highly conserved motifs, ERF(W)NIN and GNFD. The second group comprises cathepsin B variants from different sources with ~60 residues long propeptides lacking the former conserved motif. Cathepsins F and W have alternatively been proposed to form a separate, third subgroup of cathepsin F-like proteases with a conserved ERFNAQ-motif in the propeptide (Wiederanders, 2000). Propeptides of the other cathepsins do not show any homology with either cathepsin L-, B- and F-like proregions or to each other.

#### Tertiary fold

The high sequence homology of members of the C1 family of cysteine proteases results in similar tertiary structures (Fig. 2). A polypeptide chain of a typical endopeptidase of the family is folded into two domains, separated by the deep "V"-shaped active-site cleft (Kirschke *et al.*, 1995; Turk *et al.*, 1997, 1998). The L-domain consists of most of the N-terminal half of the protease polypeptide chain and has an  $\alpha$ -helical structure. The R-domain is formed by the extreme N-terminal residues and the C-terminal half of the chain and is folded into a  $\beta$ -barrel, covered at the top and bottom by shorter  $\alpha$ -helical motifs. The essential catalytic residues, Cys25 and His159, are situated on the opposite sides of the cleft in the L- and R-domains, respectively, on the rim between a shallow, wide part and a narrow, deep part of the crevice. Residues that form the more open part of the active-site cleft, comprising the S1 and S2 subsites, are highly conserved in members of the C1 family.

Exopeptidases of the papain family have the same common fold as endopeptidases. However, in the former, the access to the substrate binding sites is restricted by additional structural elements, which determines the exopeptidase activity of these enzymes. In particular, in cathepsin B an insertion of about 20 residues forms a loop that covers the top of the active-site cleft and partially occludes the S2' and S3' substrate binding subsites (Musil *et al.*, 1991; Nägler *et al.*, 1997). The presence of this loop greatly reduces the endopeptidase activity of cathepsin B. Moreover, the positively charged imidazole side chains of His110 and His111 in the loop can bind the negatively charged C-terminal carboxylic group of a polypeptide substrate and direct the latter into position for hydrolysis



*Fig. 2.* Structural similarities between papain-like cysteine proteases, illustrated by a stereo view of the superposition of C $\alpha$  atom traces of nine members of the papain family. The structures are shown in the standard view with the R- and L-domains at the right and left, respectively, and with the active-site clefts at the top. The catalytic Cys25 of papain is marked. The structure distinguished from the others by the presence of the loop on top of the active-site cleft is that of cathepsin B. (Reprinted, with permission, from Turk *et al.* (1998). Copyright 1998, Walter de Gruyter & Co, Berlin–New York.)

of the C-terminal dipeptide, thereby accounting for the dipeptidylcarboxypeptidase activity of cathepsin B (Musil *et al.*, 1991; Illy *et al.*, 1997).

In cathepsin H, a glycosylated fragment of the propeptide, called the mini-chain, is attached to the mature enzyme through a disulfide bond in such a manner that three of its residues occupy the S2 and S3 substrate-binding subsites in the active-site cleft (Guncar *et al.*, 1998; McGrath, 1999). Another part of the active-site cleft of cathepsin H is partially filled with the carbohydrate rings of the mini-chain. The negatively charged C-terminal carboxyl group of the mini-chain has been suggested to interact with the positively charged N-terminal amino group of a substrate. The mini-chain and part of its carbohydrate chain, together with a short four-residue insertion loop in the active-site cleft, are thus responsible for the aminopeptidase activity of cathepsin H that results in cleavage of a single residue from the N-terminal end of the substrate. Similarly, the active-site cleft in cathepsins X and C, which are true exopeptidases, is partially occluded by a mini-loop and a propeptide segment, respectively (Nägler *et al.*, 1999; Klemencic *et al.*, 2000; Sivaraman *et al.*, 2000; Olsen *et al.*, 2001; Turk *et al.*, 2001a).

#### Cathepsins in physiological and pathophysiological states

Lysosomal proteases play important roles in terminal degradation of proteins and generation of amino acids for synthesis of new proteins. Most papain-like cathepsins, such as cathepsins B, C, F, H, L, O, and X, are ubiquitously expressed. Therefore, a common role associated with a nonspecific intracellular "housekeeping" function, *i.e.* breakdown of proteins targeted to lysosomes, has been ascribed to these cathepsins. In support of such a nonspecific role,

experiments with knock-out of the genes for cathepsins B and L showed that these two cathepsins are not essential for normal foetal growth. Cathepsin B- or L-deficient mice were normal at birth, and no defects in protein degradation were observed (Deussing *et al.*, 1998; Nakagawa *et al.*, 1998; Roth *et al.*, 2000), indicating that other proteases could take over most functions of the absent cathepsins.

In addition to nonspecifically degrading proteins, papain-like cathepsins also participate in protein processing by limited proteolysis, thereby regulating many important physiological processes (Turk *et al.*, 2000, 2001b). In these reactions, cathepsins may act very specifically. In cathepsin C-deficient mice, the activation of certain proteases was thus impaired (Pham & Ley, 1999). Furthermore, lack of cathepsin L in mice resulted in periodic hair loss and epidermal hyperplasia, suggesting that cathepsin L is essential for regulation of normal proliferation of hair follicle epithelial cells and basal keratinocytes of the epidermis (Roth *et al.*, 2000). Mice deficient in cathepsin L also had impaired processing of the invariant chain of major histocompatibility complex (MHC) class II molecules in cortical thymic epithelial cells (Nakagawa *et al.*, 1998). Moreover, cathepsin F was suggested to be important for this process in macrophages (Shi *et al.*, 2000).

The relatively recent discovery of cathepsins with expression restricted to specific tissues or cell types has suggested new roles for these proteases (Chapman *et al.*, 1997a; McGrath, 1999; Buhling *et al.*, 2000; Turk *et al.*, 2000). Thus, cathepsin S is found predominantly in lymphatic tissues such as spleen, lymph nodes and peripheral leukocytes, cathepsin K is highly expressed in osteoclasts at sites of bone resorption, cathepsin V is thymus- and testis-specific and cathepsin W is predominantly expressed in lymphocytes. Studies with cathepsin S-deficient mice demonstrated that this protease is specifically required for processing of the invariant chain in peripheral antigen presenting cells such as macrophages, B-cells and dendritic cells (Shi *et al.*, 1999; Villadangos *et al.*, 1999). Cathepsin V was suggested to be responsible for catalysis of the same reaction in human thymic epithelial cells (Turk *et al.*, 2000). Different cathepsins have thus been shown to be linked to antigen presentation, possibly acting through different pathways (Turk *et al.*, 2002c). Furthermore, cathepsin K was found to be critical for bone resorption, a part of the dynamic remodeling of bone (Saftig *et al.*, 2000; Turk *et al.*, 2000). Knockout of the cathepsin K gene in mice thus caused an impaired ability of osteoclasts to resorb bone during growth, which led to osteopetrosis, or increased bone density. In this state, bone demineralization proceeds normally but not digestion of the bone matrix. Another tissue-specific cathepsin, cathepsin V, was indicated to be involved in T-cell selection in humans, whereas a specific function in T-cell-mediated cytotoxicity was proposed for cathepsin W (Bhandoola *et al.*, 2000; Turk *et al.*, 2000). A number of novel, tissue-specific cathepsins in other mammals have also been reported, *i.e.* rat and mouse cathepsins Q, M, P, and R, which are primarily expressed in placenta and therefore may play important roles in implantation and embryonic development (Sol-Church *et al.*, 2002).

Extracellular cathepsins are potentially very destructive. In particular, abnormally high levels of extracellular papain-like cathepsins have been observed during development of various cancer types and metastasis (Calkins & Sloane, 1995;

Henskens *et al.*, 1996; Friedrich *et al.*, 1999; Fernandez *et al.*, 2001; Colella *et al.*, 2002). In cancer progression, such cathepsins are responsible for activation of other proteases and for degradation of components of the extracellular matrix (Turk *et al.*, 2000, 2002c). However, intracellular cathepsin B has also been implicated to have specific roles in reducing malignancy, in particular by acting as an execution protease in tumor cells (Foghsgaard *et al.*, 2001; Szpaderska & Frankfater, 2001). In agreement, lysosomal cathepsins have been indicated to participate in apoptosis also under other circumstances (Turk *et al.*, 2000, 2002a). Lysosomal cathepsins may also be secreted by other than cancer cells, *i.e.* by activated macrophages, osteoclasts and fibroblasts. Papain-like cathepsins, often extracellular or extralysosomal, have thus been linked to development of such diseases as inflammation, rheumatoid arthritis and osteoarthritis, gingivitis, osteoporosis, asthma, pulmonary emphysema, Alzheimer's disease, multiple sclerosis, muscular dystrophy, pancreatitis, myocardial disorders and glomerulonephritis (Turk *et al.*, 1997, 2000, 2002b).

Genetic defects in cathepsins may cause severe physiological disorders. Mutations in the human cathepsin K gene that lead to a total loss of enzyme activity were found to be associated with pycnodysostosis, a genetic disorder with an osteopetrotic phenotype (Saftig *et al.*, 2000), in agreement with the knock-out experiments in mice. Moreover, a loss-of-function mutation in the cathepsin C gene was shown to cause Papillon-Lefevre syndrome, characterized by early-onset periodontitis and palmo-plantar keratosis (Toomes *et al.*, 1999).

#### Regulation of activity of cathepsins

The activity of lysosomal papain-like cathepsins is regulated in many ways, *i.e.* at the gene translation and expression levels, by localization, by synthesis of the proteases as inactive precursors, by posttranslational modifications and by degradation (Chapman *et al.*, 1997b; Turk *et al.*, 1997, 2000). Alternative splicing of the cathepsins B and L genes has been observed to result in highly elevated translation and increased protease stability, as well as in an altered targeting mechanism and increased secretion. An essential regulating factor of the proteolytic activity of lysosomal cathepsins is the pH of the environment. Thus, the decrease in pH occurring during maturation of lysosomes presumably weakens the interaction between the propeptide and the proenzyme, favoring the activation process. Furthermore, cathepsins B, S and L are indicated to be irreversibly denatured in lysosomes towards the end of the maturation process, when pH has decreased to a value of 3.8. In addition, many cathepsins, *viz.* cathepsins L, B, H, K, V and F, are unstable at neutral pH and therefore are less active outside lysosomes. Finally, endogenous protein inhibitors are considered to be common and most important means of regulating the activity of mature cathepsins that have escaped lysosomes, as well as of exogenous cysteine proteases released by different infectious microorganisms (Turk *et al.*, 2001b, 2002b).

### **Protein inhibitors of cysteine proteases of the papain family**

Depending on their major physiological function, protein inhibitors of proteases may be subdivided into emergency and modulatory inhibitors (Turk *et al.*,

2002b). Protein inhibitors of the first type defend living cells and organisms in acute cases of occasionally increased proteolysis. They rapidly bind to their target proteases and inactivate these either irreversibly or for sufficiently long times. Two kinetic parameters have been suggested as criteria for whether protein inhibitors are capable of fulfilling this role. These are the delay time, which determines the time needed for ~99% completion of the binding reaction, and, in the case of reversible interactions, the stability time, during which an inhibitor—protease complex remains essentially intact and which is defined as 1/10th of the half life of the bimolecular complex (Bieth, 1984). Protein inhibitors that act with a delay time of <1 s and have a stability time of >10 min are regarded as emergency inhibitors. Normally, inhibitors of this type are localized separately from their target proteases and are in molar excess over the protease when performing their protective function. By contrast, modulatory inhibitors are colocalized with their target proteases and regulate the endogenous activity of the latter at different levels, either by irreversible or pseudo-irreversible slow binding of activated proteases (delay type) or by rapid binding of these proteases in short-lived complexes (buffer type). Threshold inhibitors are also modulatory inhibitors that act similarly to the emergency inhibitors, although in lower physiological concentrations compared with those of the proteases. A change of conditions, such as salt concentration and pH, or the presence of cofactors can alter the properties of protein inhibitors and convert them from one type to another.

### *Cystatins*

Cystatins are protein inhibitors of cysteine proteases of the papain family (Barrett, 1986; Turk & Bode, 1991; Turk *et al.*, 1997). Cystatins inactivate target proteases reversibly and competitively by indirect blockage of the catalytic centers of these enzymes, thereby preventing substrate docking and cleavage. In general, cystatins bind target proteases with rate constants as high as  $10^4$  to  $10^7$   $M^{-1}s^{-1}$  and dissociation constants in the nM to fM range. They form very stable bimolecular complexes with the proteases, keeping these inactive for hours and weeks. Hence, cystatins can function as either emergency or threshold inhibitors. Cystatins are widely distributed in animals, plants and protozoa, both intracellularly and extracellularly, which indicates an important physiological role (Barrett, 1986; Turk & Bode, 1991; Irvine *et al.*, 1992). They are believed to protect cells and organisms from uncontrolled activity of endogenous cysteine proteases released from lysosomes of dying or deceased cells. In addition, cystatins have been shown to be potent inhibitors of cysteine proteases used by microorganisms and parasites for invading host cells and entire organisms.

Cystatins are evolutionarily related proteins with a conserved physiological function. The cystatin superfamily comprises proteins with diverse primary structures, sizes and distributions. In particular, members of this superfamily can be one- or multidomain proteins, with or without disulfides, as well as glycosylated or nonglycosylated. Nevertheless, statistical analysis of their amino-acid sequences suggests that cystatins originate from a common ancestor protein with no disulfide bonds and that a precursor containing such bonds appeared about 1000 million years ago (Rawlings & Barrett, 1990). Further evolution of cystatins occurred, presumably by insertion and deletion of genetic material in small blocks



and also later by gene duplication. Proteins of the cystatin superfamily are traditionally subdivided according to their sequence similarities into three major families (Barrett *et al.*, 1986a). Originally, closely related proteins, diverging by less than a half of their amino-acid residues, were grouped into these individual families (Barrett, 1986; Rawlings & Barrett, 1990). However, later discovered cystatins have also been added to these families, despite a sequence similarity between some new family members and the old ones of substantially lower than 50%. Major characteristics for the classification of cystatins have instead become the size of the polypeptide chain and the presence or absence of internal disulfides. This simplification has led to uncertainties regarding the assignment of some cystatins to a particular family, and revision of the classification established to date, as well as introduction of new cystatin families or subfamilies, therefore appears to be necessary.

#### Family 1 cystatins

Members of this family, also called stefins, are the smallest in the cystatin superfamily, having a molecular mass of ~11 kDa. The single polypeptide chain of these proteins, consisting of ~100 residues, is folded into one domain. Family 1 cystatins lack both disulfide bridges and carbohydrates, thus representing the most primitive form of cystatins (Turk & Bode, 1991). This family includes cystatins A and B, found in different mammals: man (Machleidt *et al.*, 1983; Green *et al.*, 1984; Ritonja *et al.*, 1985; Abrahamson *et al.*, 1986), cow (Turk *et al.*, 1992, 1995a), rat (Takio *et al.*, 1983, 1984; Takeda *et al.*, 1985), sheep (Ritonja *et al.*, 1996), pig (Lenarcic *et al.*, 1993, 1996) and mouse (Tsui *et al.*, 1993), as well as bovine stefin C (Turk *et al.*, 1993) and pig stefin D (Lenarcic *et al.*, 1996). Alignment of sequences of family 1 cystatins reveals three evolutionarily highly conserved regions, *viz.* a Gly residue in the N-terminal region, a QVVAG segment in the middle of the polypeptide chain and a LP pair in the C-terminal part of the chain (Fig. 3).

Family 1 cystatins are located primarily intracellularly, in the cytosol (Turk & Bode, 1991), although they have also been detected in low concentrations in different extracellular fluids (Abrahamson *et al.*, 1986). Cystatins A and B are the most studied members of the family. Cystatin A is found predominantly in various epithelial cells and in polymorphonuclear leukocytes (Brzin *et al.*, 1983; Green *et al.*, 1984; Järvinen *et al.*, 1987), whereas cystatin B is widespread among different cells and tissues (Davies & Barrett, 1984; Katunuma & Kominami, 1985; Barrett, *et al.*, 1986b; Lenarcic *et al.*, 1986; Henskens *et al.*, 1996). These differences in distribution may reflect different physiological roles of the two cystatins. Cystatin A is often considered as a first-line guardian against cysteine proteases released by infectious microorganisms and parasites, whereas cystatin B presumably has a general protective function.

#### Family 2 cystatins

Family 2 members, uniformly termed cystatins, are also single-chain, one-domain proteins, although somewhat, ~2 kDa, larger than the family 1 inhibitors. The polypeptide chains of cystatins of family 2 are ~120 residues long and contain two characteristic disulfide bridges, at conserved positions in the C-terminal part of the

molecule (residues 71–81 and 95–115 in chicken cystatin numbering) (Rawlings & Barrett, 1990) (Fig. 3). In addition to the four cysteines, three other conserved regions are characteristic for members of cystatin family 2. A Gly residue in the N-terminal region, a QxVxG sequence in the middle of the chain and a PW dipeptide in the C-terminal part of a molecule are present in all cystatins of this family (Fig. 3). Most cystatins known are nonglycosylated, although mouse and rat cystatin C and also cystatins E/M and F are glycoproteins (Esnard *et al.*, 1990; Solem *et al.*, 1990; Ni *et al.*, 1997, 1998). Family 2 cystatins are synthesized as preproteins with a signal peptide and are present predominantly in extracellular fluids, although some have also been found intracellularly (Abrahamson *et al.*, 1986; Halfon *et al.*, 1998; Abrahamson, 1994; Nathanson *et al.*, 2002). The family includes mammalian cystatins: cystatin C of several different species (Poulik *et al.*, 1981; Grubb & Löfberg, 1982; Hirado *et al.*, 1985; Esnard *et al.*, 1988; Solem *et al.*, 1990; Tu *et al.*, 1990), cystatin D (Freije *et al.*, 1991, 1993), cystatin E/M (Ni *et al.*, 1997; Sotiropoulou *et al.*, 1997; Zeeuwen *et al.*, 2001), cystatin F/leukocystatin (Halfon *et al.*, 1998; Ni *et al.*, 1998; Morita *et al.*, 1999), cystatins S, SA and SN (Isemura *et al.*, 1984, 1986, 1987), and also chicken egg-white cystatin (Turk *et al.*, 1983).

#### Family 1

**CA** MIP**G**GLSEAK PATPEIQEIV DKVKPQLEEK TNETYGKLEA **40**  
**CB** MMCGAPSATQ PATAETQHIA DQVRSQLEEK ENKKFPVFKA

**CA** VQYKT**QVVAG** TNYIYKVRAG DNKYMHLKVF K**S**L**P**GGQNE**D**L **80**  
**CB** VSFKS**QVVAG** TNYFIKVVHG DEDFVHLRVF **Q**S**L**P**H**ENK**P**L

**CA** VLTGYQVDKN KDELDTGF  
**CB** TLSNYQTNKA KHDELTYF

#### Family 2

**CC** SSPGKPPRLV **G**GPMDASVEE EGVRRALDFA VGEYNKASND **40**  
**EWC** SEDRSRL **G**APVPVDEND EGLQRALQFA MAEYNRASND

**CC** MYHSRALQVV RARK**QIVAGV** NYFLDVELGR TTCTKTQPNL **80**  
**EWC** KYSSRVVRI SAKR**QIVS**GI KYILQVEIGR TTCPKSSGDL

**CC** DNCPFHDQPH LKRKAFCSFQ IYAV**PW**QGT**M** TLSKSTCQDA **120**  
**EWC** QSCEFHDPEE MAKYTTCTFV VYSIP**PW**LNQ**I** KLLESK**C**Q

Fig. 3. Amino-acid sequences of representatives of families 1 and 2 of the cystatin superfamily. CA, human cystatin A; CB, human cystatin B; CC, human cystatin C; EWC, chicken egg-white cystatin. Protein sequences were taken from the SWISS-PROT database (accession codes CYTA\_HUMAN, CYTB\_HUMAN, CYTC\_HUMAN and CYT\_CHICK, respectively) and aligned within each family. Signal peptides of family 2 cystatins are not shown. The motifs conserved within each family are in bold.

The recently reported cystatins E/M and F have only ~30% sequence identity with other cystatins of family 2. Nevertheless, they have been included in this family, mainly on the basis of their size and the presence of two conserved disulfide bonds, as well as the other evolutionarily conserved regions of the primary structure that are typical for family 2 cystatins. Cystatin F differs from the other family members by the presence of one additional disulfide bridge in the N-terminal part of the molecule. Chicken cystatin and cystatins C, E/M and F, besides being inhibitors of cysteine proteases of family C1, like all cystatins, also inhibit an entirely different enzyme, mammalian legumain, of family C13 of cysteine proteases (Chen *et al.*, 1997; Alvarez-Fernandez *et al.*, 1999).

The best studied representatives of family 2 cystatins are chicken cystatin and human cystatin C, the latter being the most abundant of the inhibitors of the cystatin superfamily in all human body fluids examined, with the highest concentration in seminal plasma and cerebrospinal fluid (Abrahamson *et al.*, 1986; Olafsson *et al.*, 1988; Abrahamson, 1994). The broad distribution of this protein implies its essential role in protection of the organism against exogenous and endogenous, occasionally released, cysteine proteases. Cystatins of the S-type are found predominantly in saliva but also in tears, urine and seminal plasma (Isemura *et al.*, 1984; Abrahamson *et al.*, 1986).

The distribution of cystatins D, E and F has not been investigated as extensively as that of cystatin C. Cystatin D has been found in saliva and tears and cystatin E in small amounts in urine (Freije *et al.*, 1993; Ni *et al.*, 1997). Cystatin F has been detected in low concentrations in blood and shown to be secreted in trace amounts by B cells and in significant amounts by T cells in culture (Ni *et al.*, 1998). Most information available concerns the expression pattern of the genes for these cystatins obtained by DNA and RNA techniques. The cystatin D gene has been shown to be expressed only in the parotid glands (Freije *et al.*, 1991). Data on expression of the gene for cystatin E/M are inconsistent. Different groups have reported diverse expression patterns of this cystatin, with expression being highest in liver, ovary and pancreas (Ni *et al.*, 1997), in primary breast cancer tumors with no expression in liver or brain (Sotiropoulou *et al.*, 1997) or restricted to epithelial cells and sweat glands (Zeeuwen *et al.*, 2001). Expression of the cystatin F gene has been detected in a number of tissues, the highest levels being found in peripheral blood leukocytes and spleen (Ni *et al.*, 1998) and also in hematopoietic cells and some human cancer cell lines (Halfon *et al.*, 1998; Morita *et al.*, 1999).

### Family 3 cystatins

Family 3, or the kininogen family, comprises the largest and most complex inhibitors of the cystatin superfamily, with molecular masses of ~50 or ~70–80 kDa (Barrett *et al.*, 1986a; Turk *et al.*, 2002b). The single chain of a kininogen is folded into three family 2-cystatin-like domains, all of them being glycosylated. In addition to the six conserved disulfide bridges, two in each domain, there are two additional disulfides in the N-terminal region of each of the kininogen domains 2 and 3. Kininogens are synthesized with a signal peptide and found extracellularly, with the highest concentrations, ~10  $\mu$ M, in blood plasma and synovial fluid (Abrahamson *et al.*, 1986; Abrahamson, 1994). There are three

kininogen types known, H-, L- and T-kininogens. High- and low-molecular-weight kininogens, H- and L-kininogen, respectively, are encoded by the same gene (Kitamura *et al.*, 1985) and are identical in their N-terminal parts, but H-kininogen is longer, containing an unrelated C-terminal region due to alternative splicing. T-kininogens, also called thiostatins, are found only in rats (Müller-Esterl *et al.*, 1986). Kininogens are multifunctional proteins and were first identified as precursors to vasoactive peptides, the kinins. The kinin sequence is located in the C-terminal part of a kininogen, and release of the kinin fragment by kallikrein cleavage converts the parent H- or L-kininogen to a two-chain form, stabilized by one interchain disulfide bond. The first domain of a kininogen is inactive in inhibition of cysteine proteases, whereas the two other domains can simultaneously bind two molecules of papain-like proteases (DeLa Cadena & Colman, 1991; Turk *et al.*, 1995b, 1996a). The second domain also inhibits calpain (Salvesen *et al.*, 1986).

#### Other cystatins and cystatin-related proteins

Many other proteins with inhibitory activity against papain-like cysteine proteases and with sequences related to well-studied mammalian cystatins have been discovered. Among them, a large group consists of defensive inhibitors from plants, phytocystatins, which have structural features intermediate between those of families 1 and 2 cystatins (Rawlings & Barrett, 1990; Brown & Dziegielewska, 1997; Margis *et al.*, 1998). Phytocystatins also possess a specific consensus sequence that is absent in other cystatins. Therefore, classification of plant cystatins as an independent family of the cystatin superfamily has been suggested. Cystatins from rice (Abe *et al.*, 1988; Kondo *et al.*, 1989, 1990, 1991), corn (Abe *et al.*, 1992), potato (Waldron *et al.*, 1993), avocado (Kimura *et al.*, 1995), papaya (Song *et al.*, 1995), soy (Misaka *et al.*, 1996), chelidonium majus (Rogelj *et al.*, 1998), sunflower (Kouzuma *et al.*, 2001) and lima bean (Lawrence & Nielsen, 2001) have been described. Related inhibitors have also been found in various lower organisms, such as African puff adder (Ritonja *et al.*, 1987), carp (Tsai *et al.*, 1996), horseshoe crab (Agarwala *et al.*, 1996), *Sarcophaga peregrina* larvae (Suzuki & Natori, 1985; Saito *et al.*, 1989), *Drosophila* (Delbridge & Kelly, 1990) and nematodes (Manoury *et al.*, 2001; Hartmann *et al.*, 2002). Cystatins in such lower organisms are more related to family 2 cystatins than the phytocystatins are (Turk *et al.*, 1997).

Several novel members of the cystatin superfamily have been identified on the basis of sequences derived from mouse cDNA libraries, *viz.* a cystatin-related epididymal and spermatogenic protein, termed *Cres*, and also *Cres2*, *Cres3*, testatin and cystatins T, TE-1 and SC (Cornwall *et al.*, 1992; Tohonen *et al.*, 1998; Kanno *et al.*, 1999; Shoemaker *et al.*, 2000; Li *et al.*, 2002; Hsia & Cornwall, 2003). Additionally, cDNAs and mRNAs for the human counterparts of mouse *Cres*, *Cres2* (termed cystatin 11 in human) and testatin were described (Cornwall *et al.*, 1999; Hamil *et al.*, 2002; Wassler *et al.*, 2002). These proteins, specifically expressed in the reproductive and neuroendocrine systems, were suggested to represent a new *Cres* subgroup of cystatin family 2 (Hsia & Cornwall, 2003; Cornwall & Hsia, 2003), as they have the conserved PW motif but lack the two other conserved motifs. A further search of DNA databases has

yielded a number of new *Cres*-like sequences in species ranging from nematode to human (Cornwall & Hsia, 2003). The mouse *Cres* protein was shown to not inhibit the cysteine proteases papain and cathepsin B but the calcium-dependent serine protease, prohormone convertase 2 (Cornwall *et al.*, 2003).

A further novel human cystatin-like molecule (CLM), ubiquitously expressed in normal tissues, has been shown to be related to cystatins C, D, S, SA and SN and especially to mouse and human testatins (>50% similarity) (Sun *et al.*, 2003). However, CLM lacks any conserved cystatin motifs, and as yet no protease-inhibitory activity has been demonstrated for this protein. In addition, other proteins with cystatin-related sequences but without inhibitory activity have been identified (Brown & Dziegielewska, 1997). Among them are fetuin and histidine-rich glycoprotein, both containing two cystatin-like domains with the conserved disulfide bonds but lacking the conserved sequence motifs of cystatins. Furthermore, another family of inactive cystatin-like proteins is represented by mammalian cathelins, the overall fold of which is indicated to be very similar to that of cystatins despite of a low degree of sequence identity (Kopitar *et al.*, 1989; Lenarcic *et al.*, 1993; Yang *et al.*, 2003). Inactive cystatin-like proteins may represent families that by divergence and/or gene duplication have evolved from the archetypal cystatins.

#### *Other protein and peptide inhibitors of cysteine proteases of the papain family*

A number of other proteins exhibit inhibitory activity against lysosomal papain-like cysteine proteases.  $\alpha$ -Macroglobulins unspecifically trap endopeptidases of different types, blocking the access of protein substrates to the active site of the trapped proteinases without inactivating them (Sottrup-Jensen, 1989; Armstrong, 2001). Thyropins, which are homologous to the thyroglobulin type-1 domain, are another class of inhibitors (Lenarcic & Bevec, 1998; Lenarcic & Turk, 1999). Cathepsin L is specifically inhibited by a mammalian representative of this class, the major histocompatibility complex (MHC) class II-associated p41 invariant chain fragment (Bevec *et al.*, 1996; Guncar *et al.*, 1999). The serpin, squamous cell carcinoma antigen 1, has also been shown to inhibit proteases of the papain family (Takeda *et al.*, 1995; Schick *et al.*, 1998). Cytotoxic T-lymphocyte antigen-2 beta (CTLA-2 beta), which is 36% identical to the proregion of mouse cathepsin L, has been shown to inhibit cathepsin L-like proteases (Delaria *et al.*, 1994). CTLA-2 beta has been suggested to represent a novel class of cysteine protease inhibitors. In addition, peptides constituting the N-terminal proregions that are cleaved off on activation of proforms of cysteine proteases may act as potent reversible inhibitors of the mature proteases (Cygler & Mort, 1997).

Chagasin from *Trypanosoma cruzi* has been shown to inhibit a number of papain-like proteases (Monteiro *et al.*, 2001). Chagasin-like proteins were found to be encoded by genomes of several eukaryotes, bacteria and archaea, and thus appear to be the first protein inhibitors of cysteine proteases identified in prokaryotes (Rigden *et al.*, 2002). Chagasins are the only protease inhibitors known that adopt an immunoglobulin-type fold (Rigden *et al.*, 2001).

### *Physiological functions of cystatins*

A major role in protecting the organism against the harmful activities of cysteine proteases, originally attributed to cystatins on the basis of their high inhibitory potential, has been supported by many experimental studies. Colocalization of cathepsins and cystatins as well as significantly altered levels of the inhibitors have thus been observed in many pathological states, including cancer, myoclonal epilepsy, cardiovascular disease, and several autoimmune diseases, such as asthma (Hopsu-Havu *et al.*, 1983a, b; Lah *et al.*, 1990; Friedrich *et al.*, 1999; Kos *et al.*, 2000a, b; Cimerman *et al.*, 2001; Locsey *et al.*, 2001; Riccio *et al.*, 2001; Zore *et al.*, 2001; Colella *et al.*, 2002). In general, cystatin concentrations are elevated in patients suffering from inflammatory diseases and autoimmunity (reviewed in Barrett, 1986; Hansen *et al.*, 2000; Mangge *et al.*, 2000). In contrast, development of cancer may be promoted by expression of cystatins in lower levels or as less active isoforms (Lah *et al.*, 1992a, b; Zore *et al.*, 2001; Zajc *et al.*, 2002). Cystatin A was thus not detected in a number of epimeroid carcinomas and prostate adenocarcinomas, contrary to in normal tissues (Rinne *et al.*, 1984a, b; Söderström *et al.*, 1995; Mirtti *et al.*, 2003), and cystatins B and E/M were found to be downregulated in prostate and esophageal carcinomas and breast cancer, respectively (Sotiropoulou *et al.*, 1997; Shiraishi *et al.*, 1998; Mirtti *et al.*, 2003). Such lower concentrations of the inhibitors of harmful cathepsins associated with cancer progress may promote extracellular matrix destruction and tumor invasion (Friedrich *et al.*, 1999; Yano *et al.*, 2001). In agreement with this proposal, cystatin C, the most potent and abundant of human cystatins, was shown to prevent tumor cell invasion and metastasis both *in vitro* and *in vivo* (Kolkhorst *et al.*, 1998; Cox *et al.*, 1999b). Nevertheless, increased levels of expression of cystatins by malignant cells have also been reported (Nishida *et al.*, 1984; Calkins & Sloane, 1995; Ebert *et al.*, 1997; Mirtti *et al.*, 2003;). Such elevated extracellular levels of cystatins A, B and C in melanoma and colorectal cancer patients strongly correlated with shorter survival (Kos *et al.*, 2000a, b). Similarly to cancer progression, psoriasis was proposed to be associated with a decreased inhibitory ability of cystatin A produced by epidermal cells (Ohtani *et al.*, 1982).

Antiviral and antibacterial activities of cystatins that could be related to the inhibition of exogenous proteases have also been demonstrated. The growth of polio and corona virus in cultured human cells could be retarded by addition of chicken cystatin or cystatins C or D (Korant *et al.*, 1985; Cimerman *et al.*, 1996; Collins & Grubb, 1998). Moreover, peptide inhibitors with sequences mimicking the inhibitory center of cystatin C suppressed streptococcal infections in mice and the replication of corona and herpes simplex virus in cell cultures (Björck *et al.*, 1989, 1990; Kasprzykowski *et al.*, 2000). In addition, growth of *Porphyromonas gingivalis* and *Staphylococcus aureus* was inhibited by cystatins A and S (Takahashi *et al.*, 1994; Naito *et al.*, 1995), and the recently discovered cystatin 11 was similarly indicated to have antimicrobial activity against *Escherichia coli* (Hamil *et al.*, 2002). Cystatins were also found to inhibit cysteine proteases from the parasites *Trypanosoma cruzi*, *Trypanosoma congolense* and *Trypanosoma brucei brucei* (Stoka *et al.*, 1995; Troeberg *et al.*, 1996; Chagas *et al.*, 1997).

Furthermore, phycocystatins have been indicated to be insecticides, inhibiting digestive papain-like cysteine proteinases required for insect feeding and plant viral infection (Aoki *et al.*, 1995; Urwin *et al.*, 1995b; Koiwa *et al.*, 2000).

Apart from a general protective function, regulatory roles exerted by cystatins appear also to be essential for normal development and functioning of the organism in many cases. In particular, cystatin A was proposed to be involved in differentiation of the epidermis (Rinne *et al.*, 1984a; Järvinen *et al.*, 1987). Similarly, cystatin B was implicated in the control of myofibril turnover in the cytoplasm of muscle fibers (Spanier & Bird, 1982). In addition, cystatins were suggested to participate in regulation of cell proliferation (Hiwasa *et al.*, 1995), as overexpression of cystatin A was found to induce increased levels of cyclin B1, which is involved in cell-cycle regulation. Furthermore, cystatin C was shown to be involved in local and precise balancing of the activity of testicular cathepsin L, restricting the action of the enzyme to a specific step of spermatozoa maturation (Peloille *et al.*, 1997). A regulatory function was also proposed for salivary cystatins. Besides protecting periodontal tissues and presumably exerting antibacterial and antiviral activity (Abrahamson *et al.*, 1997), these cystatins may modulate the mineralization process at the saliva-enamel interface (Tseng *et al.*, 2000). *Cres* protein was also implied to be a regulatory protein, controlling prohormone and proprotein processing within the reproductive and neuroendocrine systems (Cornwall *et al.*, 2003). Similarly, cystatin F appeared to play a role in immune regulation through inhibition of a unique target in the hematopoietic system (Halfon *et al.*, 1998). Interestingly, cystatins isolated from parasites have shown immunomodulatory activities affecting host immune systems (Manoury *et al.*, 2001; Vray *et al.*, 2002). Nippocystatin secreted by an intestinal nematode, *Nippostrongylus brasiliensis*, was thus implicated in modulation of antigen processing in antigen-presenting cells in mice (Dainichi *et al.*, 2001). Furthermore, filarial nematode cystatins were reported to be efficient in upregulating nitric oxide production by macrophages in mice, a mediator known to be an effector molecule against filarial worms (Hartmann *et al.*, 2002). In addition, cystatins may have other functions, non-related to their cysteine-protease inhibitory activity. Thus, carp ovarian cystatin was found to be important for preventing polyspermy in carp eggs, acting via electrostatic interaction as an agglutinating factor of spermatozoa (Wang & Huang, 2002).

Certain pathological conditions are related to congenital cystatin mutations. A hereditary form of human epilepsy, progressive myoclonus epilepsy (EPM1) or Unverricht-Lundborg disease, was found to be associated with either of several mutations in the gene for cystatin B (Pennacchio *et al.*, 1996; Lalioti *et al.*, 1997a, b; Alakurtti *et al.*, 2000). In addition, knock-out of the cystatin B gene in mice resulted in development of symptoms similar to those in the human disorder and in a broad neuronal atrophy due to apoptosis (Pennacchio *et al.*, 1998; Shannon *et al.*, 2002). A recent study provided evidence that EPM1 pathogenesis is linked to an unbalanced cysteine protease activity due to the lack of cystatin B inhibitory activity (Kinne *et al.*, 2002). Significantly increased levels of cathepsins B, L and S activities and highly reduced levels of cystatin B mRNA were thus detected in lymphoblastoid cells of patients suffering from EPM1. In addition, a multiprotein complex of cystatin B with several non-protease proteins

has been implicated to have a specific cerebellar function (Di Giamo *et al.*, 2002). Loss of this function as a consequence of the absence of cystatin B was proposed also to contribute to EPM1 development. Another congenital defect is responsible for hereditary cystatin C amyloid angiopathy (HCCAA). This human disease is caused by a point mutation, Leu68 to Gln, in cystatin C, the inhibitor variant being deposited in amyloid fibrils in the cerebral arteries. Such amyloid formation leads to brain hemorrhage in young adults, resulting in paralysis, dementia and eventual death (Abrahamson *et al.*, 1992; Olafsson *et al.*, 1996). The L68Q mutation was found to decrease the stability of the cystatin C molecule, this instability being a critical factor for the development of the amyloidosis in HCCAA patients (Gerhartz *et al.*, 1998; Gerhartz & Abrahamson, 2002). Cystatin C has also been shown to be colocalized with amyloid deposits in brain arterioles in patients with Alzheimer's disease or with hereditary cerebral hemorrhage associated with amyloidosis-Dutch type (Wattendorff *et al.*, 1982; van Duinen *et al.*, 1987; Vinters *et al.*, 1990). However, the role of cystatin C in these disorders is unknown. No mutations in the cystatin C gene in patients with Alzheimer's disease-related cerebral amyloid angiopathy were found (Nagai *et al.*, 1998). Notably, however, in the neurons most susceptible to cell death in Alzheimer's disease, the expression of cystatin C was elevated and cystatin C distribution was limited and colocalized with that of cathepsin B (Deng *et al.*, 2001).

### *Three-dimensional structure of cystatins and mode of inhibition of papain-like cysteine proteases by cystatins*

#### Overall fold of free cystatins

Three-dimensional structures of several cystatins of families 1 and 2, either free or in complex with a target protease, have been determined by X-ray crystallography or NMR spectrometry. The structures of free chicken egg-white cystatin, free human cystatin A, human cystatin B in complex with papain and human cystatin A in complex with cathepsin H are available to date (Bode *et al.*, 1988; Stubbs *et al.*, 1990; Dieckmann *et al.*, 1993; Martin *et al.*, 1995; Tate *et al.*, 1995; Craven *et al.*, 2000; Jenko *et al.*, 2003;). In addition, a partially solved NMR structure of free human cystatin C has been presented (Ekiel *et al.*, 1997). The rather compact globules of all these cystatins share a common global fold, consisting of an antiparallel, five-stranded  $\beta$ -sheet wrapped around a central, five-turn  $\alpha$ -helix (Fig. 4). The evolutionarily conserved sequence motifs discussed above, which are distant in the primary structure, are adjacent along one edge of the natively folded molecule. The highly conserved Gly is located in an N-terminal segment that has no direct contacts with the main body of the free cystatin. The QxVxG and LP or, alternatively, PW regions are found in two loops connecting the second and third  $\beta$ -strands and the fourth and fifth  $\beta$ -strands, respectively. These loops in the cystatin molecule are often referred to as the first and second hairpin loops. In most three-dimensional structures of cystatins, with the exception of one structure of free cystatin A discussed later, the first hairpin loop is flanked by the N-terminal segment and the second loop. The NMR structure of free chicken cystatin and the most reliable NMR structure of free cystatin A (see below) have revealed that the N-terminal segment, at least up to the conserved Gly, has no stable regular conformation in solution and is highly flexible (Dieckmann *et al.*, 1993; Martin *et*



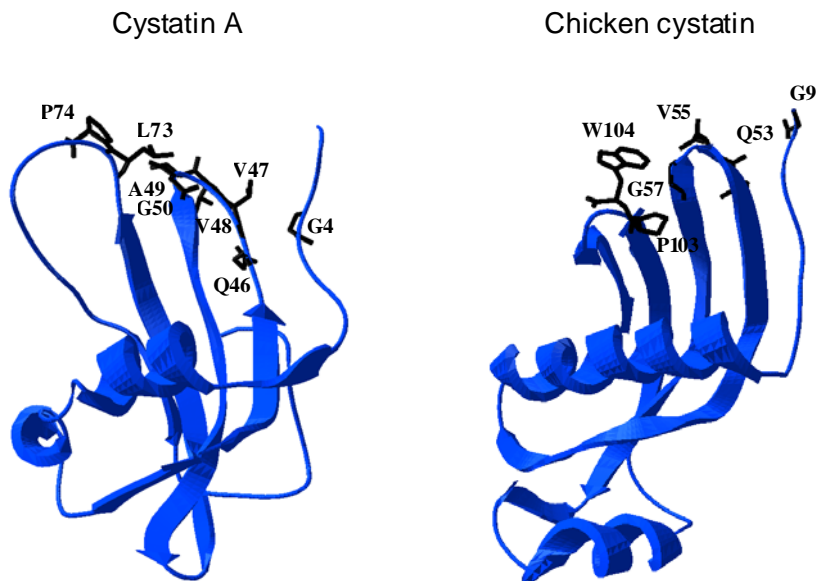


Fig. 4. Ribbon representation of the three-dimensional structures of cystatin A of family 1 and chicken egg-white cystatin of family 2. The side-chains of the residues conserved in the respective family are shown. The structures were drawn with the use of the Swiss-PdbViewer program (url: <http://www.expasy.ch/spdbv/>) and are based on the minimized averaged NMR structure of human cystatin A (Martin *et al.*, 1995; PDB code 1DVC) and the X-ray crystal structure of chicken egg-white cystatin truncated by eight residues from the N-terminus (Bode *et al.*, 1988; PDB code 1CEW). The two structures are in somewhat different orientations to allow an optimal view of the side-chains.

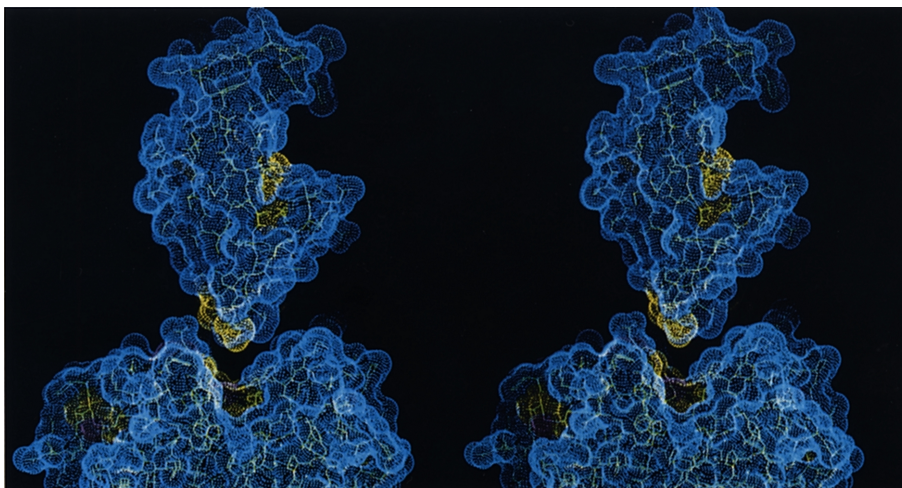
*al.*, 1995). Moreover, the first hairpin loops in the two structures are similar, tight and almost rigid. The second hairpin loop of chicken cystatin is also tight, with a well-defined conformation, but the corresponding loop in cystatin A, being two residues longer, is wide and appreciably mobile (Dieckmann *et al.*, 1993; Martin *et al.*, 1994, 1995). Another notable difference between the conformations of the two cystatins is the longer C-terminal region of cystatin A, which runs along the convex face of the  $\beta$ -sheet. This C-terminal segment of eight residues is conformationally restricted in cystatin A, although it does not adopt any regular structure. In addition, there are minor differences in the orientation of the central  $\alpha$ -helices of the two cystatins, this helix in cystatin A being almost perpendicular to the direction of the  $\beta$ -strands (Martin *et al.*, 1995). These structural differences presumably are characteristic of family 1 and 2 cystatins in general.

The X-ray structure of free chicken cystatin (Bode *et al.*, 1988) is generally similar to that obtained by NMR measurements (Dieckmann *et al.*, 1993) but differs substantially from the latter in the region comprising residues 69 to 90, which is absent in family 1 cystatins. In the X-ray structure, this region contains an additional  $\alpha$ -helix, whereas it is more disordered in the NMR structure (Engh *et al.*, 1993). This difference may be due to crystal packing effects. No evidence for an  $\alpha$ -helical conformation in the corresponding region of the related family 2 inhibitor, human cystatin C, was obtained by NMR (Ekiel *et al.*, 1997).

The cystatin A solution structure was originally elucidated by NMR methods by two different groups, although with two slightly different forms of the protein (Martin *et al.*, 1995; Tate *et al.*, 1995). The two structures have the same general fold, but the specific conformations of several regions differ significantly. The structure of authentic cystatin A, obtained at pH 5.5, agrees well with the structures of free chicken cystatin and cystatin C and, in particular, with that of the closely related cystatin B in complex with papain (Stubbs *et al.*, 1990; Martin *et al.*, 1995). However, the structure of an M65L cystatin A variant lacking the N-terminal Met, which was determined at pH 3.8, deviates substantially from the other cystatin structures, in particular in the N-terminal region and the second hairpin loop (Tate *et al.*, 1995). However, in a recent NMR study (Craven *et al.*, 2000) of both wild-type cystatin A and the M65L mutant, no differences between the structures of the two forms were found. Both structures were in accord with the data of the first report for wild-type cystatin A (Martin *et al.*, 1995) and with the other cystatin structures. Moreover, the recent investigation demonstrated that neither the M65L mutation and the lack of the N-terminal Met, nor the different pH-values of the analyses, could have been responsible for the unique structural features of  $\Delta$ M/M65L-cystatin A observed by Tate *et al.* (1995). Thus, the common fold of cystatins, including the similar structure in the regions containing the conserved sequence motifs, now appears well established.

#### Model of the interaction of cysteine proteases with protein inhibitors of the cystatin superfamily

A general model for the inhibition of cysteine proteases by cystatins, often called the "trunc model", was originally proposed on the basis of computer docking experiments with the structures of chicken cystatin, N-terminally truncated by eight residues, and papain solved by X-ray crystallography (Bode *et al.*, 1988, 1990; Machleidt *et al.*, 1989). The N-terminal region and the first and second hairpin loops of the cystatin were seen to be arranged in a wedge that was highly complementary to the active-site cleft of papain (Fig. 5). According to the model, papain is inhibited by this tripartite wedge entering into the active-site cleft of the protease, thereby preventing binding and cleavage of substrates. The interaction was concluded to be predominantly hydrophobic, with no need for conformational rearrangement of either of the two molecules. In the model, the two cystatin hairpin loops ("the binding loops") made extensive interactions with papain residues that line the wide, more open part of the active-site cleft and are relatively remote from the catalytic Cys25. The residues in papain interacting with the two loops are highly conserved in all related cysteine proteases and represent S' substrate binding subsites (Fig. 1). The N-terminal segment of the cystatin was proposed to make contacts with the residues in the narrow, bottom part of the papain active-site cleft. The two residues preceding the conserved Gly in the N-terminal region were implicated to bind to the S2 and S3 binding subsites of papain, although in a manner placing the peptide bond between the subsequent pair of residues in an inappropriate position for cleavage. Such binding in a substrate-like manner was concluded to be possible by the N-terminal residues of chicken cystatin from Leu7 to Ala10 forming a tight turn, allowed by the conserved Gly.



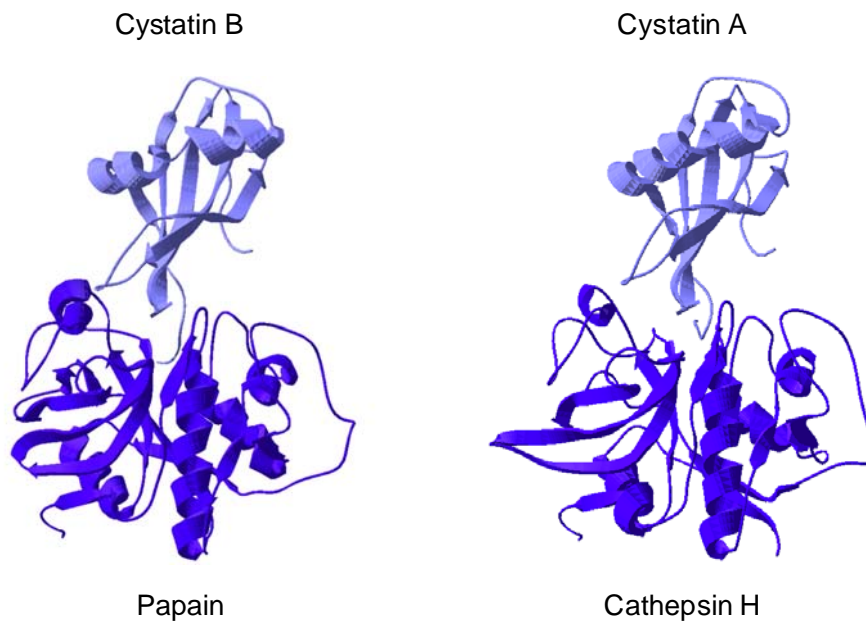
*Fig. 5.* Stereo image demonstrating the shape complementarity between the inhibitory wedge of chicken egg-white cystatin (top) and the active-site cleft of papain (bottom). (Reprinted, with permission, from Bode *et al.* (1988). Copyright 1988, The EMBO Journal.).

Although this model was proposed to be generally valid for cystatin inhibition of target proteases, the interaction of cystatins with exopeptidases, such as cathepsins B, H and X, would be expected to differ somewhat. The active-site cleft of these proteases is partially blocked by either the occluding loop, the mini-chain or the mini-loop, respectively (Musil *et al.*, 1991; Guncar *et al.*, 1998; Nägler *et al.*, 1999), which presumably would interfere with cystatin binding. Docking experiments aimed at modeling the complexes of cathepsin B with chicken cystatin or cystatin B indeed revealed that the second binding loop of either inhibitor would collide with the occluding loop of the protease in the complex (Musil *et al.*, 1991). This steric hindrance could be relieved by simultaneous tilting and rotation of either inhibitor, although the fit achieved was not as perfect as in the complexes with papain. In cathepsin H, the mini-chain, which restricts substrate binding to subsites amino-terminal of the S1 pocket, was similarly predicted to clash with the N-terminal region of cystatin B (Guncar *et al.*, 1998). Analogously, substrate docking experiments with the crystal structure of cathepsin X showed that the presence of the mini-loop substantially decreases the size of the S1' and S2' binding subsites of this protease (Nägler *et al.*, 1999; Sivaraman *et al.*, 2000). The modeling suggested that the restricted S2' subsite was the main reason for the lower inhibitory activity of cystatin C against cathepsin X than against endopeptidases. These observations, together with the still reasonably high affinities of cystatins for exopeptidases, indicate that cystatin binding to such enzymes might involve more dramatic changes in protein structure than the binding to endopeptidases.

### Three-dimensional structures and intermolecular interactions of cystatins in complexes with target proteases

The crystal structures of complexes of a C3S variant of human cystatin B with a carboxymethylated form of the endopeptidase, papain, and of human cystatin A with the aminopeptidase, cathepsin H, confirm the general model of inhibition of cysteine proteases by inhibitors of the cystatin superfamily and provide further insight into the binding (Stubbs *et al.*, 1990; Jenko *et al.*, 2003). The two structures thus verify that the complexes are formed by insertion of the narrow inhibitor wedge composed of the N-terminal region and the two binding loops into the V-shaped active-site crevice of the protease (Fig. 6). Cystatin B in complex with papain (Stubbs *et al.*, 1990) retains the overall cystatin backbone topology, although certain minor structural alterations are observed, especially in the protease binding regions. The structure of the complex reveals ~130 intermolecular interactions <4 Å in the interface region between cystatin B and papain, which are predominantly hydrophobic. A number of direct polar contacts and possible contacts mediated via solvent molecules also contribute to the intermolecular interaction. The N-terminal region and the first hairpin loop of the inhibitor provide the major number of interactions with the protease, the second binding loop being of minor importance. In addition, the C-terminal region of cystatin B contributes further interactions with the enzyme. The N-terminal region residues Met2–Ala5 form a rather open type II turn, which is in agreement with the predicted turn in the N-terminal segment of chicken cystatin in the model of the complex with papain (Bode *et al.*, 1988). This conformation of the N-terminal region has not been seen in free cystatins in solution and therefore presumably is induced by the interaction with papain. The conformation of the first binding loop of cystatin B in the complex is very similar to that of this loop in the free inhibitors analyzed, with the exception of the crystal structure of chicken cystatin, possibly due to crystal contacts having influenced the latter structure (Bode *et al.*, 1988; Stubbs *et al.*, 1990; Dieckmann *et al.*, 1993; Martin *et al.*, 1995). The reactive thiol group of Cys25 of papain in the complex is completely enclosed by Ser3 and Gly4 in the N-terminal region and Gln46 and Val 47 in the first binding loop of the inhibitor. The second binding loop of cystatin B has a helix-like conformation in the complex that is not found in the NMR structures of the related cystatin A in free form. This observation indicates that this binding loop, in addition to the N-terminal region, undergoes a certain structural adaptation to be accommodated in the complex. The second binding loop cystatin B interacts with the protease mainly through Leu73 and His75.

The recently published X-ray structure of human cystatin A in complex with porcine cathepsin H provides the first structural data on the binding of an exopeptidase by a cystatin (Jenko *et al.*, 2003). As in the complex of cystatin B with the endopeptidase, papain, both proteins in the complex exhibit the same overall fold as in the free state. However, the conformation of the binding sites of the two proteins, in particular the organization of the active-site region in cathepsin H, are appreciably altered in the complex. Notably, the mini-chain of cathepsin H is pushed away from its position in the free protease, as predicted by the modeling discussed above, and the so called insertion loop of the enzyme is



*Fig. 6.* Ribbon representation of the crystal structures of the complexes between cystatin B and papain (Stubbs *et al.*, 1990; PDB code 1STF) and between cystatin A and cathepsin H (Jenko *et al.*, 2003; PDB code 1NB3). Proteases are in black and cystatins in grey. The view is along the protease domain interface, with the R- and L-domains at the left and right, respectively. The structures were drawn with the use of the Swiss-PdbViewer program.

distorted by binding of the cystatin A N-terminal segment. The structure further illustrates how the flexible N-terminal segment of free cystatin A is adapted to the presence of the cathepsin H mini-chain in the complex. The four N-terminal residues of the inhibitor are thus bent in a short turn, with Ile2 and Pro3 occupying the positions of two C-terminal residues of the mini chain in the free protease. The first binding loop of cystatin A in the complex with cathepsin H has essentially the same conformation as that of cystatin B bound to papain. Unfortunately, however, the second binding loop of cystatin A is poorly resolved in the structure of the complex. Remarkably, the binding loops of cystatin A are inserted into the cathepsin H active-site cleft up to 0.8 Å deeper than the corresponding loops in the cystatin B—papain complex. Moreover, there are no water-mediated contacts in the interface between cystatin A and cathepsin H. Such a tighter fit may be a bona-fide feature of the inhibition of papain-like proteases by cystatins that was obscured by the carboxymethylation of papain in the previous study (Stubbs *et al.*, 1990).

#### *Contributions of individual binding regions of cystatins to binding of target proteases*

Equilibrium and kinetic investigations of the inhibition of cysteine proteases by cystatins have provided further evidence for the validity of the docking model for the binding proposed by Bode *et al.* (1988) and have also elucidated the

specificity and mechanism of the interactions. Cystatins have thus been shown to bind a number of proteases in tight equimolar complexes, often with association rates close to those of diffusion-controlled reactions (Björk *et al.*, 1989, 1994; Lindahl *et al.*, 1992a, b; Björk & Ylinenjärvi, 1990; Olsson *et al.*, 1999; Pol & Björk, 2003). In general, cystatins have been found to be rather non-specific inhibitors that virtually distinguish only endopeptidases and exopeptidases, the affinities of cystatins for endopeptidases being  $\sim 10^3$ – $10^6$ -fold higher than those for exopeptidases. During the last decade, the roles of the individual binding regions of cystatins, as well as the contribution of the conserved residues within these regions, have also been extensively investigated by comparative studies of natural inhibitors and their truncated or mutated variants.

#### The N-terminal region

Characterization of the binding of several plant and mammalian cysteine proteases by the family 2 cystatins, chicken cystatin and cystatins C and D, and by their N-terminally truncated forms has elucidated the importance of the N-terminal region for the inhibitory function of cystatins of this family (Abrahamson *et al.*, 1991; Lindahl *et al.*, 1992b; Hall *et al.*, 1993, 1998; Björk *et al.*, 1994). In particular, the N-terminal segments of chicken cystatin and cystatin C have been shown to account for  $\sim 40$  % of the total free energy of the binding of the inhibitors to papain (Machleidt *et al.*, 1991; Lindahl *et al.*, 1992b; Björk *et al.*, 1994). However, several studies have revealed that the contribution of the N-terminal segment of cystatins to the binding varies with the target protease. Thus, this region has been found to be of much lesser importance for the binding of cystatin C to cathepsin H than to cathepsins B and L (Abrahamson *et al.*, 1991; Hall *et al.*, 1995). In contrast, removal of the N-terminal segment of cystatin D abolishes the inhibitory activity of this cystatin towards all cathepsins assayed (Hall *et al.*, 1998). The N-terminal segment and the remainder of the binding region of cystatin C have been shown to contribute to the energy of binding of cysteine proteases in a nearly additive fashion (Lindahl *et al.*, 1992b; Björk *et al.*, 1994; Hall *et al.*, 1995, 1998).

The N-terminal region of family 2 cystatins has been concluded to bind in a substrate-like manner into the substrate-binding pockets of the cysteine proteases studied (Grubb *et al.*, 1990; Abrahamson *et al.*, 1991; Hall *et al.*, 1993; Björk *et al.*, 1994; Lindahl *et al.*, 1994), in agreement with the docking model and the structure of the cystatin B—papain complex (Bode *et al.*, 1988; Stubbs *et al.*, 1990). The two residues preceding the conserved Gly of chicken cystatin and human cystatin C have been shown to account for the major part of the binding energy of the entire N-terminal segment, the residue adjacent to the Gly contributing most (Lindahl *et al.*, 1992b; Hall *et al.*, 1995). These two residues are assumed to interact with the S3 and S2 substrate-binding subsites of the proteases, as the analogous residues of cystatin B do in the complex with papain, and are thus termed the P3 and P2 residues of the inhibitor. In addition to provide major binding energy, these residues have been shown also to confer biological selectivity to the inhibition of different endogenous cathepsins (Hall *et al.*, 1995, 1998; Mason *et al.*, 1998). In particular, the side chain of Leu9 of cystatin C (in the P3 position) makes the inhibitor selective for cathepsins B and L but is

unfavorable for binding of cathepsin H (Hall *et al.*, 1995). Moreover, the different P2 residues of the closely homologous cystatins C and D have been suggested to be responsible for the different inhibitory profiles of the two cystatins (Hall *et al.*, 1998). In agreement with these observations, site-directed mutagenesis of the P3 and P2 residues of family 2 cystatins can alter both the inhibitory potency and the specificity of the inhibitors (Lindahl *et al.*, 1994; Mason *et al.*, 1998). The P4 residue expected to contact the putative S4 subsite is apparently of minor importance for the binding of endopeptidases by cystatin C but may be significant for cathepsin B inhibition (Hall *et al.*, 1995; Cimerman *et al.*, 1999).

In contrast to the results for family 2 cystatins, the published data on the importance of the N-terminal region of family 1 cystatins for the inhibition of cysteine proteases have been discrepant. Early work indicated that this region of the family 1 cystatins is not important for the inhibition and questioned a substrate-like binding of the N-terminal residues of these cystatins to proteases (Thiele *et al.*, 1988, 1990; Machleidt *et al.*, 1991). However, a number of later studies provided evidence that this region plays a substantial role in the binding (Shibuya *et al.*, 1995a, b; Estrada *et al.*, 1998). Moreover, in addition to Paper I of this thesis, more recently published work on cystatin B unequivocally established the critical role of residues at the N-terminus of family 1 cystatins in the high-affinity binding of cysteine proteases (Pol & Björk, 2001, 2003). In addition, these studies established the primary importance of the residue preceding the conserved Gly also for family 1 cystatins. The apparent contradictions between the results of the later and the previous investigations are most certainly due to difficulties in quantifying the high affinities of cystatins for their target proteases. Moreover, irreversible oxidation of the P2 residue, Cys3, of cystatin B in the earlier studies could also have caused an underestimation of the role of the N-terminal segment of this inhibitor (Pol & Björk, 2003). Notably, a study subsequent to Paper I found that different tripeptides at the N-terminus of cystatin A may contribute equally well as the wild-type sequence to the high-affinity binding of papain (Ylinenjärvi *et al.*, 1999).

Several investigations have focused on the role of the evolutionarily conserved Gly in the N-terminal region of cystatins. This Gly in cystatins A and C, of families 1 and 2, as well as that in oryzacystatin-I, a plant cystatin from rice, has been substituted by a number of bulkier residues by site-directed mutagenesis, and the affinities and kinetics of inhibition of different cysteine proteases by the mutant inhibitors have been studied (Hall *et al.*, 1993; Björk *et al.*, 1995; Shibuya *et al.*, 1995b; Urwin *et al.*, 1995a; Estrada *et al.*, 1998). In general, the affinities were found to decrease proportionally to the size of the side chain introduced. On the basis of these results, the conserved Gly was deduced to be a crucial residue for the high protease-binding affinity of cystatins, functioning as a hinge that allows the N-terminal region to adopt to the binding pockets of proteases. This conclusion is in agreement with the conformation of the N-terminal region in the X-ray structure of the cystatin B—papain complex. However, the relative importance of a high flexibility of the N-terminal segment and of the absence of a side chain at the position of the conserved Gly for the binding has been shown to vary with the protease (Estrada *et al.*, 1998; Tseng *et al.*, 2000). In support of the crucial role of the evolutionarily conserved Gly4 in cystatin A, mutant forms of this inhibitor

with high affinity for papain, obtained by random phage-display mutagenesis of the four N-terminal residues, predominantly contained Gly in position 4 (Ylinenjärvi *et al.*, 1999).

Comparisons of the kinetics of binding of different cysteine proteases by the family 2 cystatins, chicken cystatin and cystatin C, have indicated that the N-terminal regions of these inhibitors contribute to the binding in different ways, depending on structure of the active-site region of the proteases (Björk *et al.*, 1994, 1995, 1996). The N-terminal segment was found to interact with endopeptidases at a rate apparently identical with that of the remainder of the binding region and to be essential for stabilizing the complexes by anchoring the protease to the inhibitor, thereby reducing the dissociation rate. In the inhibition of the exopeptidase, cathepsin B, however, the N-terminal region was shown to contribute to the high affinity by increasing the rate of association with the protease. It was therefore suggested that the N-terminal segment of the two cystatins interacts with cathepsin B in an initial binding step, thereby facilitating subsequent docking of the remaining protease binding site into the active-site crevice. The N-terminal region of the cystatins might thus promote a proper orientation of the two binding loops for optimal complex formation with cathepsin B. Later work on papain, cathepsin L and cathepsin B inhibition by Gly4-mutants of cystatin A implicated similar different roles of the N-terminal regions of family 1 cystatins in inhibition of endopeptidases and cathepsin B (Estrada *et al.*, 1998).

#### The first binding loop

An important role of the first binding loop of family 2 cystatins in the inhibition of papain-like proteases appears to be well established. The first binding loop of chicken cystatin and human cystatin C has thus been concluded to account for ~40–60% of the total energy of the binding of the inhibitors to papain and cathepsin B (Machleidt *et al.*, 1991; Björk *et al.*, 1996). Moreover, a number of equilibrium and kinetic studies with recombinant variants of cystatins have shown the importance of residues in the conserved QxVxG motif in this loop for the binding. In particular, the point mutations Q53E, Q53N and G57A, as well as several double substitutions within the loop, in chicken cystatin led to ~10–1000-fold reductions in the affinities for papain, actinidin and cathepsin B (Auerswald *et al.*, 1992). Deletion of Val55 in the first binding loop of chicken cystatin also resulted in a lower affinity for papain and cathepsins L and B, and deletion of the entire loop in cystatin SN abolished the inhibitory activity toward papain (Auerswald *et al.*, 1995; Hiltke *et al.*, 1999). Furthermore, removal of the side chains of the first three residues in the first binding loop of cystatin SN by Gly substitutions significantly reduced the binding affinities for papain and cathepsin C (Tseng *et al.*, 2000). Strong evidence for the importance of the conserved QVVAG motif in the first binding loop of phytocystatins for the effective interaction with papain has also been provided by phage display studies (Koiwa *et al.*, 2001). All soyacystatin variants selected to be functional in binding to papain from a library in which the first and second binding loops were randomly mutated thus invariantly had the QVVAG sequence in the first binding loop.



Interestingly, some deletions or point mutations in the first binding loop of chicken cystatin were shown to lead to a distorted binding of the N-terminal region of the inhibitor to papain and cathepsin L, resulting in slow cleavage of the covalent bond between Gly9 and Ala10 and release of active enzyme, so called temporary inhibition (Auerswald *et al.*, 1995; Machleidt *et al.*, 1995). The appropriate sequence of the loop thus appears to be critical for optimal binding of the cystatin N-terminal segment.

The mutations in the first binding loop of chicken cystatin predominantly affected  $k_{\text{diss}}$ , indicating that this region primarily functions by stabilizing the cystatin—protease complexes (Auerswald *et al.*, 1992). It is apparent, however, that the first binding loop of family 2 cystatins contributes to varying extents to the inhibition of different proteases (Auerswald *et al.*, 1992, 1995; Hall *et al.*, 1995; Tseng *et al.*, 2000).

In contrast to the work on family 2 cystatins, the results of similar studies of family 1 cystatins have been inconsistent. Early work indicated the first binding loop of cystatins A and B to be insignificant for the inhibition of papain (Nikawa *et al.*, 1989; Jerala *et al.*, 1990). These results are in apparent contradiction with the crystal structures of the complexes of cystatins A and B with cysteine proteases that are now available, in which the first binding loop of the inhibitor is shown to be tightly bound into the protease active-site crevice (Stubbs *et al.*, 1990; Jenko *et al.*, 2003). Difficulties in measurements of very low  $K_{\text{d}}$  values presumably are the main reason for the underestimated role of the first binding loop of family 1 cystatins in the early work. Moreover, the choice of substitutions in this loop of cystatin B in one early study (Jerala *et al.*, 1990) could have allowed intermolecular interactions similar to those in the wild-type inhibitor, which would not have resulted in a detectable change of the affinity. In agreement with the X-ray structures, the first binding loop of cystatin B was later implicated to be involved in papain binding, as a V48D mutation in this loop lowered the affinity for this enzyme by 240-fold (Machleidt *et al.*, 1991).

#### The second binding loop

The importance of the second binding loop in cystatins of family 2 for the binding of papain-like proteases has been verified in a number of studies. Early work concluded that it contributes ~13 % of the total energy of complex formation between chicken cystatin and papain (Machleidt *et al.*, 1991). Moreover, deletion of the second binding loop in chicken cystatin was shown to result in reduced affinities for papain and cathepsins B and L (Auerswald *et al.*, 1995). Several investigations have established that the evolutionarily conserved Trp in this loop in the family 2 cystatins is particularly important for efficient inhibition. Early studies with chicken cystatin demonstrated that chemical modification of this Trp substantially lowered the affinity for papain and also implicated a change of the environment of the Trp on binding of proteases, indicating that this residue is involved in the interaction (Lindahl *et al.*, 1988; Nycander & Björk, 1990). Further work showed that removal of the side chain of Trp106 in the loop in human cystatin C by a Gly substitution dramatically decreased the affinities for papain, actinidin and cathepsins B and H, whereas the affinities for cathepsins L

and S were only moderately reduced (Hall *et al.*, 1995; Björk *et al.*, 1996; Mason *et al.*, 1998). However, replacement of Trp106 by Phe reduced the affinity to a lower extent than the Gly replacement, implying that the hydrophobic side chain of Phe can partly compensate for that of Trp in protease binding (Björk *et al.*, 1996). On the basis of these studies, it was concluded that Trp106 in the second binding loop of cystatin C accounts for ~20–30% of the energy of binding of the inhibitor to papain, actinidin and cathepsins B and H (Björk *et al.*, 1996). A subsequent investigation demonstrated that substitution of both residues of the conserved PW motif in the second binding loop of human salivary cystatin SN with a GG segment appreciably reduced cathepsin C inhibition, although no effect on papain inhibition was noticed (Tseng *et al.*, 2000). The significance of the conserved Trp in the second binding loop of phytocystatins for protease binding has also been shown by phage display selection of variants with high affinity for papain from the soyacystatin library with the two binding loops randomly mutated (Koiwa *et al.*, 2001). All selected variants thus had Trp in the second binding loop, whereas other residues, both hydrophobic and basic, were found N-terminally of this Trp.

The importance of the second binding loop of a family 1 cystatin, cystatin B, in the inhibition of different cysteine proteases has been shown in one study (Pol & Björk, 1999). Leu73 and His75 of this loop were demonstrated to be important for high-affinity binding of papain and cathepsins B and H, Leu73 being most essential. The two residues together were concluded to contribute similarly as the Trp in the second binding loop of cystatin C to papain and cathepsin H binding, *i.e.* ~20–30% of the total binding energy.

Kinetic studies of cysteine protease inhibition by natural and mutant forms of cystatins have revealed that the second binding loop of the inhibitors is primarily essential for the stabilization of the complexes with the proteases by maintaining a low  $k_{\text{diss}}$  (Björk *et al.*, 1996; Pol & Björk, 1999). However, varying contributions of the loop to the binding have been observed for different proteases, presumably reflecting a diversity of structures of and around the active-site clefts of the enzymes (Auerswald *et al.*, 1995; Hall *et al.*, 1995; Björk *et al.*, 1996; Mason *et al.*, 1998; Pol & Björk, 1999; Tseng *et al.*, 2000).

#### The C-terminal region

As mentioned earlier, cystatins of family 1 have a nine-residue long C-terminal extension, compared with cystatins of family 2 (Bode *et al.*, 1988). The X-ray structure of the complex between cystatin B and papain indicated the C-terminal region of this inhibitor to be involved in interactions with the L-domain of the protease (Stubbs *et al.*, 1990). One residue of this region, Tyr97, was deduced to contribute to stabilization of the complex by an extensive net of close (<4 Å), predominantly hydrophobic, contacts with papain (Stubbs *et al.*, 1990). Despite this observation, binding of papain by cystatin B was initially reported not to be affected by truncation of 11 C-terminal residues of the inhibitor (Jerala *et al.*, 1991). In contrast, in a more recent study, site-directed mutagenesis of Tyr97 to Ala verified the importance of this residue for tight anchoring of cystatin B to papain and also, although to a lesser extent, to cathepsins L and H (Pol & Björk,

1999). Tyr97 of cystatin B was shown to contribute ~6–12% to the total energy of binding to these proteases but to be of no importance for cathepsin B inhibition (Pol & Björk, 1999).

Tyr97 is not well conserved in family 1 cystatins, being replaced by Gly in the related inhibitor, cystatin A. Moreover, the crystal structure of the complex of cystatin A with cathepsin H suggests that the C-terminal end of the inhibitor is not essential for binding to this protease (Jenko *et al.*, 2003). It is therefore uncertain whether the C-terminal end of cystatin A is of any functional importance. Unfortunately, no experimental studies on the role of the C-terminal region of other family 1 inhibitors than cystatin B in the binding of cysteine proteases have been reported.

#### The legumain-binding region

Several family 2 cystatins inhibit mammalian legumain, as mentioned above. Lysosomal cysteine proteases of the papain and legumain families are evolutionarily unrelated, having different folds and active-site motifs (Chen *et al.*, 1998). The legumain binding site of cystatins with ability to bind this protease has been found to be located on the side of the cystatin molecule directly opposite to the wedge-shaped binding site for papain-like proteases. As a consequence, these cystatins are able to form ternary complexes containing one molecule of each of the two types of cysteine proteases (Alvarez-Fernandez *et al.*, 1999). The highly conserved Asn39 (in cystatin C numbering), located in the loop connecting the central  $\alpha$ -helix with the first long strand of the  $\beta$ -sheet of the cystatin, was shown to be a key residue for legumain inhibitory function. In addition, a proper conformation of this loop was suggested to be necessary for the inhibition (Alvarez-Fernandez *et al.*, 1999). Family 1 cystatins A and B, in which this loop is two residues shorter, as well as cystatin D of family 2, which has a one-residue insertion in the loop, thus do not show legumain inhibitory activity. Presumably, other regions of a cystatin capable of binding legumain are also involved in interaction with the enzyme (Alvarez-Fernandez *et al.*, 1999).

#### *Kinetic mechanism of the interaction between cystatins and their target proteases*

Binding of cystatins to target endopeptidases that have an open active-site cleft, such as papain and cathepsin L, occurs with rate constants approaching those expected for diffusion-limited reactions, which implies that there is no need for a global conformational adjustment of either of the interacting proteins on complex formation (Björk *et al.*, 1989, 1994; Björk & Ylinenjärvi, 1990; Lindahl *et al.*, 1992a, b; Turk *et al.*, 1994; Olsson *et al.*, 1999; Pol & Björk, 2003). Moreover, a simple, one-step mechanism of the interaction of cystatins with such proteases has been indicated by stopped-flow kinetics studies of the binding at high inhibitor concentrations (Björk *et al.*, 1989; Björk & Ylinenjärvi, 1990; Lindahl *et al.*, 1992b; Turk *et al.*, 1994; Pol *et al.*, 1995). These observations are in agreement with the docking model for complex formation based on the crystal structures of chicken cystatin and papain that was discussed earlier (Bode *et al.*, 1988). However, more refined rapid kinetics analyses of papain binding to

cystatin A with its N-terminus labeled with a fluorescent group have strongly implicated a two-step mechanism of the cystatin—papain interaction, with an independent binding of the N-terminal region of the inhibitor subsequent to that of the two binding loops (Estrada *et al.*, 2000). The rates of these two binding steps were concluded to be undistinguishable in conventional studies of the binding kinetics, accounting for the apparent one-step reaction observed in such studies (Estrada *et al.*, 2000).

The presence of the occluding loop on top of the cathepsin B active-site cleft was proposed to be responsible for the generally lower affinities of cystatins for this exopeptidase, compared with those for endopeptidases with open active sites (Musil *et al.*, 1991; Henskens *et al.*, 1996). The interference of the occluding loop with the interaction between the cystatin binding wedge and the cathepsin B active-site cleft, predicted by the computer docking experiments (see above), was confirmed by site-directed mutagenesis. An engineered partial deletion of the occluding loop of cathepsin B thus resulted in >40-fold higher affinity for cystatin C (Illy *et al.*, 1997). The occlusion of part of the substrate binding site by the loop results in a mechanism of cystatin binding to cathepsin B deviating from that of the binding to endopeptidases. As discussed above, the N-terminal region of cystatins was shown to be essential for a high rate of association of the inhibitors with cathepsin B, in contrast to the binding to endopeptidases. Furthermore,  $k_{\text{obs}}$  derived from stopped-flow kinetics analyses of the interaction of cystatin C with recombinant cathepsin B was shown to depend hyperbolically on inhibitor concentration (Nycander *et al.*, 1998). This behavior is compatible with a two-step interaction, with a conformational change in one or both proteins in the second, rate-limiting step (Fersht, 1985). On the basis of these results, Nycander *et al.* (1998) suggested that the N-terminal region of cystatin C interacts rapidly with the protease in the first step of the reaction, followed by a slower displacement of the cathepsin B occluding loop by the remainder of the binding region of cystatin C entering the active-site cleft of the enzyme.

## Present investigation

Before this investigation was started, the general features of the binding and inhibition of papain-like cysteine proteases by family 2 cystatins had been well established. In particular, several studies had elucidated the critical roles of three binding regions of these cystatins, *viz.* the N-terminal region and the first and second hairpin loops, in the inhibition. The importance and roles of the corresponding structural elements in cystatins of family 1 had not been studied as extensively and, moreover, discrepant data concerning these aspects had been reported. Thus, further work was required to clarify whether the main features of cystatin inhibition of target proteases are general for different cystatin families. To this end, the bulk of the work in this investigation was done with the family 1 cystatin, cystatin A. A major aim was to elucidate the specific roles of the N-terminal region and the second binding loop of cystatin A in the binding to different cysteine proteases (Papers I and II). Gaining a deeper understanding of the complicated mode of inhibition of cathepsin B, one of the most abundant mammalian proteases, by cystatins was another goal of this study (Papers III and IV). The relationship between the inhibitory abilities of cystatins and the structures of their individual binding regions was a further important aspect of the investigation (especially Paper IV).

### Contribution of cystatin A N-terminal residues to the inhibition of different cysteine proteases (Paper I)

Previous investigations had indicated the importance of the N-terminal region of the family 1 cystatin, cystatin A, for the high affinity of this inhibitor for target proteases (Takeda *et al.*, 1985; Pol *et al.*, 1995; Shibuya *et al.*, 1995a, b; Estrada *et al.*, 1998;). However, much of the data reported was uncertain due to experimental problems (see below) and concerned only papain inhibition. Only one study had unequivocally verified the importance of the evolutionarily conserved Gly4 to the binding of cystatin A to different target proteases (Estrada *et al.*, 1998). The work in Paper I elucidated the role of the flexible cystatin A N-terminal segment in the inhibition by characterizing the contribution of the first three N-terminal residues to the binding to papain, as well as to the more physiologically relevant proteases, cathepsins L and B. Three cystatin A mutants, lacking either one, two or three residues from the N-terminus ( $\Delta M$ -,  $\Delta MI$ - and  $\Delta MIP$ -cystatin A, respectively), were obtained by recombinant protein technology, and their inhibitory properties were examined by equilibrium and kinetic methods. Far-UV circular dichroism spectra of the three mutants did not indicate any conformational changes in cystatin A due to the deletions at the N-terminus. This finding was consistent with previous NMR studies of cystatin A that had revealed a high flexibility of the first five residues of the protein and a lack of interactions between these residues and the rest of the polypeptide chain (Martin *et al.*, 1995). Hence, deletions within this flexible segment would be expected to have no influence on the conformation of the rest of cystatin molecule. Therefore, any decreased affinity of the truncated forms of cystatin A for proteases should well

reflect the loss of the contributions of the corresponding N-terminal residues to the binding.

Sequential deletion of the three residues at the N-terminus led, in general, to a progressive decrease of cystatin A affinity for all three proteases studied, but the extent of this reduction varied substantially with the protease. The removal of Met1 had a minimal effect on the affinities for papain and cathepsin B, as well as on the corresponding binding kinetics. The truncation by two residues, Met1 and Ile2, resulted in the affinities for papain and cathepsin B being decreased by 900- and 200-fold, respectively. Unfortunately, however, less precise information could be obtained for cathepsin L. Since only upper limits of  $K_d$  of  $\Delta$ M- and wild-type cystatin A for this enzyme could be estimated, only the smallest possible increase in  $K_d$ , ~3-fold, caused by the truncation of the two initial residues, could be deduced. Despite these uncertainties, the deletion of the two initial residues appeared to affect cystatin A affinity for cathepsin L to a lesser extent than the affinity for the other proteases. Removal also of Pro3 had a major effect on the inhibition of all proteases studied.  $K_d$  increased by  $2 \times 10^6$ ,  $\geq 7 \times 10^4$ - and  $7 \times 10^4$ -fold for papain, cathepsin L and cathepsin B, respectively, in comparison with that of the wild-type inhibitor. This indicates that the MIP-segment is responsible for ~40% of the total free energy of binding of cystatin A to the proteases. This contribution to cathepsin B binding is appreciably higher than that of the corresponding, although five to seven residues longer, N-terminal regions of family 2 cystatins. However, the N-terminal segments of cystatin A and family 2 cystatins contribute similarly to papain binding (Björk *et al.*, 1994). As has been reported for family 2 cystatins, the two residues of cystatin A preceding the evolutionarily conserved Gly4, i.e. Ile2 and Pro3, in particular the latter, were thus found to contribute most of the N-terminal region to the binding energy. In some aspects, however, the results of the present work were discrepant with those reported earlier for cystatin A (Shibuya *et al.*, 1995a). In the previous work, the binding of cystatin A to papain was not affected by N-terminal truncations of the residues preceding Pro3, and the affinity of  $\Delta$ MIPG-cystatin A was only ~200-fold lower than that of the control variant (Shibuya *et al.*, 1995a). The apparent contradiction of these results with those of Paper I is most probably due to the inappropriate experimental approach used in the previous studies for measuring very low  $K_d$  values (Shibuya *et al.*, 1995a). Therefore, the conclusions regarding the role of residues in the N-terminal region of cystatin A in the binding to target proteases made in these studies on the basis of uncertain values of  $K_d$  are doubtful.

Analyses of the binding kinetics revealed different mechanisms behind the reduction in affinity for different proteases. The decreased affinities of the N-terminally truncated cystatin A variants for the two proteases with an open active-site cleft (papain and cathepsin L) were entirely due to increases in  $k_{diss}$ . A similar behavior had been previously observed for the binding of N-terminally truncated cystatins of family 2 (Björk *et al.*, 1994). These findings led to the suggestion that in the case of papain and cathepsin L, a cystatin A—protease complex is initially formed with unobstructed binding of the more rigid first and second binding loops into the open active-site cleft of these proteases. The flexible N-terminal segment may then bind in a second step and stabilize the complex. When this work was undertaken, no direct evidence in proof of such a two-step binding

of cystatins to proteases with an open active site was available (Björk *et al.*, 1989, 1990; Lindahl *et al.*, 1992b; Turk *et al.*, 1994; Pol *et al.*, 1995) but such evidence was obtained in later work (Estrada *et al.*, 2000). In Paper I, the N-terminal segment of cystatin A was proposed to bind to papain and cathepsin L in a substrate-like manner, with Ile2 and Pro3 contacting the S3 and S2 protease binding subsites, respectively (Stubbs *et al.*, 1990; Lindahl *et al.*, 1994; Hall *et al.*, 1995, 1998; Mason *et al.*, 1998).

By contrast, in the case of cathepsin B, in which the catalytic site is obstructed by an occluding loop, both  $k_{\text{ass}}$  and  $k_{\text{diss}}$ , in particular the former, were affected by the truncations. This observation indicated a mechanism of cathepsin B binding different from that of binding of proteases with an exposed active site. A previous study had demonstrated that cystatin C, of family 2, binds to cathepsin B in a two-step reaction and had indicated that the second step involves displacement of the occluding loop of cathepsin B to allow formation of a final, tight bimolecular complex (Nycander *et al.*, 1998). On the basis of these previous findings and the results of Paper I, it was suggested that cystatin A also binds to cathepsin B in an analogous two-step reaction. The flexible N-terminal segment of cystatin A was assumed to bind to this protease in the first step, thereby facilitating the subsequent displacement of the occluding loop in the second step (Björk *et al.*, 1994, 1995, 1996).

It should be added that the importance of the N-terminal residues of another family 1 cystatin, cystatin B, for the inhibition of a number of cysteine proteases has also been demonstrated later (Pol & Björk, 2001, 2003). The major proposals made in Paper I regarding the role of the flexible N-terminal segment of family 1 cystatins in the binding to different target proteases were verified for cystatin B and have been further corroborated in subsequent investigations (Estrada *et al.*, 2000; Pol & Björk, 2001, 2003; Paper IV).

## **Importance of the second binding loop of cystatin A for stabilization of the complexes with different target proteases (Paper II)**

Several studies had indicated an essential role of the second binding loop of cystatin B of family 1 and of cystatin C and chicken cystatin of family 2 in inhibition of papain-like cysteine proteases before this work was started (Nycander & Björk, 1990; Auerswald *et al.*, 1995; Björk *et al.*, 1996; Pol & Björk, 1999). However, certain structural peculiarities of the corresponding loop in cystatin A cast some doubts on the importance of this loop for tight binding of this inhibitor to target proteases. The loop in cystatin A thus lacks not only the Trp that is essential for the inhibition by family 2 cystatins but also the important His in cystatin B. In the X-ray crystal structure of the complex between cystatin B and papain, His75 in the second binding loop of the inhibitor accounts for a number of close contacts with papain (Stubbs *et al.*, 1990). Moreover, an important role of His75 in stabilizing complexes of cystatin B with different cysteine proteases has also been unambiguously demonstrated by mutagenesis experiments (Pol & Björk, 1999). However, in cystatin A this His is replaced by Gly. This

substitution would be expected to lead to loss of the contacts with the protease donated by the side chain of His75 in the cystatin B complex and, consequently, to a decreased contribution of the second binding loop of cystatin A to the binding. Furthermore, the high flexibility of this loop in cystatin A (Martin *et al.*, 1995) could either be detrimental to the binding or, conversely, promote the binding by allowing an orientation of the loop that is beneficial for the interaction with target proteases. These uncertainties concerning the role of the cystatin A second binding loop prompted the examination in Paper II of the contribution of side chains of individual residues within the most flexible part of the loop to the binding of different cysteine proteases. Four residues of the loop, Leu73, Pro74, Gln76, and Asn77 were concluded to be likely candidates contributing to the intermolecular interaction. Therefore, four recombinant cystatin A variants, each having a single residue within the loop substituted with Gly, were constructed by site-directed mutagenesis. The binding of the L73G-, P74G-, Q76G-, and N77G-cystatin A variants to papain, cathepsin L and cathepsin B was then studied by equilibrium and kinetic methods. The contribution of the remaining residue of the loop, Gly75, could not be investigated by the approach taken in this work.

Only replacement of Leu73, highly conserved in family 1 cystatins, with Gly affected the affinity of cystatin A for the proteases studied to any appreciable extent. The increase in  $K_d$  for papain and cathepsin B was ~320- and ~4000-fold, respectively. The other three mutations did not have any significant effect on cystatin A binding to papain and cathepsin B, with the exception of the P74G mutation, which resulted in  $K_d$  for the complex with cathepsin B being increased by ~10-fold. Unfortunately, the effects of the mutations could not be properly quantified in the case of cathepsin L, for which only a lower limit of the decrease in affinity due to the L73G mutation of ~10-fold could be estimated. No indications of any significant change in the affinity for this protease were observed on introduction of the three other mutations in cystatin A. The losses in binding affinity indicate that the side chain of Leu73 is responsible for ~18 and ~34% of the total unitary free energy change for binding of cystatin A to papain and cathepsin B, respectively. This contribution of Leu73 of cystatin A to the free energy of binding to papain is similar to that of Leu73 in the second binding loop of cystatin B (Pol & Björk, 1999) and to that of essential Trp106 in the second hairpin loop of cystatin C of family 2 (Björk *et al.*, 1996). However, in the case of cathepsin B inhibition, the contribution of Leu73 is substantially higher than that of Trp106 of cystatin C (Björk *et al.*, 1996). Together, Leu73 and Pro74 account for ~45% of the total energy of binding of cystatin A to the latter protease. In contrast to these two residues, Gln76 and Asn77 contribute negligibly to the binding of either enzyme. Taken together, the results manifest the importance of the second binding loop of cystatin A for the inhibition of different cysteine proteases, although the contributions of this loop vary with the protease, being especially pronounced for cathepsin B. It is noteworthy that the flexible N-terminal segment and the second binding loop of cystatin A together contribute almost 90% of the energy of binding of this protein inhibitor to cathepsin B, whereas the joint contribution of the corresponding regions of cystatin C appears to be appreciably smaller (Björk *et al.*, 1994, 1996; Paper I). This observation



implies that the first binding loop of cystatin A contributes appreciably less to cathepsin B binding than the corresponding loop of cystatin C.

Analysis of the binding kinetics revealed that  $k_{\text{ass}}$  was unaffected by the mutations and that the decreased affinities were exclusively due to an increased  $k_{\text{diss}}$ . These results indicate that the role of the second binding loop of cystatin A is to stabilize the complex between cystatin A and its target protease after this complex has already been formed. Hence, the role of this loop of cystatin A is the same as that of the corresponding loops of cystatins B and C (Björk *et al.*, 1996; Pol & Björk, 1999). Leu73 presumably keeps cystatin A attached to the proteases by directly interacting with the latter. Pro74 could contribute to stabilizing the complex with cathepsin B either by direct hydrophobic contacts with the enzyme or, alternatively, by maintaining Leu73 in an orientation beneficial for its interaction with the protease.

To provide further insight into the structural implications of these results, a computer model of the complex between human cystatin A and papain was built on the basis of the X-ray crystal structure of human C3S-cystatin B in complex with S-(carboxymethyl)papain (Stubbs *et al.*, 1990). The cystatin A polypeptide chain could be accommodated within the main-chain coordinates of cystatin B without any sterical problems in the region of the second binding loop. The computer modeling largely corroborated the results of the experimental work. In the model of the complex, Leu73 of cystatin A thus exhibited six close hydrophobic contacts (within 4Å) with Trp177, highly conserved in papain-like cysteine proteases. Such interactions were experimentally supported by fluorescence difference spectra implying that the environment of Trp177 is less hydrophobic in the complex with L73G-cystatin A, i.e. in the absence of the Leu73 side chain, than in the complex with the wild-type inhibitor. The model indicated only a small contribution of Pro74 to stabilization of the cystatin A—papain complex. In complete agreement with the experimental data, no close intermolecular interactions of Gln76 and Asn77 with papain were seen in the modeled complex.

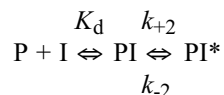
### **Evidence for dislocation of the occluding loop of cathepsin B in the second step of the binding by cystatins (Paper III)**

Cathepsin B differs from other members of the papain family, such as papain and cathepsin L, by the presence of an occluding loop (residues 108–119) that partially blocks one end of the substrate binding cleft. In the X-ray structure of cathepsin B, this loop is stabilized by a number of interactions, including two salt bridges, between His110 and Asp22 and between Arg116 and Asp224 (Musil *et al.*, 1991). Earlier work had demonstrated that the occluding loop normally restricts the access of cystatins to the active site but is a mobile element and may move, e.g. when an endopeptidase substrate binds to cathepsin B (Musil *et al.*, 1991; Cygler *et al.*, 1996; Turk *et al.*, 1996b; Illy *et al.*, 1997; Nägler *et al.*, 1997). Previous to the studies in Paper III, a complex, two-step mechanism of cathepsin B binding by cystatin C had been revealed by stopped-flow kinetic studies (Nycander *et al.*, 1998). The second step of the interaction, involving a conformational change, had

been interpreted to be due to displacement of the occluding loop by cystatin C (Nycander *et al.*, 1998).

The aim of this work was to investigate whether this two-step mode of binding to cathepsin B is general for different cystatins and to provide unequivocal evidence for dislocation of the occluding loop on formation of the tight complex in the second step. To this end, the salt bridge between His110 and Asp22 of the enzyme was disrupted by mutating His110 to Ala. This mutation had been shown previously to increase the mobility of the loop (Nägler *et al.*, 1997) and, consequently, would be expected to relieve the interference of the loop with cystatin binding. Since a H110A mutant of active cathepsin B was found to be unstable due to autoproteolysis, an additional C29A mutation that rendered cathepsin B enzymatically inactive was also introduced, the resulting H110A/C29A-cathepsin B double mutant being used in this study. This construction necessitated control experiments with C29A-cathepsin B, lacking the catalytic cysteine. Stopped-flow kinetics analyses of the binding of cystatins A and C, of families 1 and 2, to C29A- and H110A/C29A-cathepsin B at high inhibitor concentrations were monitored by changes of intrinsic fluorescence accompanying complex formation.

For the binding of either cystatin C or A to the C29A-cathepsin B control, the progress curves were exponential and the dependence of  $k_{\text{obs}}$  on inhibitor concentration was hyperbolic. This behavior is indicative of a two-step reaction with a conformational change in the second step for both cystatins. The simplest kinetic mechanism compatible with the data (Fersht, 1985) involves the formation of a weak, initial complex (PI in the scheme below) in the first step of the binding reaction. This complex is assumed to be in a fast equilibrium with free protease and inhibitor (P and I, respectively). The first step is followed by a reversible conformational change leading to formation of a final, tight complex (PI\*) in the second step:



In this mechanism,  $K_1$  is the dissociation equilibrium constant of the initial complex and  $k_{+2}$  and  $k_{-2}$  are the forward and reverse rate constants of the second, rate-limiting step.  $K_1$  and  $k_{+2}$  were obtained by fitting the hyperbolic dependence of  $k_{\text{obs}}$  vs inhibitor concentration to the equation for the proposed mechanism (Fersht, 1985). The overall association rate constant,  $k_{\text{ass}}$ , was determined from the initial slope of the hyperbolic dependence as  $k_{+2}/K_1$ , whereas the overall dissociation rate constant,  $k_{\text{diss}}$ , which is identical to  $k_{-2}$  in this mechanism, was measured by displacement experiments.

The C29A mutation in cathepsin B was found not to alter the major characteristics of the two-step binding of wild-type cathepsin B by cystatin C. A corresponding comparison was not possible in the case of cystatin A, as the stopped-flow kinetics of cystatin A binding to wild-type cathepsin B could not be quantified due to a low fluorescence change accompanying the binding and a pronounced photodecomposition. Nevertheless, comparison with the results of inhibition

kinetics, monitored by the loss of enzyme activity, for wild-type cathepsin B showed that the C29A mutation only moderately affected the overall rate constants of cystatin A binding. Taken together, the results indicate that C29A-cathepsin B is a satisfactory model for characterizing the mechanism of cystatin binding to wild-type cathepsin B. The studies strongly support the contention that cystatin A binds to wild-type cathepsin B in the same manner as to C29A-cathepsin B and similar to the manner in which cystatin C binds to the two enzyme forms. Therefore, the two-step mechanism of binding to cathepsin B discussed above was suggested to be common for all cystatins.

For both cystatin C and A binding to H110A/C29A-cathepsin B, the dependence of  $k_{\text{obs}}$  on inhibitor concentration was linear up to ~80–100  $\mu\text{M}$ , these concentration limits being determined by the amounts of inhibitor available. This behavior is characteristic of a one-step reaction. However, another alternative that is consistent with the results is that both  $K_1$  and  $k_{+2}$  for the binding of both inhibitors were increased substantially by the H110A mutation, and the data obtained corresponded to the initial, approximately linear segments of hyperbolic concentration dependencies. Simulations indicated that in this case  $K_1$  must have increased  $\gg 2$ –3-fold for both inhibitors and  $k_{+2}$  must have increased  $\gg 2$ -fold for cystatin C and  $\gg 10$ -fold for cystatin A. Both alternative explanations clearly demonstrate that disruption of the salt bridge between His110 and Asp22 in cathepsin B, leading to an increased mobility of the occluding loop, causes a striking change in the kinetics of cystatin binding. They also strongly support the proposal that the second step of cystatin binding to wild-type cathepsin B involves displacement of the occluding loop. It is possible that the loop no longer occludes the S' subsites in H110A/C29A-cathepsin B, which is consistent with cystatins binding to this cathepsin B variant in the same one-step manner as to target proteases with an exposed active site. Alternatively, the loop could have become sufficiently mobile in H110A/C29A-cathepsin B to be easily displaced by cystatins in a two-step reaction. In this case, dislocation of the occluding loop in the second binding step was presumably promoted by an initial weak interaction of either cystatin with the cathepsin B mutant, most likely mediated by the N-terminal regions of the inhibitors, as proposed for the binding to wild-type cathepsin B (Björk *et al.*, 1996; Paper I).

The analyses of the rapid kinetics of binding of cystatins A and C to C29A-cathepsin B provided further insight into the differences between the inhibitory properties of the two inhibitors against cathepsin B. They revealed that the forward rate constant of the dislocation of the occluding loop of the enzyme,  $k_{+2}$ , is markedly lower for cystatin A than for cystatin C. Cystatin A is thus much less effective than cystatin C in displacing the occluding loop in the second step. This low rate of loop dislocation is the main reason for the considerably lower overall rate constant of cathepsin B inhibition by cystatin A, compared with that for inhibition by cystatin C.

## **Improvement of the inhibitory characteristics of cystatin A by grafting of the protease binding regions of more potent inhibitors (Paper IV)**

This investigation was inspired by the results of the previous work in Paper III. In particular, this work had shown that cystatin A is considerably less effective than cystatin C in displacing the occluding loop of cathepsin B in the second step of complex formation, which accounts for the slow overall rate of inhibition of cathepsin B by cystatin A. The latter inhibitor also binds other papain-like cysteine proteases less tightly than cystatin C. The present study was therefore undertaken to identify the structural features of cystatin C that make it a better displacer of the occluding loop of cathepsin B and also a stronger inhibitor of other cysteine proteases. To this end, the influence of introducing individual binding regions of cystatin C onto the framework of cystatin A on the affinities and kinetics of the binding to papain and cathepsin B was investigated in this work.

Cystatin A variants in which either the N-terminal segment preceding the conserved Gly or the first or second binding loops were made identical with or similar to the corresponding regions of cystatin C (Fig. 3) were obtained by site-directed mutagenesis. In two N-terminal region mutants, N(1–10)CC- and N(8–10)-CC-cystatin A, the authentic N-terminal segment, MIP, was replaced by cystatin C residues 1–10 or 8–10. In V47I-cystatin A, the first binding loop was made identical with that of cystatin C by a single point mutation, Val47 to Ile. In G75W-cystatin A, the PW motif, highly conserved in the second binding loop of family 2 cystatins, was introduced into the corresponding loop of cystatin A. In addition, in G75H-cystatin A the second binding loop, which is the binding region of cystatin A differing mostly from those in cystatin B, was engineered to be similar to that of the latter inhibitor by a Gly75 to His substitution.

The affinities of N(1–10)CC- and N(8–10)CC-cystatin A for cathepsin B were higher, by ~15- and >3-fold, respectively, than that of the wild-type inhibitor, predominantly due to an increased  $k_{\text{ass}}$ . As a consequence, the affinities of the N-terminal cystatin A mutants for cathepsin B exceeded those of any natural cystatin known. To elucidate the reason for the higher  $k_{\text{ass}}$ , the stopped-flow kinetics of binding of the N(1–10)CC mutant to an inactive form of cathepsin B, in which the catalytic Cys was substituted by Ala, were analyzed at high inhibitor concentrations. In previous work (Paper III), the C29A mutation had been shown not to affect the mechanism and kinetics of cathepsin B binding by cystatins. As in the case of wild-type cystatin A in the earlier work,  $k_{\text{obs}}$  for the N(1–10)CC-cystatin A association with C29A-cathepsin B showed a hyperbolic dependence on the inhibitor concentration. This behavior is characteristic for a two-step binding mechanism with a conformational change occurring in the second, rate-limiting step (see Scheme in Paper III). The N-terminal substitution thus did not change the mechanism of cathepsin B binding. However, it markedly affected the rate of the second step of the interaction, the forward rate constant for this step,  $k_{+2}$ , being increased by ~100-fold, approaching that of cystatin C binding. Therefore, the results indicated that the N-terminal segment of cystatin C in the chimeric inhibitor, and presumably also in intact cystatin C, contributes to increasing the

rate of cathepsin B binding by efficiently promoting the displacement of the occluding loop of cathepsin B in the second binding step, presumably by appropriately orienting the two binding loops. Nevertheless, the cystatin A mutant harboring the entire N-terminal segment of cystatin C did not attain the full rate of cystatin C association with cathepsin B, possibly due to other residues of the N-terminal region, C-terminal of the conserved Gly, being of importance in cystatin C.

Both mutations in the second binding loop favorably affected both  $k_{\text{ass}}$  and  $k_{\text{diss}}$  of papain binding by cystatin A, which resulted in ~10-fold higher affinities of these two mutants than of the wild-type inhibitor for papain. The G75H mutation predominantly decreased the rate of dissociation of the complexes with papain, whereas the G75W mutation mainly accelerated complex formation. Presumably, both mutations caused local conformational changes in the second binding loop of cystatin A, which is highly flexible in the solution structure of the free wild-type inhibitor (Martin *et al.*, 1995). This restriction in flexibility could promote a faster association with papain. Moreover, computer modeling indicated that the His or Trp introduced into the loop could contribute to an increased stabilization of the complexes with papain by direct interactions with residues of the protease active-site cleft.

In contrast to these results, grafting of the N-terminal residues of cystatin C onto the cystatin A scaffold did not improve papain inhibition. Similarly, the mutations in the second binding loop of cystatin A that introduced essential residues of cystatins C and B did not enhance the inhibition of cathepsin B. None of these mutations appreciably influenced either the affinities or the kinetics of binding of the corresponding protease, with the exception of the G75W mutation. The Trp side chain apparently disturbed the accommodation of the second binding loop of G75W-cystatin A into the active-site crevice of cathepsin B, leading to a 6-fold increase in  $k_{\text{diss}}$ . Grafting of the first binding loop of cystatin C onto the cystatin A framework did not improve the inhibitory characteristics of cystatin A for either papain or cathepsin B. Instead, the affinities of this chimeric inhibitor for the two enzymes were ~2–5-fold lower than those of wild-type cystatin A, exclusively due to a higher  $k_{\text{diss}}$ . The bulkier Ile replacing Val47 in the first binding loop of the cystatin A mutant therefore must disturb the binding of either this loop or the N-terminal segment of the inhibitor to the enzymes (Stubbs *et al.*, 1990). Presumably, the presence of Ile in the corresponding position in the first binding loop of cystatin C also contributes to the lower stability of the complex of this inhibitor with cathepsin B, compared with that of cystatin A.

The results of this study outline a strategy for engineering of new specific and highly potent inhibitors of different cysteine proteases, endopeptidases as well as exopeptidases. Such small protein inhibitors may be potentially useful for abolishing the harmful activities of cathepsins that are related to many pathophysiological disorders.

## General conclusions

This investigation has primarily elucidated the mechanism of inhibition of papain-like cysteine proteases by human cystatin A. Overall, the studies performed give evidence demonstrating that cystatin A, a representative of family 1 cystatins, interacts with target proteases in a similar manner as family 2 cystatins. The contributions of individual residues of the N-terminal region and the second binding loop of cystatin A to the inhibition of different cysteine proteases have been evaluated. These data, together with analyses of the binding kinetics, give insight into the background for the high affinity of cystatin A—cysteine protease interactions. Moreover, successful attempts at improving the inhibitory capacity of cystatin A by protein engineering have been made. The results indicate possible approaches for the design of effective inhibitors of the harmful activity of cysteine proteases in pathological conditions. The principal findings of the study are as follows:

- The flexible N-terminal segment of cystatin A is indispensable for the high affinity of this inhibitor for papain, cathepsin L and cathepsin B. Three residues at the N-terminus together account for ~40% of the total free energy of cystatin A binding to these proteases. Met1 is of minimal importance, whereas Pro3, preceding the evolutionarily conserved Gly4, is responsible for most of the contribution of the N-terminal segment to the binding. In the case of the inhibition of proteases with an easily accessible active site, such as papain and cathepsin L, the primary role of the N-terminal segment is to stabilize the cystatin A—protease complex after this complex has already been formed. The flexible N-terminal segment of cystatin A most likely binds to such proteases after the remainder of the inhibitory wedge. In the binding of cathepsin B, in which an occluding loop blocks the active site, the N-terminal segment is essential for fast association of cystatin A with the protease, although it also participates in stabilization of the bimolecular complex. The N-terminal segment of cystatin A most likely interacts with cathepsin B in the first step of a two-step binding reaction.
- The second binding loop of cystatin A is highly important for efficient inhibition of different cysteine proteases. The most flexible segment of this loop, comprising the residues from Leu73 to Asn77, contributes ~20 and ~45% of the energy of binding of cystatin A to papain and cathepsin B, respectively. The contribution of the loop to protease binding is due almost solely to Leu73, highly conserved in family 1 cystatins. This residue presumably establishes direct hydrophobic interactions with the target protease. Pro74 is also of appreciable importance for cathepsin B binding, in which this residue may either form direct contacts with the protease or assist by positioning Leu73 for appropriate interaction. The second binding loop of cystatin A participates in the inhibition of cysteine proteases by serving as an anchor, stabilizing the bimolecular complex once it is formed.
- Cystatins A and C bind to an inactive form of cathepsin B, in which the catalytic Cys29 is replaced by Ala, by a two-step mechanism, with a conformational change in the second, rate-limiting step. Strong evidence

indicates that cystatin A, like cystatin C, interacts with wild-type cathepsin B in the same manner. Increasing the mobility of the occluding loop of C29A-cathepsin B by disruption of a salt bridge between His110 in the loop and Asn22 of the main body of the protein by a H110A mutation causes a marked change in the mode of binding of both cystatins A and C, transforming it into an apparent one-step reaction. Thus, the conformational change in the second step of cathepsin B binding by cystatins most certainly involves the occluding loop being dislocated by the cystatin to allow formation of a tight complex. Cystatin A is less efficient than cystatin C in displacing the occluding loop. The lower overall rate of association of cystatin A than of cystatin C with cathepsin B is primarily a consequence of this slow displacement of the occluding loop of the enzyme in the second binding step.

- Essential inhibitory properties of the most potent mammalian inhibitor of papain-like proteases, cystatin C, can be imparted onto the weaker inhibitor, cystatin A. Replacement of the three residues preceding the conserved Gly in the N-terminal region of cystatin A by the corresponding ten residue long N-terminal segment of cystatin C results in affinities for cathepsin B exceeding that of any natural cystatin known. The high affinity is due to the cystatin A/C chimera binding to cathepsin B appreciably faster than cystatin A but retaining the low  $k_{\text{diss}}$  characteristic for the interaction of cystatin A with this protease. The N-terminal segment of cystatin C on the cystatin A framework markedly increases the rate of dislocation of the occluding loop of the inactive C29A-cathepsin B variant in the second step of complex formation, the rate constant of this step being comparable with that for the binding by cystatin C. This region of cystatin C is thus strongly indicated to be pivotal for the better ability of this inhibitor than of cystatin A to displace the occluding loop of cathepsin B and thus for the higher overall association rate. By contrast, grafting the N-terminal residues of cystatin C onto the cystatin A scaffold does not appreciably enhance the inhibition of papain. Mutations of Gly75 to Trp or His, which increase the sequence similarity of the second binding loop of cystatin A with those in cystatins C and B, respectively, significantly enhance the affinity for papain up to a level similar to that of cystatin C by favorably affecting both  $k_{\text{ass}}$  and  $k_{\text{diss}}$ . These mutations, however, are somewhat deleterious for cathepsin B binding by cystatin A. Introduction of the first binding loop of cystatin C onto the cystatin A framework does not improve the inhibitory characteristics of cystatin A against either papain or cathepsin B. Taken together, the results show that the inhibitory efficiency of cystatins can be substantially improved by protein engineering.

## References

- Abe, K., Emori, Y., Kondo, H., Arai, S. & Suzuki, K. 1988. The NH<sub>2</sub>-terminal 21 amino-acid residues are not essential for the papain-inhibitory activity of oryzacystatin, a member of the cystatin superfamily. Expression of oryzacystatin cDNA and its truncated fragments in *Escherichia coli*. *J Biol Chem* 263, 7655-7659.
- Abe, M., Abe, K., Kuroda, M. & Arai, S. 1992. Corn kernel cysteine proteinase inhibitor as a novel cystatin superfamily member of plant origin. Molecular cloning and expression studies. *Eur J Biochem* 209, 933-937.
- Abrahamson, M. 1994 Cystatins. *Methods Enzymol* 244, 685-700.
- Abrahamson, M., Barrett, A.J., Salvesen, G. & Grubb, A. 1986. Isolation of six cysteine proteinase inhibitors from human urine. Their physicochemical and enzyme kinetic properties and concentrations in biological fluids. *J Biol Chem*, 261, 11282-11289.
- Abrahamson, M., Mason, R.W., Hansson, H., Buttle, D.J., Grubb, A. & Ohlsson, K. 1991. Human cystatin C. role of the N-terminal segment in the inhibition of human cysteine proteinases and in its inactivation by leucocyte elastase. *Biochem J* 273, 621-626.
- Abrahamson, M., Jonsdottir, S., Olafsson, I., Jensson, O. & Grubb, A. 1992. Hereditary cystatin C amyloid angiopathy: identification of the disease-causing mutation and specific diagnosis by polymerase chain reaction based analysis. *Hum Genet* 89, 377-380.
- Abrahamson, M., Wikstrom, M., Potempa, J., Renvert, S. & Hall, A. 1997. Modification of cystatin C activity by bacterial proteinases and neutrophil elastase in periodontitis. *Mol Pathol* 50, 291-297.
- Agarwala, K.L., Kawabata, S., Hirata, M., Miyagi, M., Tsunasawa, S. & Iwanaga, S. 1996. A cysteine protease inhibitor stored in the large granules of horseshoe crab hemocytes: purification, characterization, cDNA cloning and tissue localization. *J Biochem (Tokyo)* 119, 85-94.
- Alakurtti, K., Virtaneva, K., Joensuu, T., Palvimo, J.J. & Lehesjoki, A.E. 2000. Characterization of the cystatin B gene promoter harboring the dodecamer repeat expanded in progressive myoclonus epilepsy, EPM1. *Gene* 242, 65-73.
- Alvarez-Fernandez, M., Barrett, A.J., Gerhartz, B., Dando, P.M., Ni, J. and Abrahamson, M. 1999. Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. *J Biol Chem* 274, 19195-19203.
- Aoki, H., Akaike, T., Abe, K., Kuroda, M., Arai, S., Okamura, R., Negi, A. & Maeda, H. 1995. Antiviral effect of oryzacystatin, a proteinase inhibitor in rice, against herpes simplex virus type 1 in vitro and in vivo. *Antimicrob Agents Chemother* 39, 846-849.
- Armstrong, P.B. 2001 The contribution of proteinase inhibitors to immune defense. *Trends Immunol* 22, 47-52.
- Auerswald, E.A., Genenger, G., Assfalg-Machleidt, I., Machleidt, W., Engh, R.A. & Fritz, H. 1992. Recombinant chicken egg white cystatin variants of the QLVSG region. *Eur J Biochem* 209, 837-845.
- Auerswald, E.A., Nägler, D.K., Assfalg-Machleidt, I., Stubbs, M.T., Machleidt, W. and Fritz, H. 1995. Hairpin loop mutations of chicken cystatin have different effects on the inhibition of cathepsin B, cathepsin L & papain. *FEBS Lett* 361, 179-184.
- Barrett, A.J. 1986 The cystatins: a diverse superfamily of cysteine peptidase inhibitors. *Biomed Biochim Acta* 45, 1363-1374.
- Barrett, A.J. & Rawlings, N.D. 2001. Evolutionary lines of cysteine peptidases. *Biol Chem* 382, 727-733.
- Barrett, A.J., Fritz, H., Grubb, A., Isemura, S., Järvinen, M., Katunuma, N., Machleidt, W., Müller-Esterl, W., Sasaki, M. & Turk, V. 1986a. Nomenclature and classification of the proteins homologous with the cysteine-proteinase inhibitor chicken cystatin. *Biochem J* 236, 312.
- Barrett, A.J., Rawlings, N.D., Davies, M.E., Machleidt, W., Salvesen, G. & Turk, V. 1986b. Cysteine proteinase inhibitors of the cystatin family. In: Barrett, A.J., Salvesen, G.



- eds. *Proteinase Inhibitors*. Amsterdam: Elsevier. 515-569. ISBN 0-444-80763-2 (U.S.)
- Barrett, A.J., Rawlings, N.D. & O'Brien, E.A. 2001. The MEROPS database as a protease information system. *J Struct Biol* 134, 95-102.
- Berti, P.J. & Storer, A.C. 1995. Alignment/phylogeny of the papain superfamily of cysteine proteases. *J Mol Biol* 246, 273-283.
- Bevec, T., Stoka, V., Pungercic, G., Dolenc, I. & Turk, V. 1996. Major histocompatibility complex class II-associated p41 invariant chain fragment is a strong inhibitor of lysosomal cathepsin L. *J Exp Med* 183, 1331-1338.
- Bhandoola, A., Kithiganahalli, B., Granger, L., Singer, A. 2000. Programming for cytotoxic effector function occurs concomitantly with CD4 extinction during CD8(+) T cell differentiation in the thymus. *Int Immunol* 12, 1035-40.
- Bieth, J.G. 1984. In vivo significance of kinetic constants of protein proteinase inhibitors. *Biochem. Med* 32, 387-397.
- Björck, L., Åkesson, P., Bohus, M., Trojnar, J., Abrahamson, M., Olafsson, I. & Grubb, A. 1989. Bacterial growth blocked by a synthetic peptide based on the structure of a human proteinase inhibitor. *Nature* 337, 385-386.
- Björck, L., Grubb, A. & Kjellen, L. 1990. Cystatin C, a human proteinase inhibitor, blocks replication of herpes simplex virus. *J Virol* 64, 941-943.
- Björk, I. & Ylinenjärvi, K. 1990. Interaction between chicken cystatin and the cysteine proteinases actinidin, chymopapain A, and ficin. *Biochemistry* 29, 1770-1776.
- Björk, I., Aliksson, E. & Ylinenjärvi, K. (1989) Kinetics of binding of chicken cystatin to papain. *Biochemistry*, 28, 1568-1573.
- Björk, I., Ylinenjärvi, K. & Lindahl, P. 1990. Equilibrium and kinetic studies of the interaction of chicken cystatin with four cysteine proteinases. *Biol Chem Hoppe Seyler* 371 Suppl, 119-124.
- Björk, I., Pol, E., Raub-Segall, E., Abrahamson, M., Rowan, A.D. & Mort, J.S. 1994. Differential changes in the association and dissociation rate constants for binding of cystatins to target proteinases occurring on N-terminal truncation of the inhibitors indicate that the interaction mechanism varies with different enzymes. *Biochem J* 299, 219-225.
- Björk, I., Brieditis, I. & Abrahamson, M. 1995. Probing the functional role of the N-terminal region of cystatins by equilibrium and kinetic studies of the binding of Gly-11 variants of recombinant human cystatin C to target proteinases. *Biochem J* 306, 513-518.
- Björk, I., Brieditis, I., Raub-Segall, E., Pol, E., Håkansson, K. & Abrahamson, M. 1996. The importance of the second hairpin loop of cystatin C for proteinase binding. Characterization of the interaction of Trp-106 variants of the inhibitor with cysteine proteinases. *Biochemistry* 35, 10720-10726.
- Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J. & Turk, V. 1988. The 2.0 Å X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. *Embo J* 7, 2593-2599.
- Bode, W., Engh, R., Musil, D., Laber, B., Stubbs, M., Huber, R. & Turk, V. 1990. Mechanism of interaction of cysteine proteinases and their protein inhibitors as compared to the serine proteinase-inhibitor interaction. *Biol Chem Hoppe Seyler* 371 Suppl 111-118.
- Brown, W.M. & Dziegielewska, K.M. 1997. Friends and relations of the cystatin superfamily--new members and their evolution. *Protein Sci* 6, 5-12.
- Brzin, J., Kopitar, M., Turk, V. & Machleidt, W. 1983. Protein inhibitors of cysteine proteinases. I. Isolation and characterization of stefin, a cytosolic protein inhibitor of cysteine proteinases from human polymorphonuclear granulocytes. *Hoppe Seylers Z Physiol Chem* 364, 1475-1480.
- Bühling, F., Fengler, A., Brandt, W., Welte, T., Ansorge, S. & Nägler, D.K. 2000. Novel cysteine proteases of the papain family. *Adv Exp Med Biol* 477, 241-254.
- Calkins, C.C. & Sloane, B.F. 1995). Mammalian cysteine protease inhibitors: biochemical properties and possible roles in tumor progression. *Biol Chem Hoppe Seyler* 376, 71-80.

- Chagas, J.R., Authie, E., Serveau, C., Lalmanach, G., Juliano, L. & Gauthier, F. 1997. A comparison of the enzymatic properties of the major cysteine proteinases from *Trypanosoma congolense* and *Trypanosoma cruzi*. *Mol Biochem Parasitol* 88, 85-94.
- Chapman, H.A., Riese, R.J. & Shi, G.P. 1997a. Emerging roles for cysteine proteases in human biology. *Annu Rev Physiol* 59, 63-88.
- Chapman, R.L., Kane, S.E. & Erickson, A.H. 1997b. Abnormal glycosylation of procathepsin L due to N-terminal point mutations correlates with failure to sort to lysosomes. *J Biol Chem* 272, 8808-8816.
- Chen, J.M., Dando, P.M., Rawlings, N.D., Brown, M.A., Young, N.E., Stevens, R.A., Hewitt, E., Watts, C. & Barrett, A.J. 1997. Cloning, isolation, and characterization of mammalian legumain, an asparaginyl endopeptidase. *J Biol Chem* 272, 8090-8098.
- Chen, J.M., Rawlings, N.D., Stevens, R.A. & Barrett, A.J. 1998. Identification of the active site of legumain links it to caspases, clostripain and gingipains in a new clan of cysteine endopeptidases. *FEBS Lett* 441, 361-365.
- Cimerman, N., Kosorok, M.D., Korant, B.D., Turk, B. & Turk, V. 1996. Characterization of cystatin C from bovine parotid glands: cysteine proteinase inhibition and antiviral properties. *Biol Chem Hoppe Seyler* 377, 19-23.
- Cimerman, N., Prebanda, M.T., Turk, B., Popovic, T., Dolenc, I. & Turk, V. 1999. Interaction of cystatin C variants with papain and human cathepsins B, H and L. *J Enzyme Inhib* 14, 167-174.
- Cimerman, N., Mesko Brguljan, P., Krasovec, M., Suskovic, S. & Kos, J. 2001. Serum concentration and circadian profiles of cathepsins B, H and L, and their inhibitors, stefins A and B, in asthma. *Clin Chim Acta* 310, 113-122.
- Colella, R., Goodwyn, E. & Gopal, P. 2002. Increased cell density decreases cysteine proteinase inhibitor activity and increases invasive ability of two prostate tumor cell lines. *Cancer Lett* 185, 163-172.
- Collins, A.R. & Grubb, A. 1991. Inhibitory effects of recombinant human cystatin C on human coronaviruses. *Antimicrob Agents Chemother* 35, 2444-2446.
- Collins, A.R. & Grubb, A. 1998. Cystatin D, a natural salivary cysteine protease inhibitor, inhibits coronavirus replication at its physiologic concentration. *Oral Microbiol Immunol* 13, 59-61.
- Cornwall, G.A. & Hsia, N. 2003. A new subgroup of the family 2 cystatins. *Mol Cell Endocrinol* 200, 1-8.
- Cornwall, G.A., Orgebin-Crist, M.C. & Hann, S.R. 1992. The CRES gene: a unique testis-regulated gene related to the cystatin family is highly restricted in its expression to the proximal region of the mouse epididymis. *Mol Endocrinol* 6, 1653-1664.
- Cornwall, G.A., Hsia, N. & Sutton, H.G. 1999. Structure, alternative splicing and chromosomal localization of the cystatin-related epididymal spermatogenic gene. *Biochem J* 340, 85-93.
- Cornwall, G.A., Cameron, A., Lindberg, I., Hardy, D.M., Cormier, N. & Hsia, N. 2003. The cystatin-related epididymal spermatogenic protein inhibits the serine protease prohormone convertase 2. *Endocrinology* 144, 901-908.
- Cox, G.A., Johnson, R.B., Cook, J.A., Wakulchik, M., Johnson, M.G., Villarreal, E.C. & Wang, Q.M. 1999a. Identification and characterization of human rhinovirus-14 3C protease deamidation isoform. *J Biol Chem* 274, 13211-13216.
- Cox, J.L., Sexton, P.S., Green, T.J. & Darmani, N.A. 1999b. Inhibition of B16 melanoma metastasis by overexpression of the cysteine proteinase inhibitor cystatin C. *Melanoma Res* 9, 369-374.
- Craven, C.J., Baxter, N.J., Murray, E.H., Hill, N.J., Martin, J.R., Ylinenjärvi, K., Björk, I., Waltho, J.P. & Murray, I.A. 2000. Wild-type and Met-65-->Leu variants of human cystatin A are functionally and structurally identical. *Biochemistry* 39, 15783-15790.
- Cygler, M. & Mort, J.S. 1997. Proregion structure of members of the papain superfamily. Mode of inhibition of enzymatic activity. *Biochimie* 79, 645-652.

- Cyglar, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A.C. & Mort, J.S. 1996. Structure of rat procathepsin B: model for inhibition of cysteine protease activity by the proregion. *Structure* 4, 405-416.
- Dainichi, T., Maekawa, Y., Ishii, K., Zhang, T., Nashed, B.F., Sakai, T., Takashima, M. & Himeno, K. 2001. Nippocystatin, a cysteine protease inhibitor from *Nippostrongylus brasiliensis*, inhibits antigen processing and modulates antigen-specific immune response. *Infect Immun* 69, 7380-7386.
- Davies, M.E. & Barrett, A.J. 1984. Immunolocalization of human cystatins in neutrophils and lymphocytes. *Histochemistry* 80, 373-377.
- DeLa Cadena, R.A. & Colman, R.W. 1991. Structure and functions of human kininogens. *Trends Pharmacol Sci* 12, 272-275.
- Delaria, K., Fiorentino, L., Wallace, L., Tamburini, P., Brownell, E. & Muller, D. 1994. Inhibition of cathepsin L-like cysteine proteases by cytotoxic T-lymphocyte antigen-2 beta. *J Biol Chem* 269, 25172-25177.
- Delbridge, M.L. & Kelly, L.E. 1990. Sequence analysis, and chromosomal localization of a gene encoding a cystatin-like protein from *Drosophila melanogaster*. *FEBS Lett* 274, 141-145.
- Deng, A., Irizarry, M.C., Nitsch, R.M., Growdon, J.H. & Rebeck, G.W. 2001. Elevation of cystatin C in susceptible neurons in Alzheimer's disease. *Am J Pathol* 159, 1061-1068.
- Deussing, J., Roth, W., Saftig, P., Peters, C., Ploegh, H.L. & Villadangos, J.A. 1998. Cathepsins B and D are dispensable for major histocompatibility complex class II-mediated antigen presentation. *Proc Natl Acad Sci U S A* 95, 4516-4521.
- Dieckmann, T., Mitschang, L., Hofmann, M., Kos, J., Turk, V., Auerswald, E.A., Jaenicke, R. & Oschkinat, H. 1993. The structures of native phosphorylated chicken cystatin and of a recombinant unphosphorylated variant in solution. *J Mol Biol* 234, 1048-1059.
- Di Giaimo, R., Riccio, M., Santi, S., Galeotti, C., Ambrosetti, D.C., Melli, M. 2002. New insights into the molecular basis of progressive myoclonus epilepsy: a multiprotein complex with cystatin B. *Hum Mol Genet* 11, 2941-2950.
- Ebert, E., Werle, B., Julke, B., Kopitar-Jerala, N., Kos, J., Lah, T., Abrahamson, M., Spiess, E. & Ebert, W. 1997. Expression of cysteine protease inhibitors stefin A, stefin B, and cystatin C in human lung tumor tissue. *Adv Exp Med Biol* 421, 259-265.
- Ekiel, I., Abrahamson, M., Fulton, D.B., Lindahl, P., Storer, A.C., Levadoux, W., Lafrance, M., Labelle, S., Pomerleau, Y., Groleau, D., LeSauter, L. & Gehring, K. 1997. NMR structural studies of human cystatin C dimers and monomers. *J Mol Biol* 271, 266-277.
- Engh, R.A., Dieckmann, T., Bode, W., Auerswald, E.A., Turk, V., Huber, R. & Oschkinat, H. 1993. Conformational variability of chicken cystatin. Comparison of structures determined by X-ray diffraction and NMR spectroscopy. *J Mol Biol* 234, 1060-1069.
- Esnard, A., Esnard, F., Faucher, D. and Gauthier, F. 1988. Two rat homologues of human cystatin C. *FEBS Lett* 236, 475-478.
- Esnard, F., Esnard, A., Faucher, D., Capony, J.P., Derancourt, J., Brillard, M. & Gauthier, F. 1990. Rat cystatin C: the complete amino acid sequence reveals a site for N-glycosylation. *Biol Chem Hoppe Seyler* 371 Suppl, 161-166.
- Estrada, S., Nycander, M., Hill, N.J., Craven, C.J., Waltho, J.P. & Björk, I. 1998. The role of Gly-4 of human cystatin A (stefin A) in the binding of target proteinases. Characterization by kinetic and equilibrium methods of the interactions of cystatin A Gly-4 mutants with papain, cathepsin B, and cathepsin L. *Biochemistry* 37, 7551-7560.
- Estrada, S., Olson, S.T., Raub-Segall, E. & Björk, I. 2000. The N-terminal region of cystatin A (stefin A) binds to papain subsequent to the two hairpin loops of the inhibitor. Demonstration of two-step binding by rapid-kinetic studies of cystatin A labeled at the N-terminus with a fluorescent reporter group. *Protein Sci* 9, 2218-2224.

- Fernandez, P.L., Farre, X., Nadal, A., Fernandez, E., Peiro, N., Sloane, B.F., Shi, G.P., Chapman, H.A., Campo, E. & Cardesa, A. 2001. Expression of cathepsins B and S in the progression of prostate carcinoma. *Int J Cancer* 95, 51-55.
- Fersht, A. 1985. *Enzyme Structure and Mechanism* 2nd ed. New York: W.H. Freeman. 128-137.
- Foghsgaard, L., Wissing, D., Mauch, D., Lademann, U., Bastholm, L., Boes, M., Elling, F., Leist, M., Jaattela, M. 2001. Cathepsin B acts as a dominant execution protease in tumor cell apoptosis induced by tumor necrosis factor. *J Cell Biol* 153, 999-1010.
- Freije, J.P., Abrahamson, M., Olafsson, I., Velasco, G., Grubb, A. & Lopez-Otin, C. 1991. Structure and expression of the gene encoding cystatin D, a novel human cysteine proteinase inhibitor. *J Biol Chem* 266, 20538-20543.
- Freije, J.P., Balbin, M., Abrahamson, M., Velasco, G., Dalbøge, H., Grubb, A. & Lopez-Otin, C. 1993. Human cystatin D. cDNA cloning, characterization of the *Escherichia coli* expressed inhibitor, and identification of the native protein in saliva. *J Biol Chem* 268, 15737-15744.
- Friedrich, B., Jung, K., Lein, M., Turk, I., Rudolph, B., Hampel, G., Schnorr, D. & Loening, S.A. 1999. Cathepsins B, H, L and cysteine protease inhibitors in malignant prostate cell lines, primary cultured prostatic cells and prostatic tissue. *Eur J Cancer* 35, 138-144.
- Gerhartz, B. & Abrahamson, M. 2002. Physico-chemical properties of the N-terminally truncated L68Q cystatin C found in amyloid deposits of brain haemorrhage patients. *Biol Chem* 383, 301-305.
- Gerhartz, B., Ekiel, I. & Abrahamson, M. 1998. Two stable unfolding intermediates of the disease-causing L68Q variant of human cystatin C. *Biochemistry* 37, 17309-17317.
- Green, G.D., Kumbhavi, A.A., Davies, M.E. & Barrett, A.J. 1984. Cystatin-like cysteine proteinase inhibitors from human liver. *Biochem J.* 218, 939-946.
- 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* 32, 504-514.
- Grubb, A. & Löfberg, H. 1982. Human gamma-trace, a basic microprotein: amino acid sequence and presence in the adenohypophysis. *Proc Natl Acad Sci US A* 79, 3024-3027.
- Grubb, A., Abrahamson, M., Olafsson, I., Trojnar, J., Kasprzykowska, R., Kasprzykowski, F. & Grzonka, Z. 1990. Synthesis of cysteine proteinase inhibitors structurally based on the proteinase interacting N-terminal region of human cystatin C. *Biol Chem Hoppe Seyler* 371, Suppl 137-144.
- 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, V. & Turk, D. 1999. Crystal structure of MHC class II-associated p41 Ii fragment bound to cathepsin L reveals the structural basis for differentiation between cathepsins L and S. *Embo J* 18, 793-803.
- Halfon, S., Ford, J., Foster, J., Dowling, L., Lucian, L., Sterling, M., Xu, Y., Weiss, M., Ikeda, M., Liggett, D., Helms, A., Caux, C., Lebecque, S., Hannum, C., Menon, S., McClanahan, T., Gorman, D. & Zurawski, G. 1998. Leukocystatin, a new Class II cystatin expressed selectively by hematopoietic cells. *J Biol Chem* 273, 16400-16408.
- Hall, A., Dalbøge, H., Grubb, A. & Abrahamson, M. 1993. Importance of the evolutionarily conserved glycine residue in the N-terminal region of human cystatin C (Gly-11) for cysteine endopeptidase inhibition. *Biochem J* 291, 123-129.
- Hall, A., Håkansson, K., Mason, R.W., Grubb, A. & Abrahamson, M. 1995. Structural basis for the biological specificity of cystatin C. Identification of leucine 9 in the N-terminal binding region as a selectivity-conferring residue in the inhibition of mammalian cysteine peptidases. *J Biol Chem* 270, 5115-5121.

- Hall, A., Ekiel, I., Mason, R.W., Kasprzykowski, F., Grubb, A. & Abrahamson, M. 1998. Structural basis for different inhibitory specificities of human cystatins C and D. *Biochemistry* 37, 4071-4079.
- Hamil, K.G., Liu, Q., Sivashanmugam, P., Yenugu, S., Soundararajan, R., Grossman, G., Richardson, R.T., Zhang, Y.L., O'Rand, M.G., Petrusz, P., French, F.S. & Hall, S.H. 2002. Cystatin 11: a new member of the cystatin type 2 family. *Endocrinology* 143, 2787-2796.
- Hansen, T., Petrow, P.K., Gaumann, A., Keyszer, G., Brauer, R. & Kriegsmann, J. 2000. Synovial giant cells in rheumatoid arthritis: expression of cystatin C, but not of cathepsin B. *Exp Toxicol Pathol* 52, 312-316.
- Hartmann, S., Schonemeyer, A., Sonnenburg, B., Vray, B. & Lucius, R. 2002. Cystatins of filarial nematodes up-regulate the nitric oxide production of interferon-gamma-activated murine macrophages. *Parasite Immunol* 24, 253-262.
- Henskens, Y.M., Veerman, E.C.I. & Nieuw Amerongen, A.V. 1996. Cystatins in health and disease. *Biol Chem Hoppe Seyler* 377, 71-86.
- Hiltke, T.R., Lee, T.C. & Bobek, L.A. 1999. Structure/function analysis of human cystatin SN and comparison of the cysteine proteinase inhibitory profiles of human cystatins C and SN. *J Dent Res* 78, 1401-1409.
- Hirado, M., Tsunasawa, S., Sakiyama, F., Niinobe, M. & Fujii, S. 1985. Complete amino acid sequence of bovine colostrum low-Mr cysteine proteinase inhibitor. *FEBS Lett* 186, 41-45.
- Hiwasa, T., Ma, J., Ike, Y., Katunuma, N. & Sakiyama, S. 1995. Increase of cyclin B by overexpression of cystatin alpha. *Cell Biochem Funct* 13, 293-296.
- Hopsu-Havu, V.K., Järvinen, M. & Rinne, A. 1983a. Separation of cysteine proteinase inhibitors from psoriatic scale. *Br J Dermatol* 109 Suppl 25, 77-85.
- Hopsu-Havu, V.K., Joronen, I.A., Järvinen, M. & Rinne, A. 1983b. Cysteine proteinase inhibitors in psoriatic epidermis. *Arch Dermatol Res* 275, 305-309.
- Hsia, N. & Cornwall, G.A. 2003. *Cres2* and *Cres3*: new members of the cystatin-related epididymal spermatogenic subgroup of family 2 cystatins. *Endocrinology* 144, 909-915.
- Illy, C., Quraishi, O., Wang, J., Purisima, E., Vernet, T. & Mort, J.S. 1997. Role of the occluding loop in cathepsin B activity. *J Biol Chem* 272, 1197-1202.
- Irvine, J.W., Coombs, G.H. & North, M.J. 1992. Cystatin-like cysteine proteinase inhibitors of parasitic protozoa. *FEMS Microbiol Lett* 75, 67-72.
- Isemura, S., Saitoh, E., Ito, S., Isemura, M. & Sanada, K. 1984. Cystatin S: a cysteine proteinase inhibitor of human saliva. *J Biochem (Tokyo)* 96, 1311-1314.
- Isemura, S., Saitoh, E. & Sanada, K. 1986. Characterization of a new cysteine proteinase inhibitor of human saliva, cystatin SN, which is immunologically related to cystatin S. *FEBS Lett* 198, 145-149.
- Isemura, S., Saitoh, E. & Sanada, K. 1987. Characterization and amino acid sequence of a new acidic cysteine proteinase inhibitor (cystatin SA) structurally closely related to cystatin S, from human whole saliva. *J Biochem (Tokyo)* 102, 693-704.
- Järvinen, M., Rinne, A. & Hopsu-Havu, V.K. 1987. Human cystatins in normal and diseased tissues--a review. *Acta Histochem* 82, 5-18.
- Jenko, S., Dolenc, I., Guncar, G., Dobersek, A., Podobnik, M. & Turk, D. 2003. Crystal structure of Stefin A in complex with cathepsin H: N-terminal residues of inhibitors can adapt to the active sites of endo- and exopeptidases. *J Mol Biol* 326, 875-885.
- Jerala, R., Trstenjak-Prebenda, M., Kroon-Zitko, L., Lenarcic, B. and Turk, V. 1990. Mutations in the QVVAG region of the cysteine proteinase inhibitor stefin B. *Biol Chem Hoppe Seyler* 371 Suppl, 157-160.
- Jerala, R., Kroon-Zitko, L., Kopitar, M., Popovic, T. & Turk, V. 1991. Deletion of the carboxy terminal part of stefin B does not have a major effect for binding to papain. *Biomed Biochim Acta* 50, 627-629.
- Kanno, Y., Tamura, M., Chuma, S., Sakura, T., Machida, T. & Nakatsuji, N. 1999. A cystatin-related gene, testatin/cresp, shows male-specific expression in germ and somatic cells from the initial stage of murine gonadal sex-differentiation. *Int J Dev Biol* 43, 777-784.

- Kasprzykowski, F., Schalen, C., Kasprzykowska, R., Jastrzebska, B. & Grubb, A. 2000. Synthesis and antibacterial properties of peptidyl derivatives and cyclopeptides structurally based upon the inhibitory centre of human cystatin C. Dissociation of antiproteolytic and antibacterial effects. *Apmis* 108, 473-481.
- Katunuma, N. & Kominami, E. 1985. Lysosomal thiol cathepsins and their endogenous inhibitors. Distribution and localization. *Prog Clin Biol Res* 180, 71-79.
- Khan, A.R. & James, M.N. 1998. Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci* 7, 815-836.
- Kimura, M., Ikeda, T., Fukumoto, D., Yamasaki, N. & Yonekura, M. 1995. Primary structure of a cysteine proteinase inhibitor from the fruit of avocado (*Persea americana* Mill). *Biosci Biotechnol Biochem* 59, 2328-2329.
- Kinne, R., Saukko, P., Järvinen, M. & Lehesjoki, A.E. 2002. Reduced cystatin B activity correlates with enhanced cathepsin activity in progressive myoclonus epilepsy. *Ann Med* 34, 380-385.
- Kirschke, H., Barrett, A.J. & Rawlings, N.D. 1995. Proteinases 1: lysosomal cysteine proteinases. *Protein Profile* 2, 1581-1643.
- Kitamura, N., Kitagawa, H., Fukushima, D., Takagaki, Y., Miyata, T. & Nakanishi, S. 1985. Structural organization of the human kininogen gene and a model for its evolution. *J Biol Chem* 260, 8610-8617.
- Klemencic, I., Carmona, A.K., Cezari, M.H., Juliano, M.A., Juliano, L., Guncar, G., Turk, D., Krizaj, I., Turk, V. & Turk, B. 2000. Biochemical characterization of human cathepsin X revealed that the enzyme is an exopeptidase, acting as carboxymonopeptidase or carboxydipeptidase. *Eur J Biochem* 267, 5404-5412.
- Koiwa, H., Shade, R.E., Zhu-Salzman, K., D'Urzo, M.P., Murdock, L.L., Bressan, R.A. & Hasegawa, P.M. 2000. A plant defensive cystatin (soyacystatin) targets cathepsin L-like digestive cysteine proteinases (DvCALs) in the larval midgut of western corn rootworm (*Diabrotica virgifera virgifera*). *FEBS Lett* 471, 67-70.
- Koiwa, H., D'Urzo, M.P., Assfalg-Machleidt, I., Zhu-Salzman, K., Shade, R.E., An, H., Murdock, L.L., Machleidt, W., Bressan, R.A. & Hasegawa, P.M. 2001. Phage display selection of hairpin loop soyacystatin variants that mediate high affinity inhibition of a cysteine proteinase. *Plant J* 27, 383-391.
- Kolkhorst, V., Sturzebecher, J. & Wiederanders, B. 1998. Inhibition of tumour cell invasion by protease inhibitors: correlation with the protease profile. *J Cancer Res Clin Oncol* 124, 598-606.
- Kondo, H., Emori, Y., Abe, K., Suzuki, K. & Arai, S. 1989. Cloning and sequence analysis of the genomic DNA fragment encoding oryzacystatin. *Gene* 81, 259-265.
- Kondo, H., Abe, K., Nishimura, I., Watanabe, H., Emori, Y. & Arai, S. 1990. Two distinct cystatin species in rice seeds with different specificities against cysteine proteinases. Molecular cloning, expression, and biochemical studies on oryzacystatin-II. *J Biol Chem* 265, 15832-15837.
- Kondo, H., Abe, K., Emori, Y. & Arai, S. 1991. Gene organization of oryzacystatin-II, a new cystatin superfamily member of plant origin, is closely related to that of oryzacystatin-I but different from those of animal cystatins. *FEBS Lett* 278, 87-90.
- Kopitar, M., Ritonja, A., Popovic, T., Gabrijelcic, D., Krizaj, I. & Turk, V. 1989. A new type of low-molecular mass cysteine proteinase inhibitor from pig leukocytes. *Biol Chem Hoppe Seyler* 370, 1145-1151.
- Korant, B.D., Brzin, J. & Turk, V. 1985. Cystatin, a protein inhibitor of cysteine proteases alters viral protein cleavages in infected human cells. *Biochem Biophys Res Commun* 127, 1072-1076.
- Kos, J., Krasovec, M., Cimerman, N., Nielsen, H.J., Christensen, I.J. & Brunner, N. 2000a. Cysteine proteinase inhibitors stefin A, stefin B, and cystatin C in sera from patients with colorectal cancer: relation to prognosis. *Clin Cancer Res* 6, 505-511.
- Kos, J., Werle, B., Lah, T. & Brunner, N. 2000b. Cysteine proteinases and their inhibitors in extracellular fluids: markers for diagnosis and prognosis in cancer. *Int J Biol Markers* 15, 84-89.
- Kouzuma, Y., Tsukigata, K., Inanaga, H., Doi-Kawano, K., Yamasaki, N. & Kimura, M. 2001. Molecular cloning and functional expression of cDNA encoding the cysteine

- proteinase inhibitor Sca from sunflower seeds. *Biosci Biotechnol Biochem* 65, 969-972.
- Lah, T.T., Kokalj-Kunovar, M. & Turk, V. 1990. Cysteine proteinase inhibitors in human cancerous tissues and fluids. *Biol Chem Hoppe Seyler* 371 Suppl, 199-203.
- Lah, T.T., Kokalj-Kunovar, M., Drobnic-Kosorok, M., Babnik, J., Golouh, R., Vrhovec, I. & Turk, V. 1992a. Cystatins and cathepsins in breast carcinoma. *Biol Chem Hoppe Seyler* 373, 595-604.
- Lah, T.T., Kokalj-Kunovar, M., Kastelic, L., Babnik, J., Stolf, A., Rainer, S. & Turk, V. 1992b. Cystatins and stefins in ascites fluid from ovarian carcinoma. *Cancer Lett* 61, 243-253.
- Lalioti, M.D., Mirotsoy, M., Buresi, C., Peitsch, M.C., Rossier, C., Ouazzani, R., Baldy-Moulinier, M., Bottani, A., Malafosse, A. & Antonarakis, S.E. 1997a. Identification of mutations in cystatin B, the gene responsible for the Unverricht-Lundborg type of progressive myoclonus epilepsy (EPM1). *Am J Hum Genet* 60, 342-351.
- Lalioti, M.D., Scott, H.S., Buresi, C., Rossier, C., Bottani, A., Morris, M.A., Malafosse, A. & Antonarakis, S.E. 1997b. Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. *Nature* 386, 847-851.
- Lawrence, J.C. & Nielsen, S.S. 2001. Partial isolation and characterization of a cysteine proteinase inhibitor from Lima bean (*Phaseolus lunatus*). *J Agric Food Chem* 49, 1020-1025.
- Lenarcic, B. & Bevec, T. 1998. Thyropins--new structurally related proteinase inhibitors. *Biol Chem* 379, 105-111.
- Lenarcic, B. & Turk, V. 1999. Thyroglobulin type-1 domains in equistatin inhibit both papain-like cysteine proteinases and cathepsin D. *J Biol Chem* 274, 563-566.
- Lenarcic, B., Ritonja, A., Sali, A., Kotnik, M., Turk, V., Machleidt, W. 1986. Properties and structure of human spleen stefin B – a low molecular weight protein inhibitor of cysteine proteinases. In: Turk, V., ed. *Cysteine Proteinases and Their Inhibitors*. Berlin: Walter de Gruyter. 473-487.
- Lenarcic, B., Ritonja, A., Dolenc, I., Stoka, V., Berbic, S., Pungercar, J., Strukelj, B. & Turk, V. 1993. Pig leukocyte cysteine proteinase inhibitor (PLCPI), a new member of the stefin family. *FEBS Lett* 336, 289-292.
- Lenarcic, B., Krizaj, I., Zunec, P. & Turk, V. 1996. Differences in specificity for the interactions of stefins A, B and D with cysteine proteinases. *FEBS Lett* 395, 113-118.
- Li, Y., Friel, P.J., Robinson, M.O., McLean, D.J. & Griswold, M.D. 2002. Identification and Characterization of Testis- and Epididymis-Specific Genes: Cystatin SC and Cystatin TE-1. *Biol Reprod* 67, 1872-1880.
- Lindahl, P., Alriksson, E., Jörnval, H. & Björk, I. 1988. Interaction of the cysteine proteinase inhibitor chicken cystatin with papain. *Biochemistry* 27, 5074-5082.
- Lindahl, P., Abrahamson, M. & Björk, I. 1992a. Interaction of recombinant human cystatin C with the cysteine proteinases papain and actinidin. *Biochem J* 281, 49-55.
- Lindahl, P., Nycander, M., Ylinenjärvi, K., Pol, E. & Björk, I. 1992b. Characterization by rapid-kinetic and equilibrium methods of the interaction between N-terminally truncated forms of chicken cystatin and the cysteine proteinases papain and actinidin. *Biochem J* 286, 165-171.
- Lindahl, P., Ripoll, D., Abrahamson, M., Mort, J.S. & Storer, A.C. 1994. Evidence for the interaction of valine-10 in cystatin C with the S2 subsite of cathepsin B. *Biochemistry* 33, 4384-4392.
- Locsey, L., Szegedi, J., Dan, A., Gorogh, S., Toths, E. 2001. Homocysteine and cystatin C level changes in haemodialysed patients and connection with cerebro- and cardiovascular complications. *Acta Physiol Hung* 88, 293-299.
- Machleidt, W., Borchart, U., Fritz, H., Brzin, J., Ritonja, A. & Turk, V. 1983. Protein inhibitors of cysteine proteinases. II. Primary structure of stefin, a cytosolic protein inhibitor of cysteine proteinases from human polymorphonuclear granulocytes. *Hoppe Seylers Z Physiol Chem* 364, 1481-1486.
- Machleidt, W., Thiele, U., Laber, B., Assfalg-Machleidt, I., Esterl, A., Wiegand, G., Kos, J., Turk, V. & Bode, W. 1989. Mechanism of inhibition of papain by chicken egg white

- cystatin. Inhibition constants of N-terminally truncated forms and cyanogen bromide fragments of the inhibitor. *FEBS Lett* 243, 234-238.
- Machleidt, W., Thiele, U., Assfalg-Machleidt, I., Forger, D. & Auerswald, E.A. 1991. Molecular mechanism of inhibition of cysteine proteinases by their protein inhibitors: kinetic studies with natural and recombinant variants of cystatins and stefins. *Biomed Biochim Acta* 50, 613-620.
- Machleidt, W., Nägler, D.K., Assfalg-Machleidt, I., Stubbs, M.T., Fritz, H. & Auerswald, E.A. 1995. Temporary inhibition of papain by hairpin loop mutants of chicken cystatin. Distorted binding of the loops results in cleavage of the Gly(9)-Ala10 bond. *FEBS Lett* 361, 185-190.
- Mangge, H., Liebmann, P., Tanil, H., Herrmann, J., Wagner, C., Gallistl, S., Schauenstein, K. & Erwa, W. 2000. Cystatin C, an early indicator for incipient renal disease in rheumatoid arthritis. *Clin Chim Acta* 300, 195-202.
- Manoury, B., Gregory, W.F., Maizels, R.M. & Watts, C. 2001. Bm-CPI-2, a cystatin homolog secreted by the filarial parasite *Brugia malayi*, inhibits class II MHC-restricted antigen processing. *Curr Biol* 11, 447-451.
- Margis, R., Reis, E.M. & Villeret, V. 1998. Structural and phylogenetic relationships among plant and animal cystatins. *Arch Biochem Biophys* 359, 24-30.
- Martin, J.R., Jerala, R., Kroon-Zitko, L., Zerovnik, E., Turk, V. & Waltho, J.P. 1994. Structural characterisation of human stefin A in solution and implications for binding to cysteine proteinases. *Eur J Biochem* 225, 1181-1194.
- Martin, J.R., Craven, C.J., Jerala, R., Kroon-Zitko, L., Zerovnik, E., Turk, V. & Waltho, J.P. 1995. The three-dimensional solution structure of human stefin A. *J Mol Biol* 246, 331-343.
- Mason, R.W., Sol-Church, K. & Abrahamson, M. 1998. Amino acid substitutions in the N-terminal segment of cystatin C create selective protein inhibitors of lysosomal cysteine proteinases. *Biochem J* 330, 833-838.
- McGrath, M.E. 1999. The lysosomal cysteine proteases. *Annu Rev Biophys Biomol Struct* 28, 181-204.
- Mirtti, T., Alanen, K., Kallajoki, M., Rinne, A. & Söderström, K.O. 2003. Expression of cystatins, high molecular weight cytokeratin, and proliferation markers in prostatic adenocarcinoma and hyperplasia. *Prostate* 54, 290-298.
- Misaka, T., Kuroda, M., Iwabuchi, K., Abe, K. & Arai, S. 1996. Soyacystatin, a novel cysteine proteinase inhibitor in soybean, is distinct in protein structure and gene organization from other cystatins of animal and plant origin. *Eur J Biochem* 240, 609-614.
- Monteiro, A.C., Abrahamson, M., Lima, A.P., Vannier-Santos, M.A. & Scharfstein, J. 2001. Identification, characterization and localization of chagasin, a tight-binding cysteine protease inhibitor in *Trypanosoma cruzi*. *J Cell Sci* 114, 3933-3942.
- Morita, M., Yoshiuchi, N., Arakawa, H. & Nishimura, S. 1999. CMAP: a novel cystatin-like gene involved in liver metastasis. *Cancer Res* 59, 151-158.
- Müller-Esterl, W., Lottspeich, F., Kellermann, J. & Henschen, A. 1986. On the evolution of mammalian kininogens. *Behring Inst Mitt* 48-61.
- Musil, D., Zucic, D., Turk, D., Engh, R.A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N. & *et al.* 1991. The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *Embo J* 10, 2321-2330.
- Nagai, A., Kobayashi, S., Shimode, K., Imaoka, K., Umegae, N., Fujihara, S. & Nakamura, M. 1998. No mutations in cystatin C gene in cerebral amyloid angiopathy with cystatin C deposition. *Mol Chem Neuropathol* 33, 63-78.
- Nägler, D.K., Storer, A.C., Portaro, F.C., Carmona, E., Juliano, L. & Ménard, R. 1997. Major increase in endopeptidase activity of human cathepsin B upon removal of occluding loop contacts. *Biochemistry* 36, 12608-12615.
- Nägler, D.K., Zhang, R., Tam, W., Sulea, T., Purisima, E.O. & Ménard, R. 1999. Human cathepsin X: A cysteine protease with unique carboxypeptidase activity. *Biochemistry* 38, 12648-12654.



- Naito, Y., Sasaki, M., Umemoto, T., Namikawa, I., Sakae, K., Ishihara, Y., Isomura, S. & Suzuki, I. 1995. Bactericidal effect of rat cystatin S on an oral bacterium *Porphyromonas gingivalis*. *Comp Biochem Physiol C Pharmacol Toxicol Endocrinol* 110, 71-75.
- Nakagawa, T., Roth, W., Wong, P., Nelson, A., Farr, A., Deussing, J., Villadangos, J.A., Ploegh, H., Peters, C. & Rudensky, A.Y. 1998. Cathepsin L: critical role in T cell degradation and CD4 T cell selection in the thymus. *Science* 280, 450-453.
- Nathanson, C.M., Wasselius, J., Wallin, H. & Abrahamson, M. 2002. Regulated expression and intracellular localization of cystatin F in human U937 cells. *Eur J Biochem* 269, 5502-5511.
- Ni, J., Abrahamson, M., Zhang, M., Fernandez, M.A., Grubb, A., Su, J., Yu, G.L., Li, Y., Parmelee, D., Xing, L., Coleman, T.A., Gentz, S., Thotakura, R., Nguyen, N., Hesselberg, M. & Gentz, R. 1997. Cystatin E is a novel human cysteine proteinase inhibitor with structural resemblance to family 2 cystatins. *J Biol Chem* 272, 10853-10858.
- Ni, J., Fernandez, M.A., Danielsson, L., Chillakuru, R.A., Zhang, J., Grubb, A., Su, J., Gentz, R. & Abrahamson, M. 1998. Cystatin F is a glycosylated human low molecular weight cysteine proteinase inhibitor. *J Biol Chem* 273, 24797-24804.
- Nikawa, T., Towatari, T., Ike, Y. & Katunuma, N. 1989. Studies on the reactive site of the cystatin superfamily using recombinant cystatin A mutants. Evidence that the QVVAG region is not essential for cysteine proteinase inhibitory activities. *FEBS Lett* 255, 309-314.
- Nishida, Y., Sumi, H. & Mihara, H. 1984. A thiol protease inhibitor released from cultured human malignant melanoma cells. *Cancer Res* 44, 3324-3329.
- Nycander, M. & Björk, I. 1990. Evidence by chemical modification that tryptophan-104 of the cysteine-proteinase inhibitor chicken cystatin is located in or near the proteinase-binding site. *Biochem J* 271, 281-284.
- Nycander, M., Estrada, S., Mort, J.S., Abrahamson, M. & Björk, I. 1998. Two-step mechanism of inhibition of cathepsin B by cystatin C due to displacement of the proteinase occluding loop. *FEBS Lett* 422, 61-64.
- Ohtani, O., Fukuyama, K. & Epstein, W.L. 1982. Biochemical properties of thiol proteinase inhibitor purified from psoriatic scales. *J Invest Dermatol* 78, 280-284.
- Olafsson, I., Löfberg, H., Abrahamson, M. & Grubb, A. 1988. Production, characterization and use of monoclonal antibodies against the major extracellular human cysteine proteinase inhibitors cystatin C and kininogen. *Scand J Clin Lab Invest* 48, 573-582.
- Olafsson, I., Thorsteinsson, L. & Jansson, O. 1996. The molecular pathology of hereditary cystatin C amyloid angiopathy causing brain hemorrhage. *Brain Pathol* 6, 121-126.
- Olsen, J.G., Kadziola, A., Lauritzen, C., Pedersen, J., Larsen, S. & Dahl, S.W. 2001. Tetrameric dipeptidyl peptidase I directs substrate specificity by use of the residual pro-part domain. *FEBS Lett* 506, 201-206.
- Olsson, S.L., Ek, B. & Björk, I. 1999. The affinity and kinetics of inhibition of cysteine proteinases by intact recombinant bovine cystatin C. *Biochim Biophys Acta* 1432, 73-81.
- Otto, H.H. & Schirmeister, T. 1997. Cysteine Proteases and Their Inhibitors. *Chem Rev* 97, 133-172.
- Pelouille, S., Esnard, A., Dacheux, J.L., Guillou, F., Gauthier, F. & Esnard, F. 1997. Interactions between ovine cathepsin L, cystatin C and alpha 2-macroglobulin. Potential role in the genital tract. *Eur J Biochem* 244, 140-146.
- Pennacchio, L.A., Lehesjoki, A.E., Stone, N.E., Willour, V.L., Virtaneva, K., Miao, J., D'Amato, E., Ramirez, L., Faham, M., Koskiniemi, M., Warrington, J.A., Norio, R., de la Chapelle, A., Cox, D.R. & Myers, R.M. 1996. Mutations in the gene encoding cystatin B in progressive myoclonus epilepsy (EPM1). *Science* 271, 1731-1734.
- Pennacchio, L.A., Bouley, D.M., Higgins, K.M., Scott, M.P., Noebels, J.L. & Myers, R.M. 1998. Progressive ataxia, myoclonic epilepsy and cerebellar apoptosis in cystatin B-deficient mice. *Nat Genet* 20, 251-258.

- Pham, C.T. & Ley, T.J. 1999. Dipeptidyl peptidase I is required for the processing and activation of granzymes A and B *in vivo*. *Proc Natl Acad Sci U S A* 96, 8627-8632.
- Pol, E. & Björk, I. 1999. Importance of the second binding loop and the C-terminal end of cystatin B (stefin B) for inhibition of cysteine proteinases. *Biochemistry* 38, 10519-10526.
- Pol, E. & Björk, I. 2001. Role of the single cysteine residue, Cys 3, of human and bovine cystatin B (stefin B) in the inhibition of cysteine proteinases. *Protein Sci* 10, 1729-1738.
- Pol, E. & Björk, I. 2003. Contributions of individual residues in the N-terminal region of cystatin B (stefin B) to inhibition of cysteine proteinases. *Biochim Biophys Acta* 1645, 105-112.
- Pol, E., Olsson, S.L., Estrada, S., Prasthofer, T.W. & Björk, I. 1995. Characterization by spectroscopic, kinetic and equilibrium methods of the interaction between recombinant human cystatin A (stefin A) and cysteine proteinases. *Biochem J* 311, 275-282.
- Poulik, M.D., Shinnick, C.S. & Smithies, O. 1981. Partial amino acid sequences of human and dog post-gamma globulins. *Mol Immunol* 18, 569-572.
- Rawlings, N.D. & Barrett, A.J. 1990. Evolution of proteins of the cystatin superfamily. *J Mol Evol* 30, 60-71.
- Rawlings, N.D. & Barrett, A.J. 1993. Evolutionary families of peptidases. *Biochem J* 290, 205-218.
- Rawlings, N.D. & Barrett, A.J. 1999. MEROPS: the peptidase database. *Nucleic Acids Res* 27, 325-331.
- Rawlings, N.D. & Barrett, A.J. 2000. MEROPS: the peptidase database. *Nucleic Acids Res* 28, 323-325.
- Rawlings, N.D., O'Brien, E. & Barrett, A.J. 2002. MEROPS: the protease database. *Nucleic Acids Res* 30, 343-346.
- Riccio, M., Di Giaimo, R., Pianetti, S., Palmieri, P.P., Melli, M. & Santi, S. 2001. Nuclear localization of cystatin B, the cathepsin inhibitor implicated in myoclonus epilepsy (EPM1). *Exp Cell Res* 262, 84-94.
- Rigden, D.J., Monteiro, A.C. & Grossi de Sa, M.F. 2001. The protease inhibitor chagasin of *Trypanosoma cruzi* adopts an immunoglobulin-type fold and may have arisen by horizontal gene transfer. *FEBS Lett* 504, 41-44.
- Rigden, D.J., Mosolov, V.V. & Galperin, M.Y. 2002. Sequence conservation in the chagasin family suggests a common trend in cysteine proteinase binding by unrelated protein inhibitors. *Protein Sci* 11, 1971-1977.
- Rinne, A., Järvinen, M., Rasanen, O. & Hopsu-Havu, V.K. 1984a. Acid and neutral cysteine proteinase inhibitor in normal uterine portio and in squamo-epithelial metaplasia, dysplasias and infiltrative carcinoma of the uterine portio. *Exp Pathol* 26, 67-70.
- Rinne, A., Rasanen, O., Järvinen, M., Dammert, K., Kallioinen, M. & Hopsu-Havu, V.K. 1984b. Occurrence of acid and neutral cysteine proteinase inhibitors in epidermal malignancies: immunohistochemical study. *Acta Histochem* 74, 75-79.
- Ritonja, A., Machleidt, W. & Barrett, A.J. 1985. Amino acid sequence of the intracellular cysteine proteinase inhibitor cystatin B from human liver. *Biochem Biophys Res Commun* 131, 1187-1192.
- Ritonja, A., Evans, H.J., Machleidt, W. & Barrett, A.J. 1987. Amino acid sequence of a cystatin from venom of the African puff adder (*Bitis arietans*). *Biochem J* 246, 799-802.
- Ritonja, A., Coetzer, T.H., Pike, R.N. & Dennison, C. 1996. The amino acid sequences, structure comparisons and inhibition kinetics of sheep cathepsin L and sheep stefin B. *Comp Biochem Physiol B Biochem Mol Biol* 114, 193-198.
- Rogelj, B., Popovic, T., Ritonja, A., Strukelj, B. & Brzin, J. 1998. Chelidocystatin, a novel phytocystatin from *Chelidonium majus*. *Phytochemistry* 49, 1645-1649.
- Roth, W., Deussing, J., Botchkarev, V.A., Pauly-Evers, M., Saftig, P., Hafner, A., Schmidt, P., Schmahl, W., Scherer, J., Anton-Lamprecht, I., Von Figura, K., Paus, R. & Peters, C.

2000. Cathepsin L deficiency as molecular defect of furless: hyperproliferation of keratinocytes and perturbation of hair follicle cycling. *Faseb J* 14, 2075-2086.
- Saftig, P., Hunziker, E., Everts, V., Jones, S., Boyde, A., Wehmeyer, O., Suter, A. & von Figura, K. 2000. Functions of cathepsin K in bone resorption. Lessons from cathepsin K deficient mice. *Adv Exp Med Biol* 477, 293-303.
- Saito, H., Suzuki, T., Ueno, K., Kubo, T. & Natori, S. 1989. Molecular cloning of cDNA for sarcocystatin A and analysis of the expression of the sarcocystatin A gene during development of *Sarcophaga peregrina*. *Biochemistry* 28, 1749-1755.
- Salvesen, G., Parkes, C., Abrahamson, M., Grubb, A. & Barrett, A.J. 1986. Human low-Mr kininogen contains three copies of a cystatin sequence that are divergent in structure and in inhibitory activity for cysteine proteinases. *Biochem J* 234, 429-434.
- Schechter, I. & Berger, A. 1967. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* 27, 157-162.
- Schick, C., Pemberton, P.A., Shi, G.P., Kamachi, Y., Cataltepe, S., Bartuski, A.J., Gornstein, E.R., Brömme, D., Chapman, H.A. & Silverman, G.A. 1998. Cross-class inhibition of the cysteine proteinases cathepsins K, L, and S by the serpin squamous cell carcinoma antigen 1: a kinetic analysis. *Biochemistry* 37, 5258-5266.
- Shannon, P., Pennacchio, L.A., Houseweart, M.K., Minassian, B.A., Myers, R.M. 2002. Neuropathological changes in a mouse model of progressive myoclonus epilepsy: cystatin B deficiency and Unverricht-Lundborg disease. *J Neuropathol Exp Neurol* 61, 1085-1091.
- Shi, G.P., Villadangos, J.A., Dranoff, G., Small, C., Gu, L., Haley, K.J., Riese, R., Ploegh, H.L. & Chapman, H.A. 1999. Cathepsin S required for normal MHC class II peptide loading and germinal center development. *Immunity* 10, 197-206.
- Shi, G.P., Bryant, R.A., Riese, R., Verhelst, S., Driessen, C., Li, Z., Brömme, D., Ploegh, H.L. & Chapman, H.A. 2000. Role for cathepsin F in invariant chain processing and major histocompatibility complex class II peptide loading by macrophages. *J Exp Med* 191, 1177-1186.
- Shibuya, K., Kaji, H., Itoh, T., Ohyama, Y., Tsujikami, A., Tate, S., Takeda, A., Kumagai, I., Hirao, I., Miura, K., Inagaki, G. & Samejima, T. 1995a. Human cystatin A is inactivated by engineered truncation. The NH<sub>2</sub>-terminal region of the cysteine proteinase inhibitor is essential for expression of its inhibitory activity. *Biochemistry* 34, 12185-12192.
- Shibuya, K., Kaji, H., Ohyama, Y., Tate, S., Kainosho, M., Inagaki, F. & Samejima, T. 1995b. Significance of the highly conserved Gly-4 residue in human cystatin A. *J Biochem (Tokyo)* 118, 635-642.
- Shiraishi, T., Mori, M., Tanaka, S., Sugimachi, K. & Akiyoshi, T. 1998. Identification of cystatin B in human esophageal carcinoma, using differential displays in which the gene expression is related to lymph-node metastasis. *Int J Cancer* 79, 175-178.
- Shoemaker, K., Holloway, J.L., Whitmore, T.E., Maurer, M. & Feldhaus, A.L. 2000. Molecular cloning, chromosome mapping and characterization of a testis-specific cystatin-like cDNA, cystatin T. *Gene* 245, 103-108.
- Sivaraman, J., Nägler, D.K., Zhang, R., Ménard, R. & Cygler, M. 2000. Crystal structure of human procathepsin X: a cysteine protease with the proregion covalently linked to the active site cysteine. *J Mol Biol* 295, 939-951.
- Söderström, K.O., Laato, M., Wu, P., Hopsu-Havu, V.K., Nurmi, M. & Rinne, A. 1995. Expression of acid cysteine proteinase inhibitor (ACPI) in the normal human prostate, benign prostatic hyperplasia and adenocarcinoma. *Int J Cancer* 62, 1-4.
- Sol-Church, K., Picerno, G.N., Stabley, D.L., Frenck, J., Xing, S., Bertenshaw, G.P., Mason, R.W. 2002. Evolution of placentally expressed cathepsins. *Biochem Biophys Res Commun* 293, 23-9.
- Solem, M., Rawson, C., Lindburg, K. & Barnes, D. 1990. Transforming growth factor beta regulates cystatin C in serum-free mouse embryo (SFME) cells. *Biochem Biophys Res Commun* 172, 945-951.
- Song, I., Taylor, M., Baker, K. & Bateman, R.C., Jr. 1995. Inhibition of cysteine proteinases by *Carica papaya* cystatin produced in *Escherichia coli*. *Gene* 162, 221-224.

- Sotiropoulou, G., Anisowicz, A. & Sager, R. 1997. Identification, cloning, and characterization of cystatin M, a novel cysteine proteinase inhibitor, down-regulated in breast cancer. *J Biol Chem* 272, 903-910.
- Sottrup-Jensen, L. 1989. Alpha-macroglobulins: structure, shape, and mechanism of proteinase complex formation. *J Biol Chem* 264, 11539-11542.
- Spanier, A.M. & Bird, J.W. 1982. Endogenous cathepsin B inhibitor activity in normal and myopathic red and white skeletal muscle. *Muscle Nerve* 5, 313-320.
- Stoka, V., Nycander, M., Lenarcic, B., Labriola, C., Cazzulo, J.J., Björk, I. & Turk, V. 1995. Inhibition of cruzipain, the major cysteine proteinase of the protozoan parasite, *Trypanosoma cruzi*, by proteinase inhibitors of the cystatin superfamily. *FEBS Lett* 370, 101-104.
- Storer, A.C. & Ménard, R. 1994. Catalytic mechanism in papain family of cysteine peptidases. *Methods Enzymol* 244, 486-500.
- Stubbs, M.T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B. & Turk, V. 1990. The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *Embo J* 9, 1939-1947.
- Sun, H., Li, N., Wang, X., Liu, S., Chen, T., Zhang, L., Wan, T. & Cao, X. 2003. Molecular cloning and characterization of a novel cystatin-like molecule, CLM, from human bone marrow stromal cells. *Biochem Biophys Res Commun* 301, 176-182.
- Suzuki, T. & Natori, S. 1985. Purification and characterization of an inhibitor of the cysteine protease from the hemolymph of *Sarcophaga peregrina* larvae. *J Biol Chem* 260, 5115-5120.
- Szpaderska, A.M. & Frankfater, A. 2001. An intracellular form of cathepsin B contributes to invasiveness in cancer. *Cancer Res* 61, 3493-3500.
- Takahashi, M., Tezuka, T. & Katunuma, N. 1994. Inhibition of growth and cysteine proteinase activity of *Staphylococcus aureus* V8 by phosphorylated cystatin alpha in skin cornified envelope. *FEBS Lett* 355, 275-278.
- Takeda, A., Kaji, H., Nakaya, K., Aoki, Y., Nakamura, Y. & Samejima, T. 1985. Amino acid sequence of derivatives of newborn rat epidermal thiol proteinase inhibitor. *Biochem Int* 11, 557-564.
- Takeda, A., Yamamoto, T., Nakamura, Y., Takahashi, T. & Hibino, T. 1995. Squamous cell carcinoma antigen is a potent inhibitor of cysteine proteinase cathepsin L. *FEBS Lett* 359, 78-80.
- Takio, K., Kominami, E., Bando, Y., Katunuma, N. & Titani, K. 1984. Amino acid sequence of rat epidermal thiol proteinase inhibitor. *Biochem Biophys Res Commun* 121, 149-154.
- Takio, K., Kominami, E., Wakamatsu, N., Katunuma, N. & Titani, K. 1983. Amino acid sequence of rat liver thiol proteinase inhibitor. *Biochem Biophys Res Commun* 115, 902-908.
- Tate, S., Ushioda, T., Utsunomiya-Tate, N., Shibuya, K., Ohyama, Y., Nakano, Y., Kaji, H., Inagaki, F., Samejima, T. & Kainosho, M. 1995. Solution structure of a human cystatin A variant, cystatin A2-98 M65L, by NMR spectroscopy. A possible role of the interactions between the N- and C-termini to maintain the inhibitory active form of cystatin A. *Biochemistry* 34, 14637-14648.
- Thiele, U., Auerswald, E.A., Gebhard, W., Assfalg-Machleidt, I., Popovic, T. & Machleidt, W. 1988. Inhibitorily active recombinant human stefin B. Gene synthesis, expression and isolation of an inhibitory active MS-2 pol-stefin B fusion protein and preparation of Des[Met1,2(2)]stefin B. *Biol Chem Hoppe Seyler* 369, 1167-1178.
- Thiele, U., Assfalg-Machleidt, I., Machleidt, W. & Auerswald, E.A. 1990. N-terminal variants of recombinant stefin B: effect on affinity for papain and cathepsin B. *Biol Chem Hoppe Seyler* 371 Suppl, 125-136.
- Tohonen, V., Österlund, C. & Nordqvist, K. 1998. Testatin: a cystatin-related gene expressed during early testis development. *Proc Natl Acad Sci U S A* 95, 14208-14213.
- Toomes, C., James, J., Wood, A.J., Wu, C.L., McCormick, D., Lench, N., Hewitt, C., Moynihan, L., Roberts, E., Woods, C.G., Markham, A., Wong, M., Widmer, R., Ghaffar,

- K.A., Pemberton, M., Hussein, I.R., Temtamy, S.A., Davies, R., Read, A.P., Sloan, P., Dixon, M.J. & Thakker, N.S. 1999. Loss-of-function mutations in the cathepsin C gene result in periodontal disease and palmoplantar keratosis. *Nat Genet* 23, 421-424.
- Troeberg, L., Pike, R.N., Morty, R.E., Berry, R.K., Coetzer, T.H. & Lonsdale-Eccles, J.D. 1996. Proteases from *Trypanosoma brucei brucei*. Purification, characterisation and interactions with host regulatory molecules. *Eur J Biochem* 238, 728-736.
- Tsai, Y.J., Chang, G.D., Huang, C.J., Chang, Y.S. & Huang, F.L. 1996. Purification and molecular cloning of carp ovarian cystatin. *Comp Biochem Physiol B Biochem Mol Biol* 113, 573-580.
- Tseng, C.C., Tseng, C.P., Levine, M.J. & Bobek, L.A. 2000. Differential effect toward inhibition of papain and cathepsin C by recombinant human salivary cystatin SN and its variants produced by a baculovirus system. *Arch Biochem Biophys* 380, 133-140.
- Tsui, F.W., Tsui, H.W., Mok, S., Mlinaric, I., Copeland, N.G., Gilbert, D.J., Jenkins, N.A. & Siminovitch, K.A. 1993. Molecular characterization and mapping of murine genes encoding three members of the stefin family of cysteine proteinase inhibitors. *Genomics* 15, 507-514.
- Tu, G.F., Cole, T., Southwell, B.R. & Schreiber, G. 1990. Expression of the genes for transthyretin, cystatin C and beta A4 amyloid precursor protein in sheep choroid plexus during development. *Brain Res Dev Brain Res* 55, 203-208.
- Turk, V. & Bode, W. 1991. The cystatins: protein inhibitors of cysteine proteinases. *FEBS Lett* 285, 213-219.
- Turk, V., Brzin, J., Longer, M., Ritonja, A., Eropkin, M., Borchart, U. & Machleidt, W. 1983. Protein inhibitors of cysteine proteinases. III. Amino-acid sequence of cystatin from chicken egg white. *Hoppe Seylers Z Physiol Chem* 364, 1487-1496.
- Turk, B., Krizaj, I. & Turk, V. 1992. Isolation and characterization of bovine stefin B. *Biol Chem Hoppe Seyler* 373, 441-446.
- Turk, B., Krizaj, I., Kralj, B., Dolenc, I., Popovic, T., Bieth, J.G. & Turk, V. 1993. Bovine stefin C, a new member of the stefin family. *J Biol Chem* 268, 7323-7329.
- Turk, B., Colic, A., Stoka, V. & Turk, V. 1994. Kinetics of inhibition of bovine cathepsin S by bovine stefin B. *FEBS Lett* 339, 155-159.
- Turk, B., Ritonja, A., Björk, I., Stoka, V., Dolenc, I. & Turk, V. 1995a. Identification of bovine stefin A, a novel protein inhibitor of cysteine proteinases. *FEBS Lett* 360, 101-105.
- Turk, B., Stoka, V., Björk, I., Boudier, C., Johansson, G., Dolenc, I., Colic, A., Bieth, J.G. & Turk, V. 1995b. High-affinity binding of two molecules of cysteine proteinases to low-molecular-weight kininogen. *Protein Sci* 4, 1874-1880.
- Turk, B., Stoka, V., Turk, V., Johansson, G., Cazzulo, J.J. & Björk, I. 1996a. High-molecular-weight kininogen binds two molecules of cysteine proteinases with different rate constants. *FEBS Lett* 391, 109-112.
- Turk, D., Podobnik, M., Kuhelj, R., Dolinar, M. & Turk, V. 1996b. Crystal structures of human procathepsin B at 3.2 and 3.3 Angstroms resolution reveal an interaction motif between a papain-like cysteine protease & its propeptide. *FEBS Lett* 384, 211-214.
- Turk, B., Turk, V. & Turk, D. 1997. Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors. *Biol Chem* 378, 141-150.
- Turk, D., Guncar, G., Podobnik, M. & Turk, B. 1998. Revised definition of substrate binding sites of papain-like cysteine proteases. *Biol Chem* 379, 137-147.
- Turk, B., Turk, D. & Turk, V. 2000. Lysosomal cysteine proteases: more than scavengers. *Biochim Biophys Acta* 1477, 98-111.
- Turk, D., Janjic, V., Stern, I., Podobnik, M., Lamba, D., Dahl, S.W., Lauritzen, C., Pedersen, J., Turk, V. & Turk, B. 2001a. Structure of human dipeptidyl peptidase I (cathepsin C): exclusion domain added to an endopeptidase framework creates the machine for activation of granular serine proteases. *Embo J* 20, 6570-6582.
- Turk, V., Turk, B. & Turk, D. 2001b. Lysosomal cysteine proteases: facts and opportunities. *Embo J* 20, 4629-4633.

- Turk, B., Stoka, V., Rozman-Pungercar, J., Cirman, T., Droga-Mazovec, G., Oreic, K. & Turk, V. 2002a. Apoptotic pathways: involvement of lysosomal proteases. *Biol Chem* 383, 1035-1044.
- Turk, B., Turk, D. & Salvesen, G.S. 2002b. Regulating cysteine protease activity: essential role of protease inhibitors as guardians and regulators. *Curr Pharm Des* 8, 1623-1637.
- Turk, V., Turk, B., Guncar, G., Turk, D. & Kos, J. 2002c. Lysosomal cathepsins: structure, role in antigen processing and presentation, and cancer. *Adv Enzyme Regul* 42, 285-303.
- Urwin, P.E., Atkinson, H.J. & McPherson, M.J. 1995a. Involvement of the NH<sub>2</sub>-terminal region of oryzacystatin-I in cysteine proteinase inhibition. *Protein Eng* 8, 1303-1307.
- Urwin, P.E., Atkinson, H.J., Waller, D.A., McPherson, M.J. 1995b. Engineered oryzacystatin-I expressed in transgenic hairy roots confers resistance to *Globodera pallida*. *Plant J* 8, 121-131.
- van Duinen, S.G., Castano, E.M., Prelli, F., Bots, G.T., Luyendijk, W. & Frangione, B. 1987. Hereditary cerebral hemorrhage with amyloidosis in patients of Dutch origin is related to Alzheimer disease. *Proc Natl Acad Sci U S A* 84, 5991-5994.
- Villadangos, J.A., Bryant, R.A., Deussing, J., Driessen, C., Lennon-Dumenil, A.M., Riese, R.J., Roth, W., Saftig, P., Shi, G.P., Chapman, H.A., Peters, C., Ploegh, H.L. 1999. Proteases involved in MHC class II antigen presentation. *Immunol Rev* 172, 109-120.
- Vinters, H.V., Nishimura, G.S., Secor, D.L. & Pardridge, W.M. 1990. Immunoreactive A4 and gamma-trace peptide colocalization in amyloidotic arteriolar lesions in brains of patients with Alzheimer's disease. *Am J Pathol* 137, 233-240.
- Vray, B., Hartmann, S. & Hoebeke, J. 2000. Immunomodulatory properties of cystatins. *Cell Mol Life Sci* 59, 1503-1512.
- Waldron, C., Wegrich, L.M., Merlo, P.A. & Walsh, T.A. 1993. Characterization of a genomic sequence coding for potato multicystatin, an eight-domain cysteine proteinase inhibitor. *Plant Mol Biol* 23, 801-812.
- Wang, S.C. & Huang, F.L. 2002. Carp ovarian cystatin binds and agglutinates spermatozoa via electrostatic interaction. *Biol Reprod* 66, 1318-1327.
- Wassler, M., Syntin, P., Sutton-Walsh, H.G., Hsia, N., Hardy, D.M. & Cornwall, G.A. 2002. Identification and characterization of cystatin-related epididymal spermatogenic protein in human spermatozoa: localization in the equatorial segment. *Biol Reprod* 67, 795-803.
- Wattendorff, A.R., Bots, G.T., Went, L.N. & Endtz, L.J. 1982. Familial cerebral amyloid angiopathy presenting as recurrent cerebral haemorrhage. *J Neurol Sci* 55, 121-135.
- Wiederanders, B. 2000. The function of propeptide domains of cysteine proteinases. *Adv Exp Med Biol* 477, 261-270.
- Yang, Y., Sanchez, J.F., Strub, M.P., Brutscher, B. & Aumelas, A. 2003. NMR Structure of the Cathelin-like Domain of the Protegrin-3 Precursor. *Biochemistry* 42, 4669-4680.
- Yano, M., Hirai, K., Naito, Z., Yokoyama, M., Ishiwata, T., Shiraki, Y., Inokuchi, M. & Asano, G. 2001. Expression of cathepsin B and cystatin C in human breast cancer. *Surg Today* 31, 385-389.
- Ylinenjärvi, K., Widersten, M. & Björk, I. 1999. Hydrophobic sequences can substitute for the wild-type N-terminal sequence of cystatin A (stefin A) in tight binding to cysteine proteinases selection of high-affinity N-terminal region variants by phage display. *Eur J Biochem* 261, 682-688.
- Zajc, I., Sever, N., Bervar, A. & Lah, T.T. 2002. Expression of cysteine peptidase cathepsin L and its inhibitors stefins A & B in relation to tumorigenicity of breast cancer cell lines. *Cancer Lett* 187, 185-190.
- Zeeuwen, P.L., Van Vlijmen-Willems, I.M., Jansen, B.J., Sotiropoulou, G., Curfs, J.H., Meis, J.F., Janssen, J.J., Van Ruissen, F. & Schalkwijk, J. 2001. Cystatin M/E expression is restricted to differentiated epidermal keratinocytes and sweat glands: a new skin-specific proteinase inhibitor that is a target for cross-linking by transglutaminase. *J Invest Dermatol* 116, 693-701.

Zore, I., Krasovec, M., Cimerman, N., Kuhelj, R., Werle, B., Nielsen, H.J., Brunner, N. & Kos, J. 2001. Cathepsin B/cystatin C complex levels in sera from patients with lung and colorectal cancer. *Biol Chem* 382, 805-810.

## Acknowledgements

This work was done at the Division of Medical Biochemistry, Department of Molecular Biosciences (formerly known as the Department of Veterinary Medical Chemistry), the Biomedical Center, Swedish University of Agricultural Sciences, Uppsala.

I am deeply grateful to my mentor, professor Ingemar Björk, for the opportunity to carry out this research. I thank you, Ingemar, for all that knowledge you generously shared with me and I would only wish that the time you spent on reading, correcting, waiting *etc.* was not so enormous.

My special gratitude to Dr. Sergio Estrada for introducing me into the world of computers, kinetics and DNA recombinant techniques. Thank you for all explanations, discussions and listening. I also wish to thank Dr. Raymond Award, Dr. Maria Nycander and Dr. Svetlana Rylova for their kind support.

I appreciate much the help and pieces of advice I got from my colleagues, Dr. Ewa Pol, Ms. Kerstin Nordling, Ms. Karin Ylinenjärvi, Mr. Piotr Wlad, Dr. Liya Wang and Dr. Göran Hjälms, and other colleagues of the department. Particularly, I thank Dr. Elena Usova for the help in preparing the printable version of this thesis.

From all my heart, I thank my family, Mikael, Julia and Sonya, for all the patience, encouragement and help during the time this thesis was written.

The work was supported by a grant from the Swedish Science Council–Medicine.