COMPERATIVE VIEW INTO PROTEASE – INHIBITOR INTERACTION

Miha Renko

Doctoral Dissertation
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia, July 2010

Evaluation Board:

Prof. Dr. Igor Muševič, Jožef Stefan Institute, Jamova 39, Ljubljana, Slovenija

Prof. Dr. Janko Kos, The Faculty of Pharmacy, University of Ljubljana, Aškerčeva 7, Ljubljana, Slovenia

Prof. Dr. Peter Hinterdorfer, Institute for Biophysics, Johannes Kepler Universität Linz, Altenbergerstrasse 69, A-4040 Linz, Austria

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

Miha Renko

COMPERATIVE VIEW INTO PROTEASE – INHIBITOR INTERACTION

Doctoral Dissertation

PRIMERJALNI VPOGLED V INTERAKCIJE PROTEAZ IN NJIHOVIH INHIVITORJEV

Doktorska disertacija

Supervisor: Prof. Dr. Dušan Turk

Ljubljana, Slovenia, July 2010

Index

Abstract

Interactions between macromolecules are the basic events of life. Their characterization is thus a key to understanding the physiology and the underlying specificity of these events. Not only understanding of the final states (where the molecules get closest together), also understanding of the course of approach is mandatory to gain insight into the dynamical picture of molecular processes in a living organism.

We have made an attempt to gain insight into the distance dependence of interactions by investigating the tight interaction between proteases and their inhibitors. Their interactions have been studied by the combined use of X-ray crystallography, protein kinetics (enzyme kinetics and surface plasmon resonance) and atom force microscopy. Atomic force measurements provided the distance between the potential minimum and the transition state in the Bell's model. Additional information about the distance dependence was obtained by the kinetics of inactivated cathepsins. These data combined with the structures and models of complexes of inactivated cathepsins with stefin A, provided insight in the energetics of the distance dependence. The interactions between the studied proteases and inhibitors are taking place in the range below 10 Å (8.5 Å was the distance measured by the atom force microscopy). In a crude approximation, the interaction energy appears to be in a linear relationship with the intermolecular distance. These results shed light also on understanding of the final binding states, the geometry of which are best revealed by X-ray crystallography. Several determined structures thus provide insight into the mechanism of binding.

3-dimensional structures of complexes of reduced cathepsin V and stefin A, the MMTS-blocked cathepsins L and V in complex with stefin A, and the complex of the exopeptidase cathepsin B with stefin A were determined. They have shown that the modification of the active site cysteine residue prevents the genuine binding of the inhibitor to the protease, however, the complex is still formed. The binding of the N-terminal trunk and the first loop of the stefin A to cathepsins is well defined and is very similar between various cathepsins, whereas the second loop in stefin A exhibits large flexibility. We have also shown that the occluding loop in the cathepsin B is flexible and can adopt different conformations, depending on the size and shape of the inhibitor bound to the active site cleft. In contrary to the three loops in stefin A, clitocypin utilizes only two loops, which bind to the active site cleft. The crystal structures of cathepsin V in the complex with clitocypin enabled us to explain the previously unknown inhibition mechanism in atomic detail. The most interesting feature of this interaction is the peptide bond flip, which occurs prior to or concurrently with the inhibitor docking and enhances the inhibition by the formation of an additional hydrogen bond.

Povzetek

Interakcije med proteini so ključni dogodki v vseh življenjskih procesih, zato je njihova karakterizacija ključna v razumevanju fiziologije in z njo povezanih procesov. Za razumevanja dinamičnih procesov v celicah ni pomembno samo razumevanje končnih stanj, kjer so molekule vezane v komplekse, ampak tudi način približevanje molekul in orientacije pred tvorbo le-tega.

Z raziskovanjem močnih interakcij med proteazami in njihovimi inhibitorji smo poskušali pridobiti vpogled v nastanek kompleksov in njihove lastnosti v odvisnosti od medsebojne razdalje. Interakcije v kompleksih smo proučevali s pomočjo proteinske kristalografije in encimske kinetike, določene na klasičen način in s pomočjo površinske plazmonske resonance. Z mikroskopijo na atomsko silo smo določili razdaljo med potencialnim minimumom in prehodnim stanjem v Bellovem modelu, ki opisuje disociacijo dveh molekul. Dodatne informacije o tej razdalji smo pridobili z kinetiko tvorbe kompleksov med inaktiviranimi katepsini in stefinom A. Kinetični podatki, združeni z določenimi kristalnimi strukturami in izračunanimi modeli so nam omogočili vpogled v energijsko bilanco interakcije v odvisnosti od razdalje. Interakcije med proučevanimi encimi in njihovimi inhibitorji se pretežno dogajajo, ko je njihova medsebojna oddaljenost pod 10 Å (s pomočjo mikroskopije na atomsko silo smo določili 8.5 Å). V grobem približku lahko povzamemo, da je energije tvorbe kompleksa v linearni odvisnosti od razdalje med obema molekulama. Ti podatki so v pomoč tudi pri razumevanju končnega stanja, o katerem največ pove proteinska kristalografija. Več določenih kristalnih struktur nam poleg teh podatkov daje tudi vpogled v mehanizem tvorbe kompleksa.

Določili smo tridimenzionalne strukture kompleksov med reduciranim katepsinom V in stefinom A, med katepsinoma L in V, blokiranima z MMTS, in stefinom A ter strukturo kompleksa med eksopeptidazo katepsinom B in stefinom A. Te strukture so pokazale, da modifikacija aktivnega cisteina zmotijo nativno vezavo, ne preprečijo pa nastanek kompleksa. Prav tako lahko opazimo, da je vezava N-terminalnega konca in prve zanke v stefinu A v aktivno mesto katepsinom dobro definirana, dočim pa je druga zanka precej fleksibilna. Prav tako smo pokazali, da lahko izključitvena zanka v katepsinu B zavzame različne konformacije, ki so odvisne od vrste inhibitorja, ki se veže v aktivno mesto. V nasprotju s stefini, ki za inhibicijo uporabljajo vezavo z tremi zankami, se klitocipin veže v aktivno mesto s pomočjo le dveh zank. Kristalna struktura katepsina V v kompleksu s klitocipinom nam je omogočila, da smo na atomarnem nivoju pojasnili prej še neznan mehanizem inhibicije cistenskih proteaz s mikocipini. Pri vezavi klitocipina je najbolj zanimiva sprememba konformacije peptidne vezi med glicinoma, ki nastane pred ali sočasno z vezavo inhibitorja v aktivno mesto.

Abbreviations

AEP	=	Asparaginyl endopeptidase or legumain
AFM	=	Atomic force microscopy
AMC	=	7-Amino-4-methylcoumarin
BA	=	Complex of cathepsin B and stefin A
BBCI	=	Bauhinia bauhinioides cruzipain inhibitor
catB	=	Cathepsin B
catH	=	Cathepsin H
catL	=	Cathepsin L
catV	=	Cathepsin V
Clt	=	Clitocypin
cmLA	=	Complex of carboxymethylated cathepsin L and stefin A
Cst	=	Cystatin
DMSO	=	Dimethyl sulfoxide
DTT	=	Dithiothreitol
EC	=	Enzyme Commission number
EDTA	=	ethylenediaminetetraacetic acid
LA	=	Complex of MMTS-blocked cathepsin L and stefin A
MAD	=	Multiple Wavelength Anomalous Diffraction
Mcp	=	Macrocypin
MHC	=	Major histocompatibility complex
MMTS	=	Methyl methanethiosulfonate
NHS	=	N-hydroxysuccinimide
OD600	=	Optical Density at 600 nm
PEG	=	Polyethylene glyco
PCR	=	Polymerase chain reaction
PDB	=	Protein Data Bank
rVA	=	Complex of reduced cathepsin V and stefin A
RMSD	=	Root mean square deviation
RU	=	Response unit
SAD	=	Single Wavelength Anomalous Diffraction
SDS-PAGE	=	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SeMet	=	Selenomethionine
SPR	=	Surface Plasmon Resonance
steA	=	Stefin A
STI	=	Soybean tripsin inhibitor
TCEP	=	tris(2-carboxyethyl)phosphine
VA	=	Complex of MMTS-blocked cathepsin L and stefin

1 Introduction

Living organisms are an exceptionally complex combination of a variety of molecular processes needed for their proper functioning. All processes in the living organisms involve proteins. Many proteins are enzymes that catalyze biochemical reactions and are vital to metabolism. Besides catalysing chemical reactions, proteins also have structural or mechanical functions, such as actin and myosin in muscle and proteins in the cytoskeleton which form a system of scaffolds that maintains the cell shape.

Enzymes are one of the most studied groups of proteins. They perform most of the reactions involved in metabolism. The rate acceleration conferred by enzymatic catalysis is often enormous — as much as 10^{17}-fold increase in rate compared to the uncatalyzed reaction in the case of orotate decarboxylase (78 million years without the enzyme, 18 milliseconds with the enzyme) [1]. On the other hand, enzymes are usually highly specific and accelerate only one or a few chemical reactions.

One of the most interesting and researched representatives among enzymes are proteases. They catalyse the hydrolysis of peptide bonds in proteins and peptides. Proteases (EC 3.4.), also known as proteinases or proteolytic enzymes, are involved in digesting long protein chains into short fragments by splitting the peptide bonds that link the amino acid residues. Some of them can detach the terminal amino acids from the protein chain (exopeptidases); others cleave internal peptide bonds of a protein (endopeptidases). Peptidases can either break specific peptide bonds (limited proteolysis), depending on the amino acid sequence of a protein, or break down a complete peptide to peptides and amino acids (unlimited proteolysis).

Proteases are divided into six major groups according to the character of their catalytic active site and conditions of action: serine proteases, cysteine proteases, aspartic proteases, threonine proteases, glutamic acid proteases and metallo proteases. Attachment of a protease to a certain group depends on the structure of the catalytic site and the amino acid (as one of the constituents of the catalytic site) essential for its activity.

Proteases are usually synthesized as precursor proteins called zymogens, such as serine protease precursors trypsinogen and chymotrypsinogen, and aspartic protease precursor pepsinogen. The proteases are activated by the removal of an inhibitory segment of protein or by a specific cleavage, followed by conformational changes. Activation occurs once the protease is delivered to a specific intracellular compartment (e.g. lysosome) or extracellular environment (e.g. stomach). This system prevents the cell that produces the protease from being damaged by it.

Proteases occur naturally in all organisms. In the human body, there are approximately 500 known proteases [2]. The estimated number of proteases is slightly higher, as the initial estimation based on the human genome sequence suggested there is between 700 and 1100 proteases [2]. Later detailed genomic analyses have revealed that the human degradome is composed between 561 [3] and 570 [4] protease-coding genes.

Proteases are involved in a multitude of physiological reactions from simple digestion of food proteins to highly regulated cascades. For example, pepsin, trypsin, chymotrypsin, elastase, carboxypeptidase A and B are responsible for the digestion of protein food in human stomach and intestine [5]. Thrombin and various factors (I, V, IX, X, XI, XII) are key proteases in the blood clothing cascade [6]. Serine proteases granzymes and cysteine proteases caspases are initiators of apoptosis, a programmed cell-death, which function is to remove the damaged cell, preventing it from sapping further nutrients from the organism, or halting further spread of viral infection [7, 8].

It is clear from the forgoing that the activities of the proteases need to be strictly regulated in order to prevent inappropriate proteolysis, which could be harmful or lethal. In addition to controlling endogenous proteases, an organism should be capable of controlling proteases released by various microorganisms or inflammatory cells. The proper functioning and regulation of the activity of proteases is a delicate balance of many factors, one of the most crucial being the protease inhibitors. Protease inhibitors are usually proteins with domains that enter or block a protease active site to prevent substrate access, therefore abolishing their activity [9].

Proteases, though indispensable to the maintenance and survival of human organisms, can be potentially damaging when over or underexpressed, mutated or present in higher concentrations. The mutation in

serine protease Factor IX causes haemophilia B, a blood clotting disorder caused by a deficiency of Factor IX [10]. The eficiency of cysteine protease cathepsin C activity secondary to mutations in the cathepsin C gene can cause Papillon-Lefèvre syndrome (PLS), an autosomal recessive disorder characterised by palmoplantar keratoderma and severe, early onset periodontitis [11]. Proteases are also associated with cancer progression because of their ability to degrade extracellular matrices, which facilitates invasion and metastasis [12].

The high number of proteases, their versatile roles, their degradation potential and their potential harmful action demand that protease interactions with inhibitors are understood in details.

1.1 Cysteine proteases

Cysteine proteases are the third most abundant in MEROPS database [13] with approximately 27 000 sequences. According to MEROPS, there are 72 different families of cysteine proteases, the most important being C1 (papain-like proteases), C2 (calpains) and C14 (caspases) [14].

The proteolytic activity of all cysteine proteases arises from the presence of the catalytic pair of Cys and His residues in the enzyme reactive site. The crucial step of the catalytic process involves the formation of a reactive thiolate/imidazolium ion pair ($Cys\text{-}S^{-}/His\text{-}Im^{+}$).

The proteases of this group are most commonly exemplified by papain, a well-described plant enzyme isolated from the latex of Carica papaya fruit. In the case of papain-like cysteine proteases, the catalytic centre is complemented with Asn that ensures an orientation of the His imidazole ring optimal for the successive stages of hydrolysis.

1.1.1 Papain-like cysteine proteases

The family of papain-like cysteine proteases is the most abundant among the cysteine proteases. The papain-like protease family is the only family from the clan CA and one of the few from all other families, represented in viruses, bacteria, archaea, protozoa, fungi, plants and animals [14].

The family consists of papain and related plant proteases such as chymopapain, caricain, bromelain, actinidin, ficin and aleurain, and lysosomal cysteine cathepsins. Most of these enzymes are relatively small proteins with a molecular weight in the range of 20 – 35 kDa, with the exception of cathepsin C, which is a tetrameric enzyme with a molecular weight of 200 kDa [15].

The papain family contains peptidases with a wide variety of activities, including endopeptidases with broad specificity (such as papain), endopeptidases with very narrow specificity (such as glycyl endopeptidases), aminopeptidases, a dipeptidyl-peptidase, and peptidases with both endopeptidase and exopeptidase activities (such as cathepsins B and H).

The structures of all known papain-like proteases (except cathepsin C) are monomeric, consisting of two domains (R- and L-domains according to their right and left positions in the standard view). The most prominent feature of the L-domain is the central helix about 20 residues long located between the active site and the bottom of the molecule, whereas the R-domain is folded into a β-barrel. The domains fold together in the form of a closed book. The domains form a V shaped active site cleft with the active site cysteine and histidine residue at the bottom [15, 16].

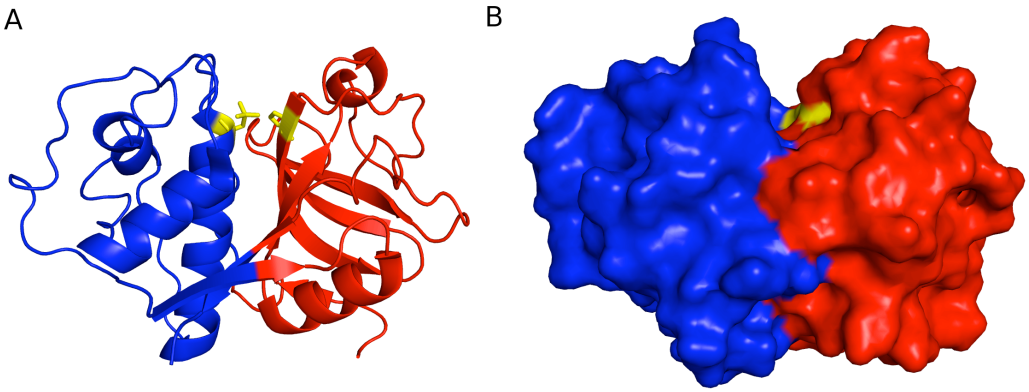


Figure 1: *Cartoon (A) and surface (B) representation of papain in standard view.* L domain is shown in blue, R domain in red and active site residues in yellow.

Active site of proteases is subdivided into subsites to which amino acid residues of the peptydic susbtrate are binding [17]. The subsites are numbered away from the catalytic site in both directions. The subsites on the C-terminal side of the bound substrate are „primed“ to differentiate them from the sites on the N-terminal side. The substrate binding sites S3, S2, and S1 of the clan CA were deduced from the structures of papain-like enzymes with substrate analogue inhibitors. The S1, S2, and S3 binding sites were deduced from the structure of papain with chloromethyl ketone inhibitor [15], whereas the S1’ and S2’ sites were deduced from the structure of cathepsin B with the epoxysuccinyl inhibitor CA030 [18].

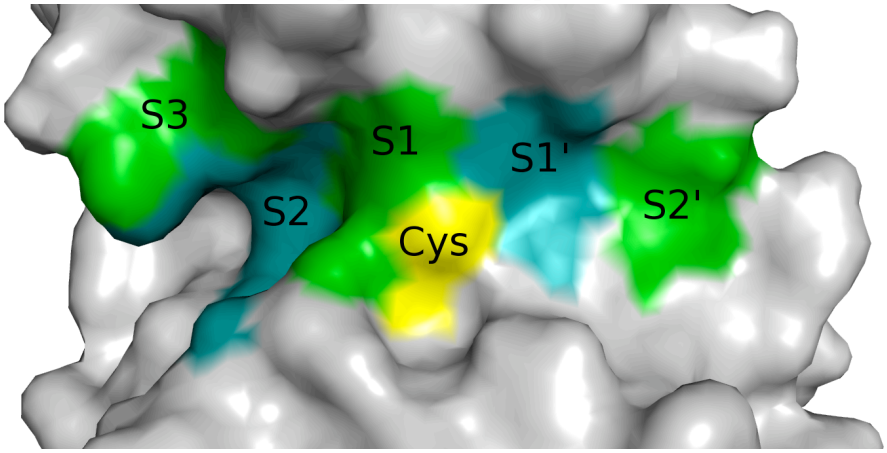


Figure 2: *Binding sites of papain-like cysteine proteases, shown in cathepsin V [19]*. The surface of cathepsin V is shown in gray, apart from the catalytic cysteine shown in yellow and the S3, S2, S1, S1’ and S2’ binding site, shown in green and cyan.

1.1.1.1 Cathepsins

Cathepsins are papain-like lysosomal cysteine proteases, optimally active in the slightly acidic, reducing environment in lysosomes.

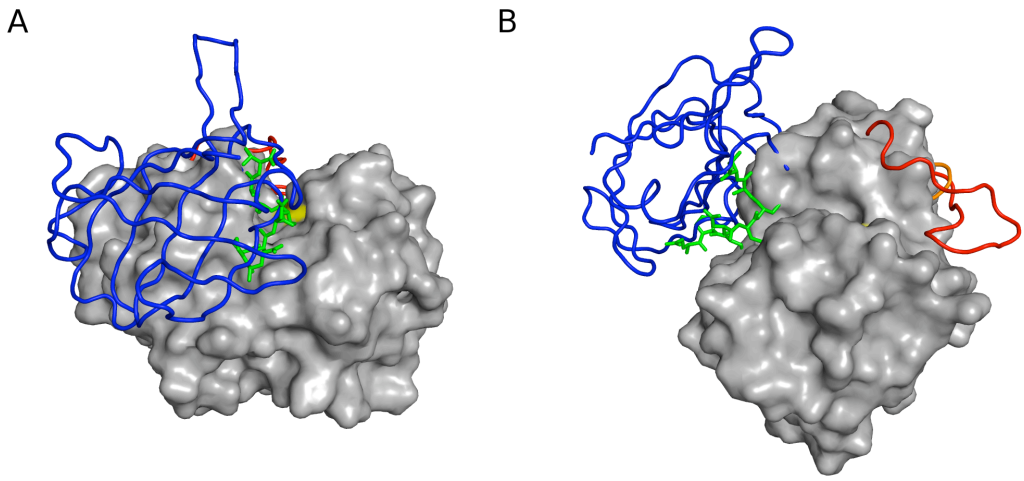


Figure 3: *Additional features of exopeptidase members of cathepsins.* A) View along active site and B) view perpendicular to the active site cleft. Papain is shown as gray surface, exclusion domain of cathepsin C as blue ribbons [23], mini loop of cathepsin X in orange [21], mini chain of cathepsin H in green sticks [24] and occluding loop as red ribbon [25].

There are 11 known members of the cathepsin group in humans. Cathepsins L, V, S, K and F are endopeptidases, whereas cathepsins B, X, H and C are exopeptidases. Cathepsins O and W still await characterization. In exopeptidases, additional structural features restrict access to the active site cleft. Cathepsin B has a short insertion, called the occluding loop, which blocks the binding sites beyond S2’ [20]. Cathepsin X has a mini-loop delimits the enzyme to a monopeptidase [21]. The part of the propeptide, covalently attached to the cathepsin H body, also called the mini chain, restricts access to nonprime binding sites [22]. Cathepsin C, the only tetrameric representative in the cathepsin group, posseses an exclusion

domain, a large insertion in the propeptide region [23].

It was long believed that the primary role of cathepsins is a nonselective protein degradation inside lysosomes. The results from a single cathepsin deficient mice showed no defects in protein degradation [26-32], suggesting that cathepsin are redundant. However, gene knockouts and genetic diseases have also revealed unknown, individual and specific functions. It was shown that cathepsin K is crucial for normal bone remodelling [33]. Cathepsin K deficient mice also developed similar symptoms as patients with pycnodysostosis [31]. The specific role of cathepsins S and L in the processing of the MHC class II-associated invariant chain, which is essential for functioning of the immune system, has also been confirmed [27, 28, 32]. Cathepsin L is also essential for epidermal homeostasis and regular hair follicle morphogenesis [30].

Cathepsin C deficient mice were unable to activate a number of serine protease (granzymes A and B, cathepsin G, neutrophil elastase and cymase) [29].

It was also shown that cathepsins play an important role in cancer. Cathepsin B has been linked to tumour progression through observations that its activity, secretion or membrane association are increased, where most malignant tumors, and specifically the cells at the invasive edge of those tumors, express the highest activity [34-36] and that the tumour progression was suppressed by addition of cathepsin B antibody or inhibitors [37]. Similar, gene expression profiling of pancreatic islet tumors in a mouse model of cancer revealed upregulation of cathepsin cysteine proteases [38].

1.2 Inhibitors of papain-like cysteine proteases

Cysteine proteases are involved in a number of important intracellular and extracellular physiological processes. It is clear from the forgoing that their activities need to be strictly regulated in order to prevent inappropriate proteolysis which could be harmful or lethal to organism. In addition to controlling endogenous proteases, an organism should be capable of controlling proteases released by various microorganisms or inflammatory cells. There are numerous ways to achieve this, one of the most important being by protease inhibitors [39]. There are several families of cysteine proteases inhibitors, the most relevant for this work are briefly presented below [39].

1.2.1 Cystatins

Cystatins are family of reversible inhibitors of papain-like cysteine proteases. They belong to I25 MEROPS family and are further divided to four groups (stefins, cysstatin, kininogens and phytocystatins) on the basis of the sequence homology [9, 14, 40, 41].

1.2.1.1 Stefins

Stefins (or type 1 cystatins) are proteins with a molecular weight of 11 kDa. They do not contain disulfide bridges and are not glycosylated. The two members found in humans are stefin A (steA) and stefin B (steB). Stefin A was found in high concentrations in epithelial cells, polymorphonuclear leukocytes, and lymphoid tissue, whereas stefin B is evenly distributed among different cells and tissues [42]. Besides protection of cytosolic and cytoskeleton proteins from the degradation by cysteine proteases accidentally released from lysosmes, several other functions have been suggested for stefins. Higher levels of stefin A in tumours have been determined in lung cancer, breast cancer, head and neck cancer and prostate cancer as well as in murine lymphosarcomas, hepatomas and Lewis lung carcinomas. These higher levels may up to a certain level, counter-balance the excessive activity of cysteine cathepsins, associated with matrix remodelling, resulting in the progression of the disease [42, 43].

1.2.1.2 Cystatins

Cystatins (or type 2 cystatins) are approximately 20-40 aminoacids longer than stefins. They are also not glycosylated, however, they have two disulfide bonds at the C-terminal part. They are mostly found in extracellular fluids [44].

The best-researched member is cystatin C, being one of the most important extracellular inhibitors of cysteine proteases. It is used as an important and effective biomarker for kidney functions [45]. Mutations in the cystatin 3 gene are responsible for the Icelandic type of hereditary cerebral amyloid angiopathy, a condition predisposing to intracerebral haemorrhage, stroke and dementia [46]. It was also discovered that certain cystatins (C, E and F) are capable of simultaneous inhibition of papain-like proteases and legumain, a cysteine protease from the C13 MEROPS family [47].

1.2.1.3 Kininogens

Kininogens (or type 2 cystatins) are multidomain proteins with a molecular weight in the range of 68 – 170 kDa. They are composed of the heavy and the light chain, connected by a disulfide bond, and the kinin domain. The heavy chain is usually composed of three cystatin-like domains, however only two of them have inhibitory activity against papain-like cysteine proteases. They are the major inhibitors of cysteine proteases in blood circulation and are also involved in inflammation, blood clotting and complement reactions [44, 48, 49].

1.2.1.4 Phytocystatins

Phytocystatins are protease inhibitors of the cystatin family, found in plants. They have been identified in rice, maize, soybean, apple fruit, carnation leaves and many other plants. Their expression is usually limited to specific organs or phases during development, such as germination, early leaf senescence, cold or salt stress [50].

1.2.2 Thyropins

Thyropins (thyroglobulin type-1 domain protease inhibitors) are inhibitors with a similar structure as the thyroglobulin type-1 domain [51]. The majority of thyropins inhibit papain-like cysteine proteases, however, a few members can inhibit aspartic protease cathepsin D [52] and proteases, for whose activity cations are needed [53].

One or more repeats of thyroglobulin type-1 domains are found in a number of functionally and structurally unrelated proteins (thyroglobulin, nidogen, testican, insulin-like growth factor-binding proteins, human pancreatic carcinoma marker protein (GA733-2), major histocompatibility complex class II-associated p41 invariant chain, saxifilin, cysteine protease inhibitor from salmon roe and equistatin) [51]. The p41 invariant chain, saxifilin, cysteine protease inhibitor from salmon roe and equistatin are reversible inhibitors of cysteine proteases [48]. These inhibtors belong to the I31 MEROPS family [14].

Crystallographic analyses of the p41 invariant chain bound to cathepsin L provided insight into the mechanism of interaction between thyropins and their target proteases [19]. The p41 fragment consists of two subdomains: the first one iscomposed of an α-helix and a ß-strand, the second subdomain being a three-stranded antiparallel ß-sheet. Similarly to cystatins, the molecule is wedge-shaped and interacts with the enzyme through three hairpin loops [19]. When compared to cystatins, the different overall structure results in additional contacts with the surface of the protease, determining the high specificity of thyropins and distinguishing them from the relatively non-selective cystatins [19, 54].

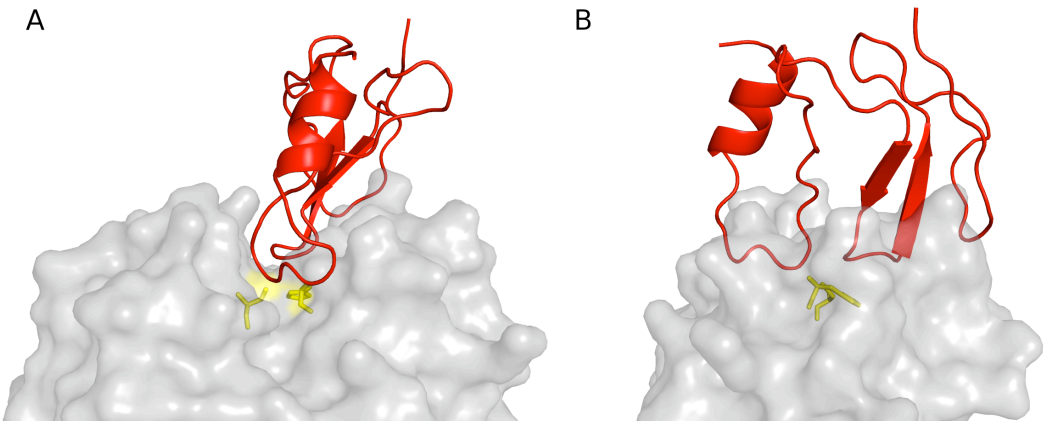


Figure 4: *Cathepsin L (gray) inhibition by p41 fragment (red).* A) View along active site and B) view perpendicular to the active site cleft. Active site residues are shown in yellow [19].

1.2.3 Chagasin

Chagasin is an inhibitor of cysteine proteases isolated from parasite *Trypanosoma cruzi*, which causes the Chagas disease in Central in Southern America. It is representative of the I42 MEROPS family [14]. Its homologues were also isolated from bacterium, archeas and other parasites (*Trypanosoma brucei*, *Leshmania* species, *Plasmodium* species), which express a variety of papain-like cysteine proteases [14, 48]. Chagasin plays an important role in the regulation of proteolytic activity of endogenous cysteine

proteases, essential for parasite differentiation and mammalian cell invasion [48, 55]. Structural studies of chagasin provide a basis for the development of synthetic drugs against different parasitic diseases (Chagas disease, malaria, trypanosomiasis, leishmaniasis) [48, 56, 57].

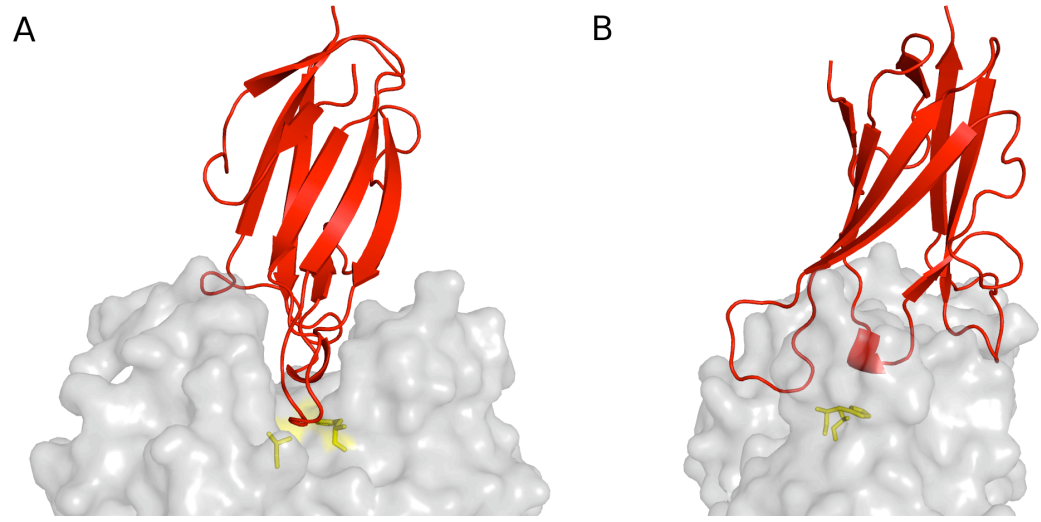


Figure 5: *Cathepsin B (gray) inhibition by chagasin (red).* A) View along active site and B) view perpendicular to the active site cleft. Active site residues are shown in yellow [58].

1.2.4 Serpins

Some representatives of serpins (serine protease inhibitors) also inhibit cysteine proteases. One of them is the squamous cell carcinoma antigen (SCCA-1). It is a member of the I4 MEROPS family [14] and inhibits trypsin, papain and cathepsins L, S and K [59] with the same inhibition mechanism [60].

1.2.5 Cysteine protease propeptides and their homologues

The majority of cysteine proteases are synthesised as inactive zymogens. Removal of the N-terminal propeptide is needed for activation of the enzyme. The propeptide can be removed autocatalytically or by other proteases. It is bound to the active site, but in the reversed orientation of a susbtrate. Its reversed orientation prevents its cleavage, while its presence prevents the approach of substrates. When enzymes are activated and the propeptide is removed, propeptide can still act as a reversible inhibitor of the active enzyme [48, 59].

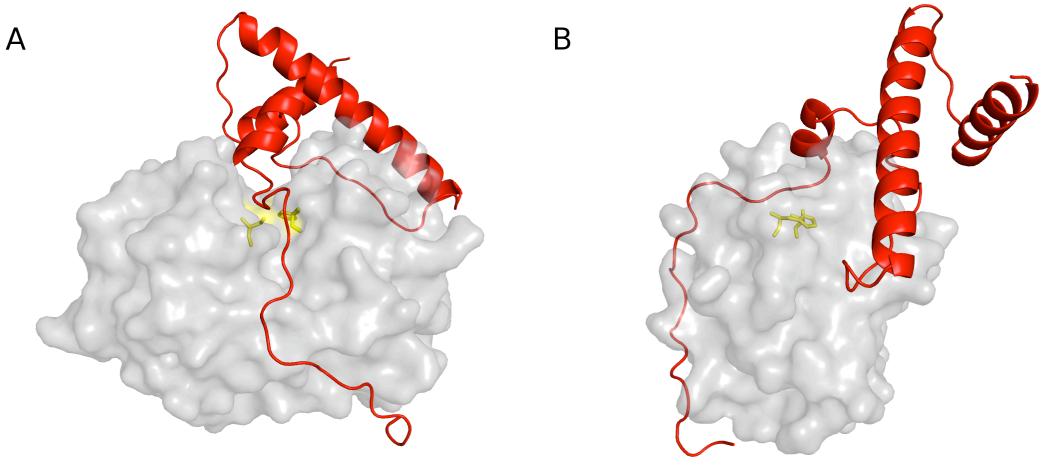


Figure 6: *Blocking of active site by propeptide in procathepsin L.* A) View along active site and B) view perpendicular to the active site cleft. Propeptide is shown in red, mature enzyme in gray and active site residues in yellow [64].

Propeptides are expressed simultaneously with the proteases, however, there are reports about propeptide homologues which are expressed independently from the proteases. The first described was the

cytotoxic T-lymphocyte antigen-2β in mice, which shows high similarity to cathepsin L propeptide and competitively inhibits cathepsins L, H and papain [48, 61, 62]. Similar homologues were also discovered in silkworm *Bombyx mori*, which inhibits cathepsins L and S, but not papain and cathepsin B. These inhibitors are classified in the I29 MEROPS family [14, 62, 63].

1.2.6 Staphostatins

Recently, a new mechanistic class of cysteine proteases inhibitors has been identified and classified in the I58 MEROPS family [14]. They were named staphostatins due to their high specificity towards staphopains, bacterial papain-like cysteine proteases [65]. Staphostatins were found to be β-barrels formed by a three-stranded mixed β-sheet and a five-stranded anti-parallel β-sheet [66].

The crystal structure of staphostatin B in complex with staphopain B revealed that their most crucial interactions occur in the region defined as the inhibitor-binding loop. The loop spans the protease active site cleft in a direction analogous to that of substrates [67].

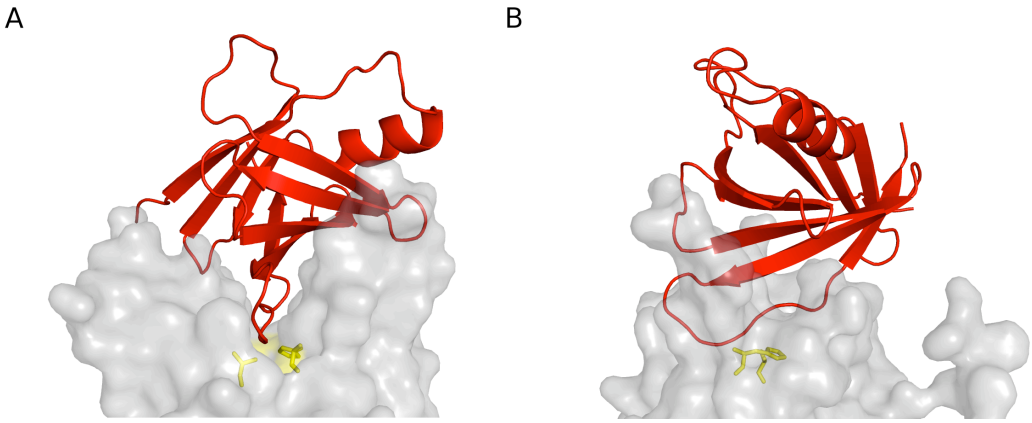


Figure 7: *Staphopain (gray) inhibition by staphostatin (red).* A) View along active site and B) view perpendicular to the active site cleft. Active site residues are shown in yellow. [67]

1.2.7 Kunitz-type inhibitors

Kuntz-type inhibitors are members of the I3 MEROPS family [14]. They are primary a serine protease inhibitors, however, a few members inhibit also proteases of other catalytic types. There are reports about inhibitors of cysteine proteases from potato [68], protease inhibitor from *Proposis juliflora* [69] and *Bauhinia bauhinioides* cruzipain inhibitor [70], however, their inhibition mechanism remains unknown.

Kunitz-type inhibitors are single-chain proteins with a molecular weight from 16 to 20 kDa. They inhibit serine proteases in canonical mode; that is binding in substrate-like manner along the active site with high affinity. At typically used concentrations and neutral pH the hydrolysis of the inhibitor is extremely slow. The system behaves as if there was a simple equilibrium between the enzyme and the free inhibitor on the one hand and the complex on the other [71, 72].

The crystal structure of the representative member soybean trypsin inhibitor (STI) in complex with serine protease trypsin [73, 74] revealed that Kunitz type inhibitors have an unusual fold which was named β-trefoil fold. The same fold is found in the inhibitor isolated from *Erythrina caffra* [75], interleukins-1α and 1β [76] and fibroblast growth factors [77].

The STI structure consists of twelve antiparallel β-strands, long loops connecting these strands and a short helical insertion. Six of the strands (β1, β4, β5, β8, β9 and β12) form a short antiparallel β-barrel, with one side of the barrel being closed by a lid consisting of the other six strands.

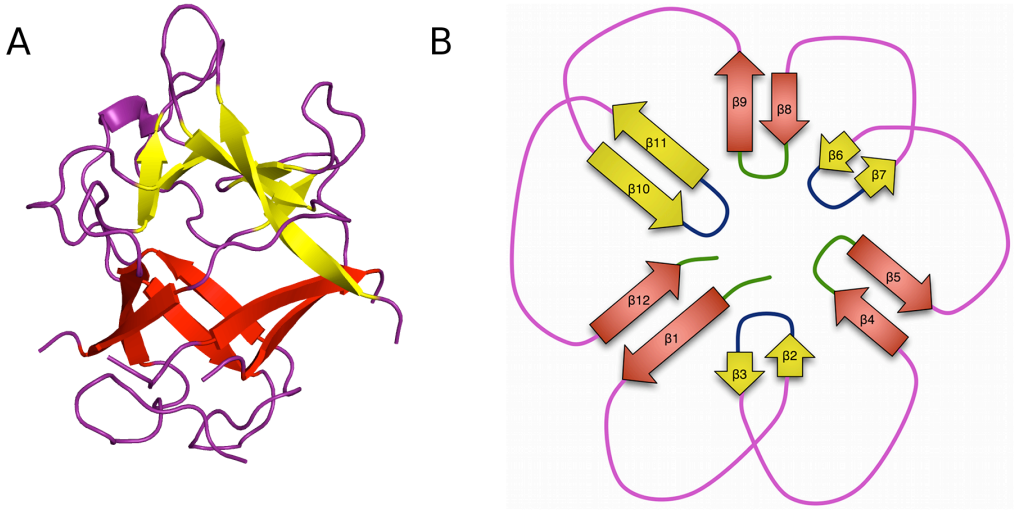


Figure 8: *A) Structure of soybean trypsin inhibitor and B) schematic representation of β-trefoil fold.* The barrel-forming strands are shown in red and lid forming strand in yellow. [74]

1.2.7.1 Inhibitors of cysteine proteases from potato

Inhibitors of cysteine proteases from potato are single-chain proteins with a molecular weight of 22 – 25 kDa, isolated from potato tubers. They inhibit cysteine proteases papain (K_i=50-100 nM), cathepsin B (K_i=90-100 nM), cathepsin H (K_i =400 nM) and cathepsin L (K_i =0.06 – 1.5 nM) [68]. They were classified to the Kunitz-type family on the basis of the sequence similarity with other members [78].

1.2.7.2 Protease inhibitor from *Proposis juliflora*

Protease inhibitor from *Proposis juliflora* was classified to the Kunitz-type family on the basis of the sequence similarity. It inhibits trypsin, papain and digestive cysteine proteases from several phytophagous pests [69]. It was suggested that the same loop is responsible for the inhibition of serine and cysteine proteases [79].

1.2.7.3 *Bauhinia bauhinioides* cruzipain inhibitor

Bauhinia bauhinioides cruzipain inhibitor BbCI, a Kunitz-type inhibitor belonging to the I3 MEROPS family, inhibits papain-like endopeptidases cruzipain and cathepsin L, serine protease trypsin, but not endopeptidase cathepsin V, and exopeptidases cathepsins B and X. It is also the first Kunitz-type cysteine protease inhibitor for which the crystal structure was determined, however, the structure of BbCI protease complexes is lacking, though it was suggested that the same reactive loop is involved in the inhibition of cysteine and serine proteases [70, 80].

1.2.8 Mycocypins

Mycocypins are a recently discovered group of cysteine protease inhibitors from basidiomycetes. Their structure and inhibition mechanism has been unknown until this work.

1.2.8.1 Clitocypin

Clitocypin (***Clito****cybe nebularis* **cy**steine **p**rotease **in**hibitor) is a strong and specific inhibitor of cysteine proteases, isolated from basidiomycete *Clitocybe nebularis* [81]. Its exact role in the basidiomycete is still unknown.

The lack of the sequence homology to other families of protease inhibitors made it the first member of the I48 MEROPS family [14, 81]. It is extremely resistant to heat, extreme pH values and proteolytic degradation. Secondary structure studies have shown that it posseses a high content of β-structures [82, 83].

Recombinant clitocypin inhibits papain (K_i=6.2 nM), cathepsins L (K_i=0.02 nM) and K (K_i=0.03 nM), legumain (K_i=21.5 nM), but not trypsin, pepsin, cathepsin H and caspases 3, 6 and 7. Contrary to cystatins and thyropins natural clitocypin also inhibits bromelain (K_i=160 nM) [81, 84].

1.2.8.2 Macrocypin

Macrocypins (***Macro****lepiota procera* **cy**steine **p**rotease **in**hibitor) are recently discovered cysteine protease inhibitors from basidiomycete *Macrolepiota procera [85]*. Members of macrocypin family are encoded by a family of genes that are divided into five groups with a more than 90% within-group sequence identity and 75–86% between-group sequence identity.

Macrocypins exhibit similar basic biochemical characteristics, stability against high temperature and extremes of pH, and inhibitory profiles similar to those of clitocypin. This suggests that they belong to the same MEROPS family of cysteine protease inhibitors, named mycocypins.

Natural and recombinant macrocypins 1, 3 and 4 inhibit papain (K_i=0.12 – 5.04 nM), cathepsins L (K_i=0.31 – 3.81 nM), V (K_i=0.45 – 12.6 nM), S (K_i=5.1 – 47.1 nM), K (K_i=4.5 – 170 nM), B (K_i > 125 nM) and H (K_i=24 – 370 nM). Natural and recombinant macrocypins 1 and 3 also inhibit legumain (K_i=3.4 – 110 nM). Macrocypin 4 does not inhibit AEP, but in contrast to others it inhibits serine protease trypsin (K_i=160 nM)[85].

1.3 Inhibition of cathepsins with stefin A

Stefins are tight and reversible inhibitors, capable of strong inhibition of cathepsins. They inhibit endopeptidases in the picomolar range (K_i= 0.05 – 130 pM) and exopeptidases in in the nanomolar range (K_i= 0.3 – 100 nM) [24, 86].

Although the crystal structures of papain had been known for a long time [87], the structural basis for the inhibition was not known until the structural characterization of chicken cystatins C [88].

On the basis of the chicken cystatin C and papain structures it has been suggested that the N-terminal segment and two cystatin loops bind into the active site cleft.

This hypothesis was later confirmed by the crystal structure of stefin B in complex with carboxymethylated papain [89], however, this structure has not explained the interactions between cathepsins and stefins in details.

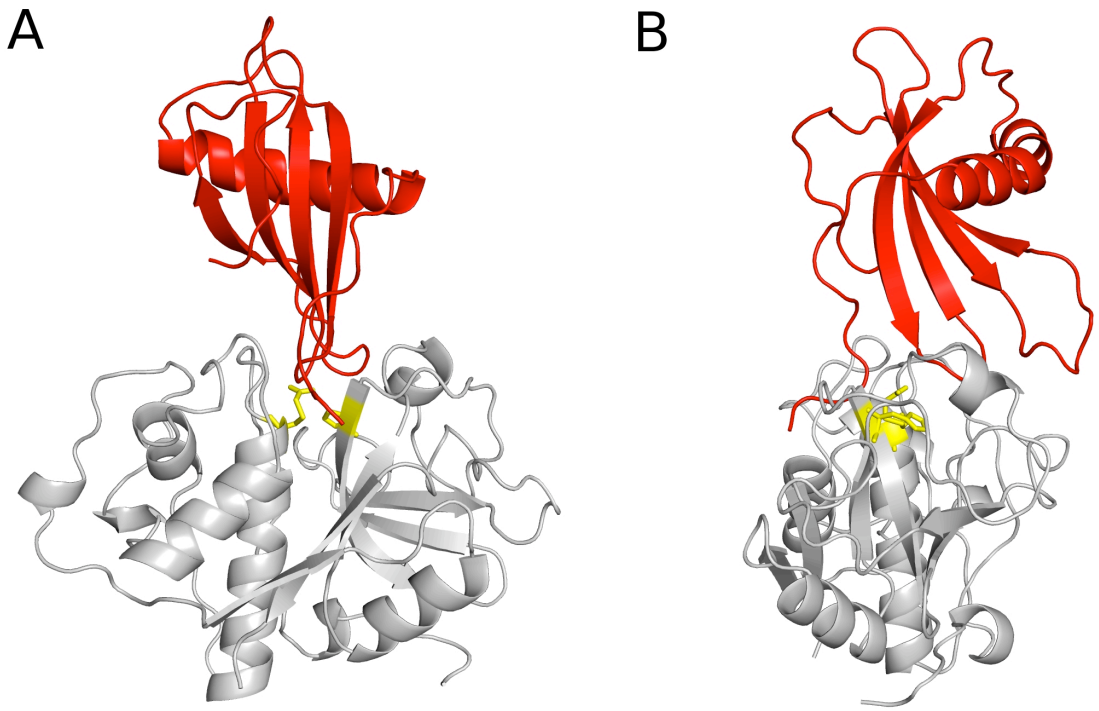


Figure 9: *A crystal structure of carboxymethylated papain with stefin B.* Papain is shown in gray, stefin B in red and active site cysteine in histidine in yellow. Stefin B fills the active site cleft in wedge like shape (A). N-terminal segment binds to the nonprimed binding sites, whereas both loops bind to primed binding sites (B).

A step further was the determination of the crystal structure of cathepsin H – stefin A complex [24], which showed that the N-terminal trunk partially displaces the mini-chain. It was also demonstrated that stefin A binds deeper into the active site cleft in the absence of carboxymethylated cysteine residue.

Extensive work was done on the characterization of important residues in stefin A and B, involved in inhibition of cathepsins and papain. It was shown that sequential deletion of N-terminal residues in stefin A reduces the affinity towards cathepsins and papain [90], whereas deletion of N-terminal methionine in stefin A did not have any effect on the inhibition of papain and cathepsins L and B. Deletion of Met1-Ile2 resulted in 900x, 3x and 200x fold reduction in affinities for papain, cathepsins L and B, respectively. Further deletion of Pro3 resulted in additional decreased affinities (2000x, 20 000x, 400x fold). The reductions in affinity shown by the latter mutant indicate that the N-terminal region contributes about 40% of the total free energy of the binding of stefin A to cysteine proteases [90].

The important role of Gly4 was confirmed by the point mutation studies in the native full-length stefin A, where even the smallest replacement, by Ala, resulted in 1000-, 10- and 6000-fold decreased affinities for papain, cathepsins L and B, respectively. Further mutation in larger aminoacids additionally reduced the affinities. The highest reduction (>105x) was observed, when Gly4 was replaced by Arg4 or Glu4 [91].

The mutation analysis of the second loop in stefin A showed that the second binding loop of stefin A plays a minor role in stabilizing the complexes with proteases by retarding their dissociation. Mutation of Leu73 decreased the affinity for papain, cathepsins L and B by ≈ 300-fold, >10-fold and ≈ 4000-fold, respectively. Mutation of Pro74 decreased the affinity for cathepsin B by ≈ 10-fold, but minimally affected the affinity for the other two enzymes. Mutation of Gln76 and Asn77 did not alter the affinity of stefin A towards any of the proteases studied. The decreased affinities were caused exclusively by the increased dissociation rate constants [92].

2 Purpose of the work

Interactions between macromolecules are the basic events of life. Their characterization is thus a key to understanding the physiology with the underlying specificity of these events at the molecular level. 3-dimensional structures determined by X-ray crystallography, electron microscopy and NMR spectroscopy reveal how molecules interact with each other at atomic detail. But not only the complexed states, also the process of approach and underlying principles are essential for understanding physiology of a living cell. Macromolecules are namely relatively large particles, therefore it is conceivable to assume that their approach is not completely random, but is a consequence of long term interactions, which orient and guide approach of the interacting partners. To confirm such hypothesis one should understand the energetics of approach of two interacting partners, when possible in intermolecular distance dependent manner. Most methods provide insight into equilibrated (steady) states of such processes, whereas the course of events - approach and separation - remain largely unexplored.

Besides revealing novel interaction mechanisms, we have made an attempt to gain insight also into the distance dependence of interactions by exploiting the tight interaction of proteases and their inhibitors. Their interactions have been studied by a combined use of X-ray crystallography, protein kinetics (enzyme kinetics and surface plasmon resonance) and atom force microscopy.

X-ray crystallography and atom force microscopy have been used to gain insight into the geometry and distance dependence of binding, whereas enzyme kinetics, sufrace plasmon resonance and have been used to gain insight into the energetics of interaction.

3 Materials and Methods

3.1 Plasmid and mutants preparation

The pPIC9 plasmid containing cDNA sequence of human procathepsin L and V were obtained from dr. Marko Mihelič. The PET3a plasmid containing cDNA sequence of procathepsin B was obtained from dr. Dejan Caglič. The PET3a plasmid containing cDNA sequence of stefin A was obtained from dr. Eva Žerovnik. The PET3a and PET11a plasmid containing cDNA sequence of clitocypin and macrocypin isoforms were obtained from dr. Jerica Sabotič.

The mutants of cathepsin L, V and stefin A were produced by site-directed mutagenesis with two mutagenic primers and two flanking primers [93].

The mutants of clitocypin and macrocypin isoforms were produced by PCR site-directed mutagenesis using appropriate pET vectors as templates, followed by digestion of templates with *DpnI* (Fermentas) and recovery of the pET vectors containing mutated inserts [94].

The cloning and expression of clitocypin and macrocypin mutants was performed in collaboration with dr. Jerica Sabotič from Department of Biotechnology, Jožef Stefan Institute.

3.2 Protein expression and isolation

Proteins were recognized as a distinct class of biological molecules in the eighteenth century. Since that time, proteins were isolated from natural sources until a few decades ago recombinant DNA technology started to evolve. Recombinant DNA technology, followed by the increasing number of known protein sequences, allowed the production of proteins which could not be isolated from natural sources. Basically, the DNA sequence of the protein is inserted into the appropriate vectors and then transferred into the expression cells, where the protein is expressed.

3.2.1 Theoretical backgrounds

3.2.1.1 *Pichia pastoris* expression

Pichia pastoris is frequently used as an expression system for the production of proteins. A number of properties make *Pichia* suited for this task: *Pichia* has a high growth rate and is able to grow on a simple, inexpensive medium. *Pichia* can grow in either shake flasks or a fermenter, and this makes it suitable for both small and large-scale production.

Pichia pastoris has two alcohol oxidase genes, AOX1 and AOX2, which have a strongly inducible promoter. These genes allow Pichia to use methanol as a carbon and energy source. The AOX promoters are induced by methanol and are repressed by e.g. glucose. Usually the gene for the desired protein is introduced under the control of the AOX1 promoter, which means that protein production can be induced by addition of methanol [95].

In a used expression vector pPIC9, the desired protein is produced as a fusion product to the secretion signal of the α-mating factor from *Saccharomyces cerevisiae* (baker's yeast). This causes the protein to be secreted into the growth medium which greatly facilitates subsequent protein purification [96-98].

3.2.1.2 Escherichia coli expression

The pET System is the most powerful system ever developed for the cloning and expression of recombinant proteins in *E. coli*. Target genes are cloned in pET plasmids under the control of strong bacteriophage T7 transcription; expression is induced by providing a source of T7 RNA polymerase in the host cell. T7 RNA polymerase is so selective and active that, when fully induced, almost all of the cell's resources are converted to target gene expression; the desired product can comprise more than 50% of the

total cell protein a few hours after induction. Although this system is extremely powerful, it is also possible to attenuate the expression level simply by lowering the concentration of inducer. Decreasing the expression level may enhance the soluble yield of some target proteins. Another important benefit of this system is its ability to maintain target genes transcriptionally silent in the uninduced state [99-101].

3.2.2 Protocols

3.2.2.1 Cathepsin L and V

The pPIC9 plasmid containing the cDNA sequence of human procathepsin L and V and their mutants were linearized with *SacI* (NEB) and electroporated in *Pichia pastroris* GS115 (Invitrogen). The recombinant yeast cells were inoculated in the BMGY medium (1 % yeast extract, 2 % peptone, 0.1M potassium phosphate, pH 6.0, 1.34 % yeast nitrogen base, 4×10-5 % biotin and 1 % glycerol). When the culture reached OD600=2.0, the medium was changed to the BMMY medium (1 % yeast extract, 2 % peptone, 0.1M potassium phosphate, pH 6.0, 1.34 % yeast nitrogen base, 4×10-5 % biotin and 1 % methanol). After 4 days the supernatant was collected, concentrated and applied to Ni-NTA (Invitrogen) affinity chromatography. The column was washed with 20 mM phosphate, 500 mM NaCl, 10 mM imidasole, pH 7.0 and proteins eluted in 20 mM phosphate, 500 mM NaCl, 500 mM imidasole, pH 7.0. Eluates containing procathepsins L and V were pooled and dialysed against 10 mM phosphate buffer, pH 7.5. The dialysed protein samples were applied to Q-sepharose equilibrated with the 10 mM phosphate buffer, pH 7.5. Proteins were eluted with 10 mM phosphate buffer, pH 7.5. The fractions containing cathepsins L and V were pooled, pH was lowered to 4.5 and 10 mM DTT was added. After the activation at 37°C had finished, the mature cathepsins were applied to SP-sepharose, equilibrated with 10 mM acetate buffer, pH 5.5. Cathepsins L and V, eluted with 10 mM acetate buffer, 500 mM NaCl, pH 5.5 were pooled, dialysed against 10 mM acetate buffer, pH 5.5 and concentrated to 1.0 mg/mL. The samples were treated with methyl methanethiosulfonate (MMTS) and frozen at -80°C until use [102].

3.2.2.2 Cathepsin B

PET3a vector containing the cDNA sequence of stefin A was transformed in *Escherichia coli* BL21(DE3). The expression was induced with 0.4 mM IPTG when OD600 reached 0.6. After 4 hours of expression, cells were centrifuged and frozen. The cells were lysed. Inclusion bodies were washed three times with 50 mM TrisHCl, 2 mM EDTA, 0.1% Triton-x-100, pH 8.0 and 50 mM TrisHCl, 2 M urea, 2 mM EDTA, 0.1% Triton-x-100, pH 8.0. The insoluble protein was dissolved in 6 M guadinine hydrochloride, 0.3 M Na2SO3, pH=8.0 and 1 mL of 50 mM Thannhauser reagent was added [103]. Proteins were than precipitated in cold water by addition of acetic acid. Precipitation after centrifugation was dissolved in 8 M urea and diluted to 0.1 mg/mL. The denatured protein was refolded by slow dialysis against 100 mM Tris-HCl, 5 mM cysteine, 5 mM EDTA. Refolded procathepsin B was activated with the addition of 5 mM DTT. After the activation had been finished, the mature cathepsin B was applied to SP-sepharose, equilibrated with the 10 mM acetate buffer, pH 5.5. Cathepsin B, eluted with 10 mM acetate buffer, 500 mM NaCl, pH 5.5 was pooled, dialysed against 10 mM acetate buffer, pH 5.5 and concentrated to 1.0 mg/mL. Samples were treated with the methyl methanethiosulfonate (MMTS) and frozen at -80°C until use [104].

3.2.2.3 Stefin A

PET3a vector containing the cDNA sequence of stefin A was transformed in *Escherichia coli* BL21(DE3) cells [105, 106]. The expression was induced with 0.4 mM IPTG when OD600 reached 0.6. After 4 hours of expression, cells were centrifuged and frozen. The soluble fraction after lysis was concentrated and applied to a size-exclusion column (Superdex 75), equilibrated with the 10 mM phosphate buffer, 200 mM NaCl, pH 6.5. Eluates containing stefin A were pooled and dialysed against 10 mM phosphate buffer, pH 7.5. The dialysed sample was applied to Q-sepharose, equilibrated with the 10 mM phosphate buffer, pH 7.5. Stefin A was eluted with 10 mM phosphate buffer, 500 mM NaCl, pH 7.5, dialysed against 10 mM acetate buffer, pH 5.5, concentrated to 1 mg/mL and frozen at -80°C until use.

3.2.2.4 Clitocypin, macrocypin and their mutants

Macrocypin, clitocypin, methionine-containing clitocypin mutant (Clt-L82M, Clt-I89M), and their mutants (Mcp-I72K+N74K, Mcp4-A70I+K72N, Mcp-DG25, Mcp- G25A, Mcp4-N74R and Clt-DG24, Clt-G24A, Clt-N70K) were expressed in *Escherichia coli* [84, 107]. The expression was induced with 0.4 mM IPTG

when OD600 reached 0.6. After 4-6 hours of expression, cells were centrifuged and frozen. The cells were lysed and inclusion bodies were washed three times with 50 mM TrisHCl, 2 mM EDTA, 0.1% Triton-x-100, pH 8.0. The inclusions bodies were repeatedly washed with increasing concentrations of urea. The proteins typically dissolved in 1-4 M urea. Fractions with the dissolved proteins were dialysed against 10 mM phosphate, 2 M urea, pH=6.5 and applied to Q-sepharose, equilibrated with the same buffer. Proteins were eluted with 10 mM phosphate, 2 M urea, 500 mM NaCl, pH=6.5. Fractions containing macrocypin, clitocypin and their mutants were pooled and dialysed against 10 mM acetate buffer, pH 5.5. Proteins were concentrated to 1 mg/mL and frozen at -80°C until use.

A selenomethionine mutant was produced using minimum autoinduction media with the addition of selenomethionine in *E. coli* BL384 cells [108]. Isolation was performed in the same way as described above, however 5 mM DTT was added in all buffers used.

3.3 Crystallization

3.3.1 Theoretical background

Crystallization is the process of formation of solid crystals which are needed for structure determination. The crystallization process consists of two major events, nucleation and crystal growth [109].

Nucleation is the step where the protein molecules dispersed in the buffer start to gather into clusters that stabilize under the current operating conditions. These stable clusters constitute the nuclei. However when the clusters are not stable, they redissolve. Therefore, the clusters need to reach a critical size in order to become stable nuclei. Such critical size is dictated by the operating conditions (temperature, supersaturation, etc.). It is at the stage of nucleation that the molecules arrange in a defined and periodic manner which defines the crystal structure.

The crystal growth is the subsequent growth of the nuclei that succeed in achieving the critical cluster size. Nucleation and growth continue to occur simultaneously while supersaturation exists. Supersaturation is the driving force of the crystallization; hence the rate of nucleation and growth is driven by the existing supersaturation in the solution. Depending upon the conditions, either nucleation or growth may be predominant over the other, and as a result, crystals with different sizes and shapes are obtained. Once the supersaturation is exhausted, the solid-liquid system reaches equilibrium and the crystallization is complete, unless the operating conditions are modified from equilibrium so as to supersaturate the solution again [110]. Figure 10 represents a simplified phase diagram, where only protein and precipitant concentration are shown. In reality, many other factors (pH, ionic strength, temperature, etc) influence the crystallization process.

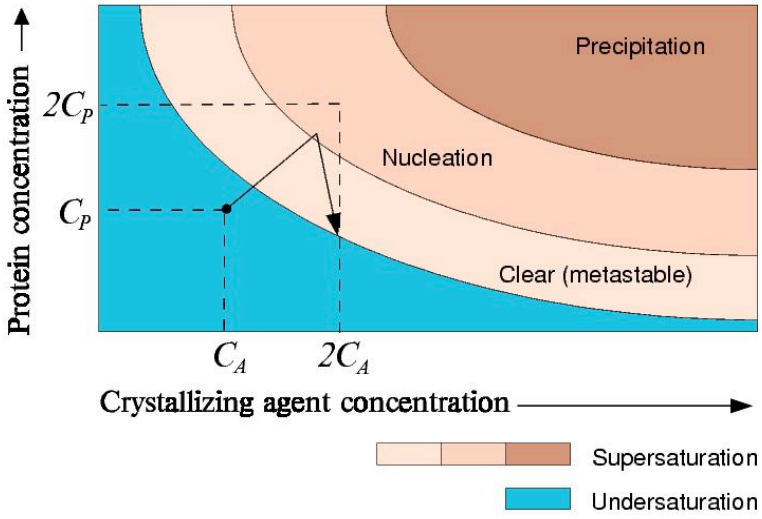


Figure 10: *Simple protein crystallization phase diagram*. Nucleation occurs in nucleation zone, but not in metastable or growing zone. Crystals are growing until protein concentration becomes undersaturated. Black arrow represents successful crystallization trial.

Two of the most commonly used methods for protein crystallization fall under the category of vapour diffusion. These are known as the hanging and sitting drop methods [111, 112]. Both entail a droplet containing purified protein, buffer, and precipitant being allowed to equilibrate with a larger reservoir

containing similar buffers and precipitants at higher concentrations. A drop composed of a mixture of sample and precipitant is placed in the vapour equilibration with the liquid reservoir of reagent. Typically the drop contains a lower reagent concentration than the reservoir. To achieve equilibrium, water vapours leave the drop and eventually end up in the reservoir. As water leaves the drop, the sample undergoes an increase in relative supersaturation. Both the sample and reagent increase in concentration as water leaves the drop for the reservoir. Equilibration is reached when the reagent concentration in the drop is approximately the same as that in the reservoir [113].

If nucleation zone is reached and stable nuclei are formed, crystals start growing.

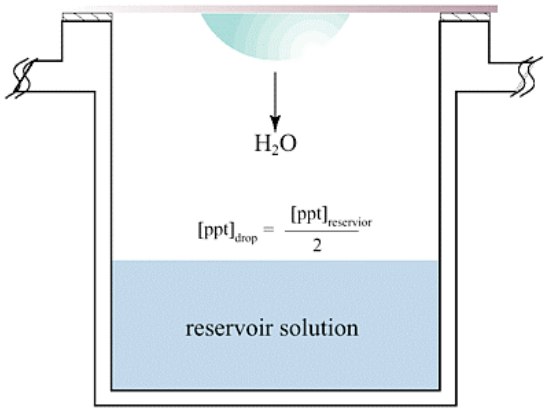


Figure 11: *Overview of vapour diffusion crystallization.* A drop composed of a mixture of protein and reservoir solution is placed in vapour equilibration with a liquid reservoir of reagent.

3.3.2 Protocols

All initial small volume crystallization trials were performed using commercial screens from Qiagen by sitting drop method in 96 well plates. Drops composed from 0.1 μL of protein and 0.1 μL of reservoir solution were equilibrated against 0.05 - 0.1 mL of reservoir solution. The optimization of initial hits was performed by sitting drop method with mixing 1 μL of protein and 1 μL of reservoir solution in 24 well plates and equilibrated against 1 mL of reservoir solution, unless otherwise stated.

3.3.2.1 Complexes of stefin A with cathepsins L, V and B

Unreduced cathepsin V and stefin A were mixed in a molar ratio of 1:1.1 and concentrated to 50 mg/mL in 10mM acetate, pH=5.5. Crystals were grown in 0.1 M Tris-HCl, pH=8.0, 12% PEG3000.

For the complex with reduced cathepsin V, cathepsin V was mixed with 5 mM DTT and incubated on 25°C for 1 hour. After stefin A was added in a molar ratio of 1:1.1, the complex was concentrated to 50 mg/mL in 10mM acetate, pH=5.5. Crystals were grown in 0.1 M Tris-HCl, pH=8.25, 16% PEG3000.

Unreduced cathepsin L and stefin A were mixed in a molar ratio of 1:1.1 and complex was purified on Superdex S75 column. Complex was concentrated to 40 mg/mL in 10mM acetate, pH=5.5. Crystals were grown in 0.1 M Tris-HCl, pH=7.0, 16% PEG3000. The crystals were frozen in liquid nitrogen prior to the diffraction data collection.

Cathepsin B and stefin A were mixed in a molar ratio of 1:1.1 and concentrated to 30 mg/mL in 10mM acetate, pH=5.5. Crystals were grown in 0.2 M sodium sulphate, 24% PEG3000. The initial crystals grown by the sitting drop method were highly mosaic, therefore useless for structure determination. However, by using the hanging drop method and controlled evaporation crystallization approach [114] we improved the crystal quality.

The crystals were soaked in mother liquor supplemented with 20-30% glycerol and frozen in liquid nitrogen prior to the diffraction data collection.

3.3.2.2 Macrocypin

Macrocypin 1 was concentrated to 30 mg/mL in 10 mM acetate buffer, 200 mM NaCl, pH 5.5. Crystals grew overnight in various conditions (BisTrisPropane buffer, pH 6.5-7.5, 100-500 mM different sodium salts, 20% PEG 3350 or 0.8-1.6M Na/K phosphate, pH 7.0-8.0). The best diffracting crystals were grown in BisTrisPropane buffer, pH 7.0, 200 mM NaI, 20% PEG3350 and BisTrisPropane buffer, pH 7.0, 200 mM

sodium citrate, 20% PEG3350.

3.3.2.3 Cathepsin V – clitocypin complex

The initial screening of clitocypin in a complex with protease was performed with cathepsins L, V, and papain. The proteins were mixed in a molar ratio of 1:1.1 and concentrated to 50 mg/mL in 10 mM acetate, 100 mM NaCl, pH 5.5. Of the complexes, only cathepsin V and the methionine mutant produced diffracting crystals. Crystals of dimensions of 0.2×0.4×0.1 mm3 were obtained in 0.4 M Li2SO4, 12% PEG800, 20% glycerol after 4 months. The selenomethionine mutant of clitocypin in complex with cathepsin V gave better diffracting crystals in a much shorter time. The crystals were directly frozen in liquid nitrogen prior to the diffraction data collection.

3.3.2.4 Clitocypin

Our group has recently reported the crystallization conditions and phasing attempts using clitocypin purified from natural sources [115]. 1 mL drops of 15 mg/ml solution of clitocypin in 15 mM MES buffer, pH 6.0, gave crystals when mixed with 1 mL of crystallization buffer (50 mM monopotassium dihydrogen phosphate, 20% (w/v) polyethylene glycol 8000, pH 3.76) using the vapour diffusion method. A number of data sets with a variety of resolutions were collected. The structure could not be solved due to the high heterogeneity of the natural clitocypin and unsuccessful derivatisation of the crystals, while the absence of significant sequence similarity [116] with other proteins with known structures has discouraged molecular replacement attempts.

In contrast to the natural clitocypin, the recombinant clitocypin, methionine and selenomethionine containing mutants alone gave no crystals.

3.4 Structure determination

3.4.1 Theoretical bacground

3.4.1.1 X-ray diffraction

Diffraction from a single molecule is too weak to be measured, therefore crystals, an ordered three-dimensional array of molecules, are used instead to intensify the signal. Once the crystals are obtained, they are exposed to X-rays. A crystal behaves like a three-dimensional diffraction grating which gives rise to both constructive and destructive interference effects in the diffraction pattern.

An intuitive understanding of X-ray diffraction can be obtained from the Bragg model of diffraction. In this model, a given reflection is associated with a set of evenly spaced planes running through the crystal. The orientation of a particular set of planes is identified by its three Miller indices (h, k, l), and their spacing by d [117].

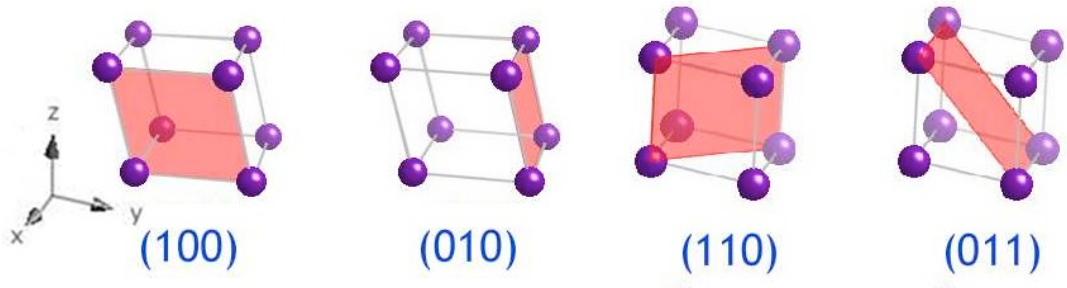
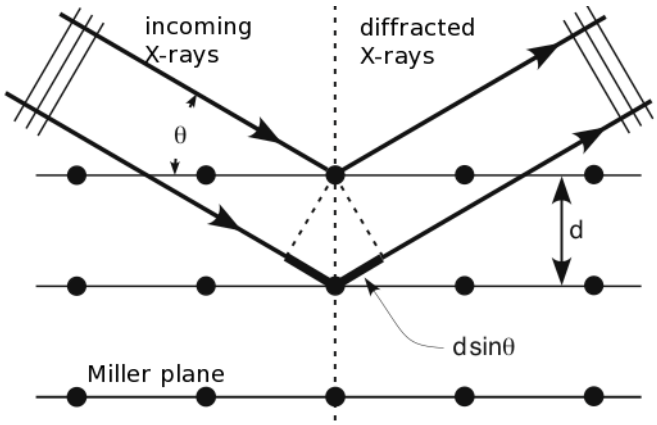


Figure 12: *Representation of Miller indices.*

William Lawrence Bragg proposed a model in which the incoming X-rays are scattered mirror-like from each plane; from that assumption, X-rays scattered from adjacent planes will combine constructively (constructive interference) when the angle θ between the plane and the X-ray results in a path-length difference that is an integer multiple n of the X-ray wavelength λ.



$$2d \times \sin\theta = n \times \lambda$$

Figure 13: *Explanation of X-ray diffraction and Bragg's law*

A diffraction image of a three-dimensional crystal appears as a pattern of series of discrete spots, known as reflections. Each reflection contains information about all atoms in the crystal structure. Conversely, each atom contributes to the intensity of each reflection.

From such an image, the crystal symmetry, the unit cell parameters (dimensions, angles), the crystal orientation and the resolution limits are determined. The data collection strategy should be designed in a way to maximize both the resolution and the completeness of the data set. After choosing an appropriate strategy, the whole data set is collected: the crystal is rotated through a small angle, typically by 1 degree, and images (X-ray diffraction patterns) are recorded. The total rotation angle depends on the crystal symmetry. The lower the symmetry, more data (larger total angle) are required [117].

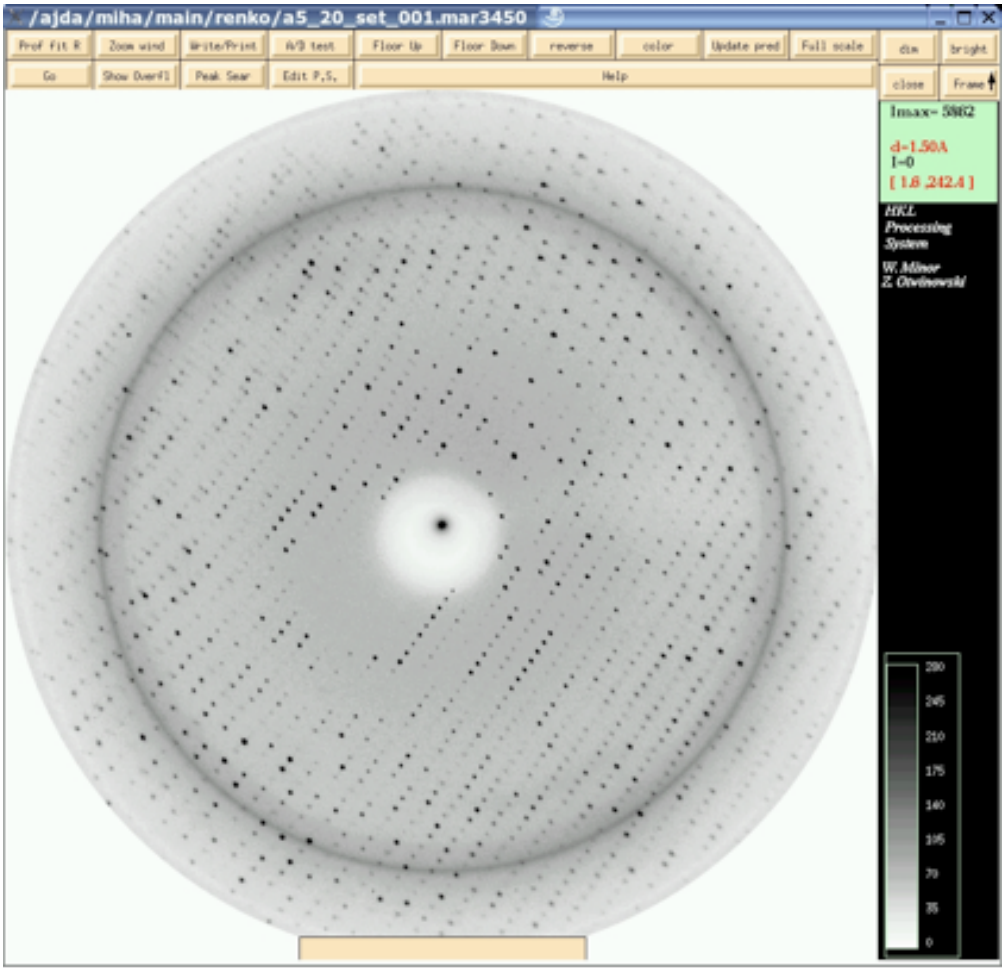


Figure 14: *A diffraction pattern.* The pattern after exposing a protein crystal to X-rays rotating the crystal by 1º viewed in HKL2000 [118].

3.4.1.2 Phase determination

Characteristic for X-rays, being an electromagnetic waving, are the amplitude and the phase. In order to recombine a diffraction pattern, both parameters are required for each reflection. During data acquisition, only the amplitudes of reflections can be recorded. Unfortunately, for the resulting electron density map the missing phase angles are more important than the amplitudes. In order to obtain an interpretable electron density map, both, the amplitude and the phase must be known. There are many different approaches to solve the phase problem [119].

3.4.1.2.1 Ab initio phasing

Ab initio phasing or direct methods are usually the method of choice for small molecules (<1000 non-hydrogen atoms), and have been used successfully to solve the phase problems for small proteins. If the resolution of the data is better than 1.4 Å, direct methods can be used to obtain phase information, by exploiting known phase relationships between certain groups of reflections [120, 121].

3.4.1.2.2 Molecular replacement

If the co-ordinates of a similar protein to the one of interest exist, the structure might be solved by the molecular replacement method. This method involves rotation and translation of an existing model in order to match it with the diffraction data of the protein of interest. Hence the name: molecular replacement. The unknown protein is replaced by a known model in order to solve the phase problem. If the method is successful, the amplitudes and phases for the electron density map of the new model can be calculated [119, 122].

3.4.1.2.3 Multiple isomorphous replacement

When electron-dense metal atoms can be introduced into the crystal, their location and subsequently initial phases can be obtained. Heavy atoms can be introduced either by soaking the crystal in a heavy atom-containing solution, or by co-crystallization (growing the crystals in the presence of a heavy atom). The changes in the scatterred amplitudes can be interpreted to yield the phases. Even though this is the original method by which protein crystal structures were solved, it has largely been largely superseded by MAD and SAD phasing with selenomethionine [123].

3.4.1.2.4 Anomalous X-ray scattering

Sometimes crystallographers can make use of the anomalous scattering of certain atoms in the lattice at or near their X-ray absorption edges to gain the phase information. Most of the heavy atoms used in isomorphous replacement provide anomalous signal. This method, termed also anomalous dispersion, relies entirely on the anomalous differences produced by one or more anomalously scattering atoms in the crystal. The list of commonly used atoms includes iodine, sulphur, manganese, iron, etc [124-126].

3.4.1.3 Refinement

Having obtained the initial phases, an initial model can be built. This model can be used to refine the phases, leading to an improved model, and so on. Refinement is a procedure during which atomic positions and their respective B-factors (accounting for the thermal motion of the atom) are adjusted to fit the observed diffraction data, ideally yielding a better set of phases. A new model can then be fitted to the new electron density map and a further round of refinement is carried out. This continues until the correlation between the diffraction data and the model is maximized [127].

3.4.2 Protocols

All diffraction data were collected at the XRD1 workstation at Synchrotrone Elletra, Trieste, unless otherwise stated. All data were processed using the HKL2000 package [118].

3.4.2.1 Complexes of stefin A with cathepsins

All structures were determined by molecular replacement using Amore [128] with cathepsin V [129], cathepsin L [19], cathepsin B [25] and stefin A [24] as a search model. The structures were refined using Refmac [130] and MAIN [131].

3.4.2.2 Macrocypin

The macrocypin structure was solved with SAD phasing from the data collected from the crystal soaked in

the saturated solution of sodium iodine. The data set was collected to 2.2 Å resolution on the in-house Rigaku rotating anode (RU 200) using Xenox mirrors. 615 images were collected from a single crystal with linear R-merge of 13% and redundancy of 30. SAD phasing was based on 15 iodine positions with occupancy ranging from 0.8 to 0.15 using automated SOLVE/RESOLVE scripts incorporated in the AutoSol module of the PHENIX suite [132]. Automated model building and docking to the macrocypin sequence gave a solution with approximately 120 out of 159 amino acids. Despite good data quality, we were unable to refine the structure, presumably due to the multiple conformations of several loop regions induced by binding iodine ions - quite a few of them with low occupancy at positions inside the protein core. Therefore another data set was collected with the crystals, grown in sodium citrate. This data set was phased with the partial structure of macrocypin using molecular replacement program Amore [133]. Cycles of manual and automated building with ARP/warp [134] and refinement with Refmac [130] and MAIN [131] were performed until all residues were built in the electron density map. The final structure was refined using MAIN against 1.64 Å resolution data [131].

3.4.2.3 Cathepsin V – clitocypin complex

The crystal structure of cathepsin V – clitocypin complex was determined by the molecular replacement method with AMoRe, using cathepsin V (PDB code: 1fh0)[129] as the search model. Four molecules of cathepsin V were positioned into the asymmetric unit. The fourfold electron density averaging in MAIN [131] produced maps that enabled us to build the substantial parts of the clitocypin structure manually. The model building was accelerated by using the similarity between the partial model of clitocypin and macrocypin scaffolds, enabling the manual superimposition of the two models. The positions of the two SeMet residues in the clitocypin sequence were helpful in the initial sequence assignment. The structure was refined using MAIN against 2.24 Å resolution data.

3.4.2.4 Clitocypin

The diffraction data was collected previously [115].

The structure of clitocypin was determined by molecular replacement using Amore [128] with the clitocypin structure in complex with cathepsin V as a search model. The structure was refined using Refmac [130] or MAIN [131].

3.5 Atomic force microscopy

3.5.1 Theoretic background

3.5.1.1 Atomic force microscope

Atomic force microscopy (AFM) or scanning force microscopy (SFM) is a very high-resolution type of scanning probe microscopy, with demonstrated resolution of fractions of a nanometer, more than 1000 times better than the optical diffraction limit. Binnig, Quate and Gerber invented the first atomic force microscope (also abbreviated as AFM) in 1986 [135]. The AFM is one of the foremost tools for imaging, measuring, and manipulating matter at the nanoscale. The information is gathered by "feeling" the surface with a mechanical probe. Piezoelectric elements that facilitate tiny but accurate and precise movements on (electronic) command enable the very precise scanning.

The AFM consists of a cantilever with a sharp tip (probe) at its end that is used to scan the specimen surface. The cantilever is typically silicon or silicon nitride with a tip radius of curvature on the order of nanometers. When the tip is brought into the proximity of a sample surface, forces between the tip and the sample lead to a deflection of the cantilever according to the Hooke's law. Depending on the situation, forces measured by AFM include mechanical contact force, van der Waals forces, capillary forces, chemical bonding, electrostatic forces, magnetic forces, etc. Typically, the deflection is measured using a laser spot reflected from the top surface of the cantilever into an array of photodiodes.

If the tip was scanned at a constant height, a risk would exist that the tip collides with the surface, causing damage. Hence, in most cases a feedback mechanism is employed to adjust the tip-to-sample distance to maintain a constant force between the tip and the sample. Traditionally, the sample is mounted on a piezoelectric tube, that can move the sample in the z direction for maintaining a constant force, and the x and y directions for scanning the sample. Alternatively a 'tripod' configuration of three piezo crystals may be employed, with each responsible for scanning in the x, y, and z directions. In newer designs, the tip is mounted on a vertical piezo scanner while the sample is being scanned in X and Y using another piezo

block. The resulting map of the area s = f(x, y) represents the topography of the sample.

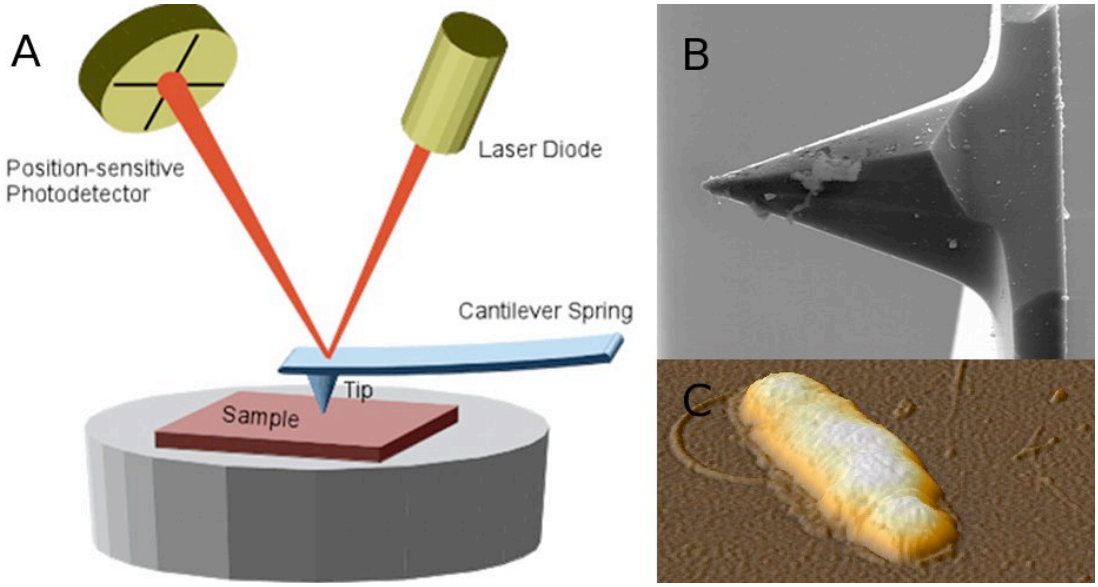


Figure 15: *Schematic representation of atomic force microscope (A), an electron microscope image of the tip (B) and topography image of bacteria, obtained by AFM (C).*

3.5.1.2 Force spectroscopy

Another major application of AFM (besides imaging) is the force spectroscopy, a direct measurement of tip-sample interaction forces as a function of the gap between the tip and sample (the result of this measurement is called a force-distance curve). For this method, the AFM tip is extended towards and retracted from the surface as the deflection of the cantilever is monitored as a function of piezoelectric displacement. These measurements have been used to measure nanoscale contacts, atomic bonding, Van der Waals forces, etc. Forces of the order of a few pico-Newtons can now be routinely measured with a vertical distance resolution of better than 0.1 nanometer [136].

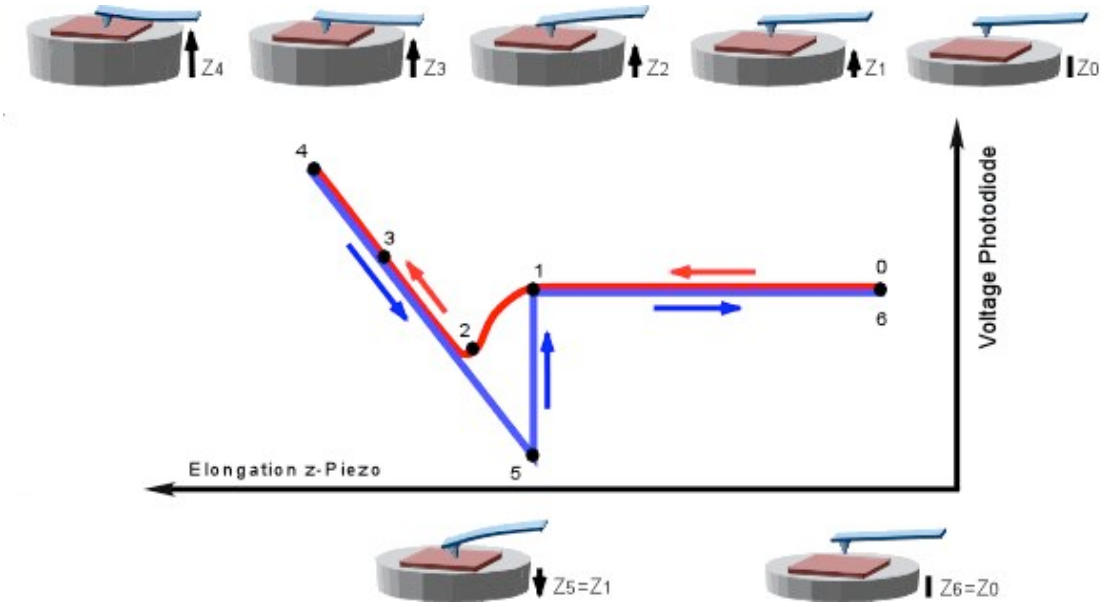


Figure 16: *Force masurements*. (0) the tip is far away from the surface, and surface forces do not act. As the tip approaches (1) to (2), it enters the range of attractive surface forces and is deflected downwards. (2) to (4): the tip is in contact with the surface and now exerts pressure while it is deflected upwards. At (4) the tip retraction started, but adhesion forces may keep the tip attached to the surface until the spring force exerted by the cantilever can overcome adhesion (5). Then the tip snaps back into its initial position and the cycle can start again (6).

When the tip and the surface are immobilized with proteins, the interaction forces can be measured. There are reports about measured forces between α5β1 integrin ligand/GRGDSP peptide receptor [137], cadherin/cadherin [138], selectin/sialyl Lewis X tetrasaccharide [139], regulatory protein $(His)_6$ExpG/DNA promoter regions [140] and others.

3.5.1.3 Bell's model

The unbinding kinetics cannot be described with a classical biochemical equation. Commonly accepted Bell's model, which is used to explain the unbinding process, treats it as an escape of a ligand from the potential well [141, 142]. The dissociation rate constant k_{off} as a function of force can be presented as:

$$k_{off}(F) = k_{off}(0) \cdot \exp\left[F x_\beta / (k_\beta T)\right] \quad (1)$$

where $k_{off}(0)$ is the natural thermal off-rate for the dissociation at zero force, k_β the Boltzmann's constant, T the temperature, F the the applied force, x_β is the distance between the potential minimum and the transition state in the Bell's model.

If we treat the bond dissociation as a random process and we use the constant loading rate v, we can describe the most probable unbinding force F^* as:

$$k_{off}(F) = k_\beta T / x_\beta \cdot \ln\left[v x_\beta / (k_{off} k_\beta T\right] \quad (2)$$

3.5.2 Protocols

3.5.2.1 Cantilever preparation

Cantilevers (Veeco) were first cleaned three times in chloroform and then dried with argon. Oxidation and additional cleaning was achieved with the exposure of tips to the ozone atmosphere in a home-build ozone cleaner. Tips were washed with ethanol, dried and immersed overnight in the ethanolamine solution (3,3 g in 6 mL DMSO). After formation of an amino terminated surface, the tips were washed with DMSO, ethanol and dried. Subsequent incubation of tips with maleimidopropionyl-PEG N-hydroxysuccinimide ester (MAL-PEG-NHS) linker (Polypure) solution in chloroform resulted in the stable binding of the NHS reactive group to the amine groups. Maleimide reactive group was used for the reaction with the cysteine residue inserted opposite to the binding site into the SteA-G34C. The reaction took one hour at room temperature in 100 mM phosphate buffer, pH=7.0. Before reaction, stefin A was reduced with 3 mM TCEP. The functionalized tips were washed 5 times (4x10 min, 1x overnight) in 10 mL 10 mM phosphate buffer, 1% Tween 20. Cantilevers were stored in the buffer solution at 4°C until further use.

3.5.2.2 Surface immobilization

Silicon nitride waffers were cut in 0.25 cm2 squares and cleaned in piranha solution (7 ml 96% H_2SO_4, 3 mL 30% H_2O_2) for 1 hour. The cleaning, ethanolamine modification and PEG linker binding was done as described for cantilevers.

Aldehyde at the end of the linker reacted with the amino group of surface lysines [143]. The functionalized tips were washed 5 times (4x10 min, 1x overnight) in 10 mL 10 mM phosphate buffer, 1% Tween 20. They were stored in the buffer solution at 4°C until further use.

3.5.2.3 Surface density determination

Surface density of the bound enzymes was determined with the classical biochemical activity test after each step of washing. Cathepsin L modified chips were immersed into a cuvette with a fluorogenic substrate z-Phe-Arg-AMC (20 μM) in 100 mM sodium acetate buffer, 3 mM DTT, 3 mM EDTA, pH=5.5. The change in fluorescence was followed for 1-5 minutes. The rate of reaction was compared with the calibration series in which known amounts of the soluble enzyme were added to the reaction mixture. After 1-5 minutes, the chip was removed from the cuvette and measurements continued for an additional minute. The washed amount of washed and immobilized enzyme was determined using the calibration series.

3.5.2.4 Force measurements

Measurements were carried out in the 100 mM sodium acetate buffer, containing 3 mM DTT, 3 mM EDTA, pH=5.5 using the PicoSPM microscope (Molecular Imaging, Tempe, AZ). Unbinding forces between tip-bound stefin A molecules and surface-bound catepsin L molecules were monitored by force-distance cycles at amplitudes between 200 and 800 nm and at frequencies between 0.5 and 3 Hz. The used cantilevers had a nominal spring constant of 0.01, 0.02 and 0,03 N/m. The retrace velocity ranged from 100 nm/s to 2000 nm/s, which resulted in loading rates from 1 nN/s to 50 nN/s.

The force distance cycles were performed at constant lateral position. The measurements were performed in sets containing 1000 force curves. Each set was recorded at the same loading rate.

The specificity of the unbinding event was verified with a blocking experiment. Cathepsin L was added into the liquid cell solution to the final concentration of 90 μM. The solution was incubated for 10 minutes, so the stefin A molecules on the tip were blocked with cathepsin L and the measurements were repeated.

3.6 Kinetic measurements

3.6.1 Theoretical background

Reversible competitive inhibitors are competing with substrates to bind to the enzyme active site, as shown in Figure 17.

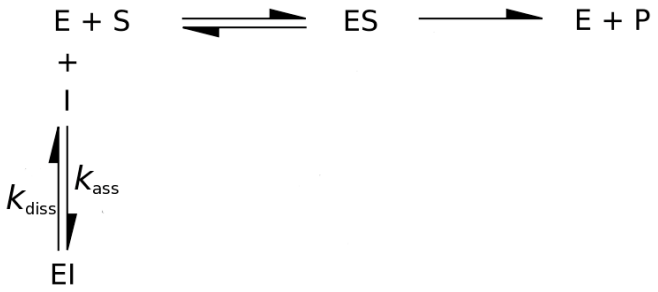


Figure 17: *Scheme of enzyme inhibition by reversible competitive inhibitors* (E – enzyme, S – substrate, P – product).

k_{diss} and k_{ass} are rate constants for inhibitor association and dissociation, respectively.

The equilibrium constant K_i is a measure of the strength of the interaction between the enzyme and inhibitor that come together to form the complex and can be calculated as:

$$K_i = \frac{k_{diss}}{k_{ass}} \tag{3}$$

When determining the equilibrium constant between enzymes and the slow tight binding inhibitors without the preequilibration step[144], the product formation is described as:

$$[P] = \frac{v_s \cdot t + (v_0 - v_s) \cdot (1 - e^{-kt})}{k} \tag{4}$$

where [P] is product concentration, vs is rate of product formation in stationary state, v_0 initial rate of product formation and k is pseudo-first-order rate constant describing the presteady state of the reaction.

Pseudo-first-order rate constant k has to be determined for a number of inhibitor concentrations and kass and kdiss are calculated with linear regression form equation:

$$k = \frac{k_{ass} \cdot [I]}{1 + [S]/K_m} + k_{diss} \tag{5}$$

3.6.2 Protocols

3.6.2.1 Determination of the inhibition kinetics

Inhibition kinetic of cathepsins and all the inhibitors were determined in continuous kinetics assays at 25°C. In all kinetic experiments, cathepins L and V were assayed using 100 mM phosphate buffer, pH 5.5, containing 3 mM EDTA and 3 mM DTT. The active concentration of the enzymes was determined by the active site titration using synthetic inhibitor E-64. All inhibition experiments were performed under the pseudo first-order conditions with inhibitor concentration at least 40-fold higher that of the enzyme concentration. Less than 10% of the substrate was hydrolyzed during these experiments. Stefin A, macrocypin, clitocypin, and their mutants were mixed at various concentrations with the substrate solution in the phosphate buffer in the fluorometric cuvette. The reaction was initiated by addition of the enzyme in a negligible volume. The progress curves were monitored at excitation and emission wavelengths of 370 and 460 nm, respectively, using a C-61 fluorimeter (Photon Technology International). Typical biphasic curves were observed and were analyzed by the method described previously [144].

The inhibition kinetic of trypsin and legumain with macrocypin, clitocypin, and their mutants were determined in the stopped kinetic assays at 25°C and analyzed as described [145].

3.6.2.2 Interactions of stefin B with the histones

Cathepsin L (final concentration 21.5 nM) was incubated for 5 minutes with different concentrations (final concentration from 10.8 nM to 2.15uM) of histones. When determining the stefin B and histon interactions, stefin B (final concentration 15 nM) was preincubated with different concentrations (final concentration from 10.8 nM to 2.15uM) of histones for 5 minut, than cathepsin L was added (final concentration 21.5 nM). After initial incubation steps substrate z-Phe-Arg-pNA (Bachem) was added to 100 uM final concentration and initial velocity was determined spectrophotometrically at 410 nM.

3.7 Surface plasmon resonance

3.7.1 Theoretical background

At the interface between two transparent media of different refractive index (glass and water), light coming from the side of higher refractive index is partly reflected and partly refracted. Above a certain critical angle of incidence, no light is refracted across the interface, and total internal reflection is observed. While incident light is totally reflected, the electromagnetic field component penetrates a short (tens of nanometers) distance into a medium of a lower refractive index creating an exponentially detenuating evanescent wave. When the interface between the media is coated with a thin layer of metal (gold), and light is monochromatic and p-polarized, the intensity of the reflected light is reduced at a specific incident angle producing a sharp shadow (called surface plasmon resonance), due to the resonance energy transfer between the evanescent wave and the surface plasmons.

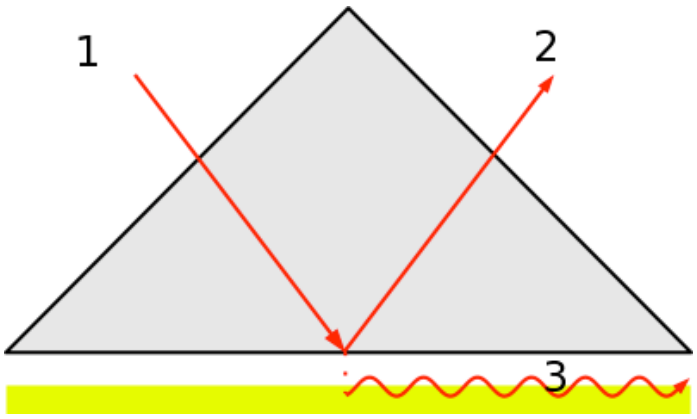


Figure 18: *Surface plasmon resonance*. Incoming light (1) deflects at the boundary between air or water (gray) and metal (yellow), where surface plasmons (3) are formed.

The resonance conditions are influenced by the material adsorbed onto the thin metal film. Satisfactory linear relationship is found between the resonance energy and the mass concentration of biochemically relevant molecules such as proteins, sugars and DNA. The SPR signal which is given in resonance units is therefore a measure of mass concentration at the chip surface of the senzor. This means that the analyte and ligand association and dissociation can be observed and ultimately rate as well as equilibrium constants can be derived.

3.8 Protocols

The surface plasmon resonance (SPR) measurements were performed using the Biacore T100 instrument (Biacore AB, Sweden) at 25°C temperature. Stefin A was immobilized on the surface of a CM5 chip to approximately 50 response units (RU) as recommended by the supplier (Biacore AB, Sweden). The buffer used was 100 mM Na-acetate, 150 mM NaCl, 3 mM EDTA, 0.005% P20, pH=5.5 and the flow rate was set to 5 µl/min. The association was followed for 1 min and the dissociation for 5 min. The regeneration of chips was achieved with trietilamine, pH 10.5 for 60s or 50 mM NaOH for 30 s.

4 Results

4.1 Protein expression and isolation

4.1.1 Cathepsins

Cathepsins L and V were successfully expressed in *P. pastoris* and cathepsin B in *E. coli*. 20 mg of cathepsin V, 10 mg of cathepsin L, and 10 mg of cathepsin B were purified from one liter of expression medium.

After purification and prior freezing at -80°C, all cathepsins were blocked with methyl methanethiosulfonate (MMTS). MMTS reacts with the sulfhydryl group of active site cysteine (R-SH + MMTS → R-S-S-CH_3) to prevent oxidation damage to the active site cysteine residue and autodegradation.

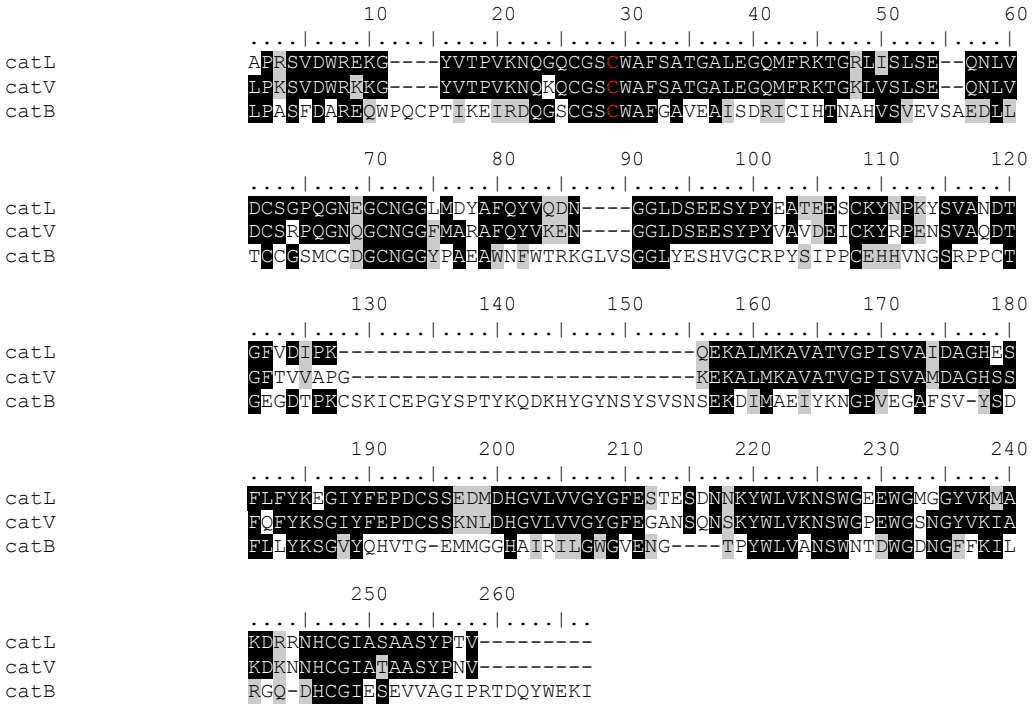


Figure 19: *Sequence alignment of expressed cathepsins L, V and B*. Active site residues are shown in red.

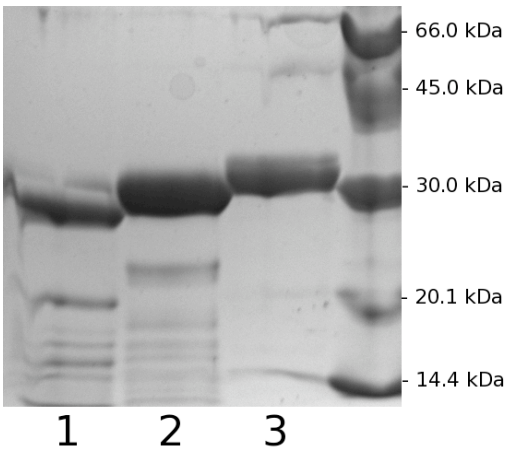


Figure 20: *SDS-PAGE of expressed and isolated cathepsins.* 1) cathepsin V, 2) cathepsin L and 3) cathepsin B.

4.1.2 Stefin A and its mutants

The native stefin A and its mutant steA-T34C were successfully expressed. More than 50 mg of both proteins were purified from one liter of expression medium.

SteA-T34C formed dimers due to the free cysteine residues in structures, so we reduced the steA-T34C with DTT or CTAB before use.

```
                    10        20        30        40        50        60
                    ....|....|....|....|....|....|....|....|....|....|....|....|
steA                MIPGGLSEAKPATPEIQEIVDKVKPQLEEKTNETYGKLEAVQYKTQVVAGTNYYIKVRAG
steA-T34C           MIPGGLSEAKPATPEIQEIVDKVKPQLEEKTNECYGKLEAVQYKTQVVAGTNYYIKVRAG

                    70        80        90
                    ....|....|....|....|....|....|....|...
steA                DNKYMHLKVFKSLPGQNEDLVLTGYQVDKNKDDELTGF
steA-T34C           DNKYMHLKVFKSLPGQNEDLVLTGYQVDKNKDDELTGF
```

Figure 21: *Sequence alignment of expressed stefin A and its mutant*. Mutated amino acids are highlighted in red.

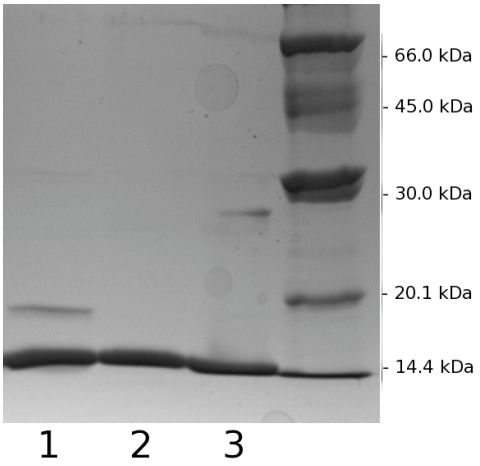


Figure 22: *SDS-PAGE of expressed and isolated stefins A.* 3) steA-T34C without DTT, 2) steA-T34C with DTT and 1) native stefin A.

4.1.3 Clitocypin and its mutants

Native clitocypin and its mutants rClt-L82M,I89M, selenomethionine mutant rClt-L82M,I89M, rClt-G24A, rClt-ΔG24 and rClt-N70K were successfully expressed. More than 50 mg of all proteins were purified from one liter of expression medium.

```
10        20        30        40        50        60
                    ....|....|....|....|....|....|....|....|....|....|....|....|
rClt                MASLEDGTYRLRAVTTSNPDPGVGGEYATVEGARQPVKAEPSTPPFFERQIWQVTRNSDG
rClt-L82M,I89M      MASLEDGTYRLRAVTTSNPDPGVGGEYATVEGARQPVKAEPSTPPFFERQIWQVTRNSDG
rClt-G24A           MASLEDGTYRLRAVTTSNPDPGVAGEYATVEGARQPVKAEPSTPPFFERQIWQVTRNSDG
rClt-Δ24G           MASLEDGTYRLRAVTTSNPDPGV-GEYATVEGARQPVKAEPSTPPFFERQIWQVTRNSDG
rClt-N70K           MASLEDGTYRLRAVTTSNPDPGVGGEYATVEGARQPVKAEPSTPPFFERQIWQVTRNSDG

                              70        80        90       100       110       120
                    ....|....|....|....|....|....|....|....|....|....|....|....|
rClt                QSTIKYQGLNAPFEYGFSYDQLEQNAPVIAGDPKEYILQLVPSTTDVYIIRAPIQRVGVD
rClt-L82M,I89M      QSTIKYQGLNAPFEYGFSYDQMEQNAPVMAGDPKEYILQLVPSTTDVYIIRAPIQRVGVD
rClt-G24A           QSTIKYQGLNAPFEYGFSYDQLEQNAPVIAGDPKEYILQLVPSTTDVYIIRAPIQRVGVD
rClt-Δ24G           QSTIKYQGLNAPFEYGFSYDQLEQNAPVIAGDPKEYILQLVPSTTDVYIIRAPIQRVGVD
rClt-N70K           QSTIKYQGLKAPFEYGFSYDQLEQNAPVIAGDPKEYILQLVPSTTDVYIIRAPIQRVGVD

                              130       140       150
                    ....|....|....|....|....|....|..
rClt                VEVGVQGNNLVYKFFPVDGSGGDRPAWRFTRE
rClt-L82M,I89M      VEVGVQGNNLVYKFFPVDGSGGDRPAWRFTRE
rClt-G24A           VEVGVQGNNLVYKFFPVDGSGGDRPAWRFTRE
rClt-Δ24G           VEVGVQGNNLVYKFFPVDGSGGDRPAWRFTRE
rClt-N70K           VEVGVQGNNLVYKFFPVDGSGGDRPAWRFTRE
```

Figure 23: *Sequence alignment of native and expressed clitocypin mutants*. Mutated amino acids are highlighted in red.

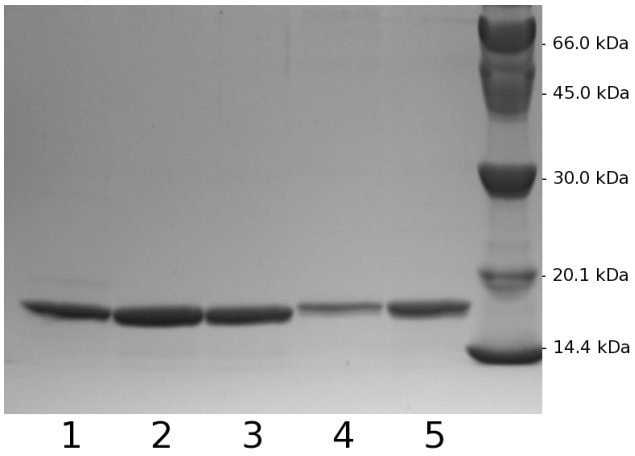


Figure 24: *SDS-PAGE of expressed and isolated clitocypin and its mutants*. 1) rClt-L82M,I89M, 2) selenomethionine mutant rClt-L82M,I89M, 3) rClt-G24A, 4) rClt-ΔG24, 5) rClt-N70K

4.1.4 Macrocypin and its mutants

Native macrocypins 1 and 4 and mutants rMcp1-G25A, rMcp1-ΔG25, Mcp4-K72R, Mcp1-I72A+N74K and Mcp4-A70I+K72N were successfully expressed. More than 50 mg of all proteins were purified from one liter of expression medium.

```
                    10        20        30        40        50        60
                    ....|....|....|....|....|....|....|....|....|....|....|....|
Mcp1                MGFEDGFYTILHLAEGQHPNSKIPGGMYASSKDGKDVPVTAEPLGPQSKIRWWIARDPQA
Mcp4                MALEDGFYTIRHLVEGQHPS--IPGGMYASSKDGKDPVTAEPLGPHSKIRWWIAAAPPA
Mcp1-Δ25G           MGFEDGFYTILHLAEGQHPNSKIP-GMYASSKDGKDVPVTAEPLGPQSKIRWWIARDPQA
Mcp1-G25A           MGFEDGFYTILHLAEGQHPNSKIPAGMYASSKDGKDVPVTAEPLGPQSKIRWWIARDPQA
```

```
Mcp4-K72R          MALEDGFYTIRHLVEGQHPS--IPGGMYASSKDGKDEPVTAEPLGPHSKIRWWIAAAPEA
Mcp1-I72A+N74K     MGFEDGFYTILHLAEGQHPNSKIPGGMYASSKDGKDVPVTAEPLGPQSKIRWWIARDPQA
Mcp4-A70I+K72N     MALEDGFYTIRHLVEGQHPS--IPGGMYASSKDGKDEPVTAEPLGPHSKIRWWIAAAPEA

                             70        80        90       100       110       120
                   ....|....|....|....|....|....|....|....|....|....|....|....|
Mcp1               GDDMYTITEFRIDNSIPGQWSRSPVETEVPVYLYDRIKAEETGYTCAWRIQPADHGADGV
Mcp4               GDDMYTITEFRADKSIPGQWARSPTEIGVPVYLYDRIKAEETGYTCVWRIQPTYEGVGGV
Mcp1-Δ25G          GDDMYTITEFRIDNSIPGQWSRSPVETEVPVYLYDRIKAEETGYTCAWRIQPADHGADGV
Mcp1-G25A          GDDMYTITEFRIDNSIPGQWSRSPVETEVPVYLYDRIKAEETGYTCAWRIQPADHGADGV
Mcp4-K72R          GDDMYTITEFRADRSIPGQWARSPTEIGVPVYLYDRIKAEETGYTCVWRIQPTYEGVGGV
Mcp1-I72A+N74K     GDDMYTITEFRADKSIPGQWSRSPVETEVPVYLYDRIKAEETGYTCAWRIQPADHGADGV
Mcp4-A70I+K72N     GDDMYTITEFRIDNSIPGQWARSPTEIGVPVYLYDRIKAEETGYTCVWRIQPTYEGVGGV

                             130       140       150       160
                   ....|....|....|....|....|....|....|....|....|....
Mcp1               YHIVGNVRIGSTDWADLREEYGEPQVYMKPVPVIPNVYIPRWFILGYEE
Mcp4               YNIMGNSRIGSTDWADLRGPDGKPQVYTKPVPVIPNVYIPRWFISEYKE
Mcp1-Δ25G          YHIVGNVRIGSTDWADLREEYGEPQVYMKPVPVIPNVYIPRWFILGYEE
Mcp1-G25A          YHIVGNVRIGSTDWADLREEYGEPQVYMKPVPVIPNVYIPRWFILGYEE
Mcp4-K72R          YNIMGNSRIGSTDWADLRGPDGKPQVYTKPVPVIPNVYIPRWFISEYKE
Mcp1-I72A+N74K     YHIVGNVRIGSTDWADLREEYGEPQVYMKPVPVIPNVYIPRWFILGYEE
Mcp4-A70I+K72N     YNIMGNSRIGSTDWADLRGPDGKPQVYTKPVPVIPNVYIPRWFISEYKE
```

Figure 25: *Sequence alignment of native and expressed macrocypin mutants.* Mutated amino acids are highlighted in red.

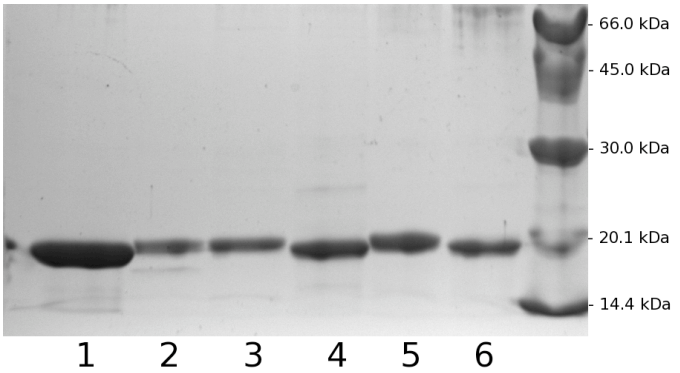


Figure 26: *SDS-PAGE of expressed and isolated macrocypin and its mutants.* 1) Mcp1, 2) Mcp-ΔG25, 3) Mcp1-G25A, 4) Mcp4-K72R 5) Mcp1-I72A+N74K, 6) Mcp4-A70I+K72N

4.2 Crystallization

Complexes of unreduced cathepsin V with stefin A (CatV-steA or VA), reduced cathepsin V with stefin A (r-CatV-steA or rVA), unreduced cathepsin L with stefin A (CatL-steAor LA), unreduced cathepsin B with stefin A (CatB-steA or BA), unreduced cathepsin V with clitocypin (CatV-clt) and macrocypin 1 alone (Mcp1) were successfully crystallized. Table 1 summarises the crystallization conditions, growth time, and the maximal size of crystals.

Table 1: Crystallization conditions, growth time and maximal size of crystals.

	Condition	Growth time	Max. size
VA	0.1 M Tris-HCl, pH=8.0 12% PEG3000	2 – 10 days	0.4×0.4×0.1 mm^3
rVA	0.1 M Tris-HCl, pH=8.25 16% PEG3000	5 – 20 days	0.3×0.3×0.1 mm^3
LA	0.1 M Tris-HCl, pH=7.0 16% PEG3000	10 – 30 days	0.1×0.1×0.1 mm^3
BA	0.2 M sodium sulphate 24% PEG3000	5 – 20 days	0.4×0.1×0.1 mm^3
CatV-clt	M Li2SO4, 12% PEG800	3 – 5 months	0.2×0.2×0.1 mm^3

Mcp1	20% glycerol BisTrisPropane buffer, pH 7.0 200mM sodium citrate 20% PEG3350	6 – 24 hours	0.5×0.5×0.5 mm^3

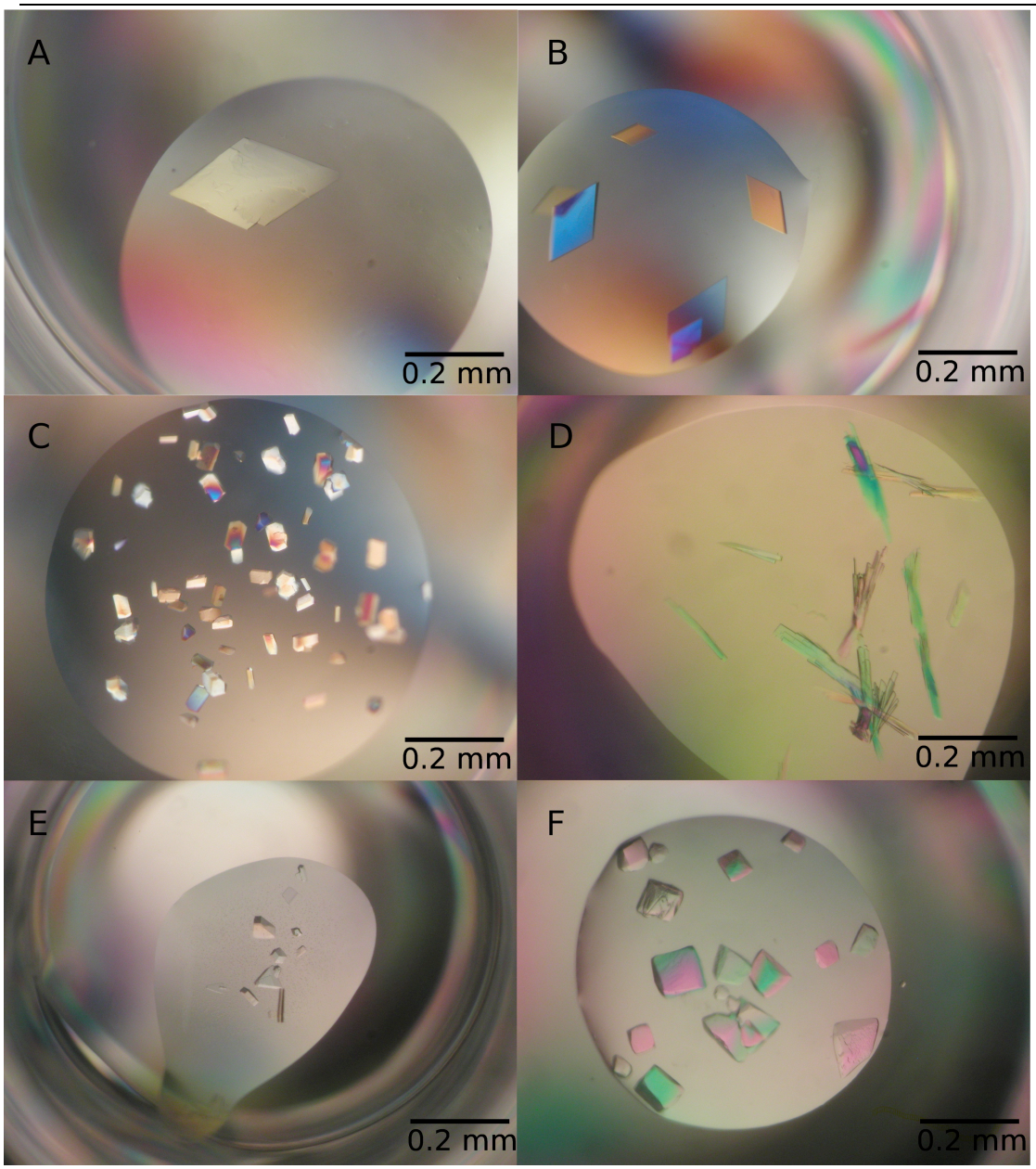


Figure 27: *Crystal images.* A) rVA, B) VA, C) LA A, D) BA, E) cathepsin V – clitocypin, F) macrocypin 1.

4.3 Data collection and refinement statistic

Data collection and refinement statistics are summarized in Table 2 for VA, rVA and LA, in Table 3 for BA, and in Table 4 for catV-clt, Mcp and Clt structures.

Table 2: Data collection and refinement statistics for complexes of unreduced cathepsin L, reduced and unreduced cathepsin V with stefin A. Numbers in parentheses are for the highest resolution shell. Only one crystal was used for each structure. No reflection cutoffs were applied. Structures are deposited in PDB database and will be freely accessible after publications will be accepted in press.

Data collection	VA	rVA	LA
PDB ID	3KSE	3A9N	3KFQ
Space group	P212121	P212121	P1
Cell dimensions			
a, b, c (Å)	59.3, 107.3, 116.0	60.0, 105.2, 110.2	35.3, 83.5, 83.9
α, β, γ (°)	90, 90, 90	90, 90, 90	118.1, 98.0, 98.0
Resolution (Å)	78.76 - 1.99	76.0 - 1.81	71.9 - 1.70
R_{merge} (%)	8.4 (47.7)	5.7 (23.1)	4.0 (27.3)
I/σI	28.7 (3.7)	39.5 (5.2)	20.1 (2.7)
Completeness (%)	99.5 (91.5)	98.1 (61.7)	95.2 (62.7)
Redundancy	8.7 (7.9)	7.2 (3.4)	2.0 (1.3)
Refinement			
Resolution	78.76 - 1.99	76.0 - 1.81	71.9 - 1.70
No. of reflections (work/free)	48400/2586	60557/3248	80215/4292
R_{work}/R_{free}	16.4/21.9	15.8/20.0	14.7/19.5
B factors			
Protein	31.2	18.6	16.6
Water	44.0	33.1	28.6
No. of atoms			
Protein	4932	6060	7516
Water	722	1057	1146
r.m.s. deviation			
Bond lengths (Å)	0.018	0.026	0.025
Bond angles (°)	1.58	2.11	2.20

Table 3: *Data collection and refinement statistics for complex of cathepsin B with stefin A.* Numbers in parentheses are for the highest resolution shell. Only one crystal was used for the structure. No reflection cutoffs were applied. Structure is deposited and freely available from the PDB database.

Data collection	BA
PDB ID	3K9M
Space group	P1
Cell dimesions	
a, b, c (Å)	62.0, 31.0, 70.9
α, β, γ (°)	90, 104.5, 90
Resolution (Å)	68.6 – 2.51
R_{merge} (%)	8.4 (20.6)
I/σI	9.5 (2.6)
Completeness (%)	92.1 (66.7)
Redundancy	2.6 (2.2)
Refinement	
Resolution	40.5 – 2.61
No. of reflections (work/free)	24360 / 713
R_{work}/R_{free}	19.8 / 25.0
B factor (Å 2)	42.0
No. of atoms	
Protein	5454
Water	127
r.m.s. deviation	
Bond lenghts (Å)	0.013
Bond angles (°)	1.71

Table 4: *Data collection and refinement statistics for clitocypin, macrocypin and complex of clitocypin with cathepsin V*. Numbers in parentheses are for the highest resolution shell [146]. Only one crystal was used for each structure. No reflection cutoffs were applied. Structures are deposited and freely available from PDB database.

Data collection	Macrocypin	Clitocypin – cathepsin V	Clitocypin
PDB ID	3H6Q	3H6S	3H6R
Space group	P3121	P21212	P21
Cell dimensions			
a, b, c (Å)	77.1, 77.1, 60.9	98.2, 177.8, 60.9	46.5, 58.0, 58.3
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 111.2, 90
Resolution (Å)	50 – 1.64	27.5-2.22	30 -1.94
R_{merge} (%)	5.0 (14.2)	3.7 (19.4)	2.6 (9.8)
I/σI	69.0 (13.1)	43.9 (6.1)	63.1 (16.8)
Completeness (%)	98.6 (86.5)	98.8 (76.8)	97.9 (85.0)
Redundancy	10.3 (8.3)	3.9 (1.8)	7.2 (6.9)
Refinement			
Resolution	27.7 – 1.64	27.5-2.22	30 -1.948
No. of reflections (work/free)	24490/1282	76800/4054	20034/1040
R_{work}/R_{free}	16.2/19.3	18.3/23.4	18.6/24.1
B factors			
Protein	19.0	26.5	28.7
Water	35.4	37.3	40.1
No. of atoms			
Protein	1424	11500	2388
Water	321	856	278
r.m.s. deviation			
Bond lengths (Å)	0.019	0.018	0.022
Bond angles (°)	2.00	1.70	2.05

4.4 Kinetics of stefin A inhibition of cysteine proteases

4.4.1 Kinetics of active cathepsins

The active concentration of enzyme was determined by the active site titration using the synthetic inhibitor E-64. Cathepsins L and V were typically 75-85% active, whereas cathepsin B had lower activity (60%). Stefin A was usually 50-70% active. The percentage varied from batch to batch and was strongly influenced by repeating freeze – thaw cycles.

Inhibition kinetic of cathepsins and all inhibitors were determined in continuous kinetics assays at 25°C [144] and with surface plasmon measurements. Data is summarized in Table 5.

Table 5: *Kinetic constants of cathepsin inhibition by stefin A.* Data published by others: 1 - [147], 2- [148], 3 - [86]

	k_{ass} (10^6 $M^{-1}s^{-1}$)	k_{diss} (10^{-4} s^{-1})	K_i (nM)
Solution			
CatL	43.2±5.0	10.5±2.9	0.24±0.01
CatV	2.4±0.2	2.9±0.7	1.24±0.31
SPR			
CatV	21.9±3.7	1.8±0.8	0.009±0.005
CatL	2.9±0.2	2.1±0.4	0.074±0.015
Published data			
CatL[1]	/	/	0.05
CatL[2]	/	/	0.21

CatL[3]	/	/	0.02
CatV[1]	/	/	0.11
CatB[3]	/	/	0.91

The inhibition constants for cathepsins L and V inhibition by stefin A are similar to those already published. There are some differences (0.24 nM vs. 0.02, 0.05 and 0.21 nM for cathepsin L, 2.0 nM vs. 0.11 nM for cathepsin V), but these can be explained by systemic errors and differences between measurements in different labs:

- proteins used in determination of constants were expressed in different expression systems (bacterial, yeast and insect)
- some measurements were done in buffers with different ionic strength and at different temperatures
- sometimes different fluorogenic substrate was used.

Surface plasmon resonance determination of kinetic constant for cathepsin L is similar to those determined in solution (0.24 nM vs. 0.074 nM), however, the inhibition constant is quite different for cathepsin V (2 nM vs. 0.009 nM). The 220 fold differences between both techniques cannot be explained. One of the reasons for such a difference can be the tendency of cathepsin V to precipitate, bind to tubes or sensor chip, as we observed lower SPR signals with identical samples after longer measurements (more than 6 hours). It was also reported that cathepsin V (but not cathepsin L) has positively charged patches, which can bind to negatively charged objects like DNA [149].

4.4.2 Influence of ionic strength to inhibition of cathepsin L with stefin A

We have also investigated the influence of different ionic strengths on inhibition. The kinetic data are summarized in Table 6.

Table 6: *Kinetic constants of cathepsin L inhibition by stefin A in different ionic strengths.*

I / M NaCl	k_{ass} (10^6 $M^{-1}s^{-1}$)	k_{diss} (10^{-4} s^{-1})	K_i (nM)
0.00	43.2±5.0	10.5±2.9	0.24±0.01
0.25	5.4±1.0	4.4±0.6	0.81±0.26
0.50	3.6±1.6	3.2±1.2	0.88±0.30
1.00	6.1±5.1	3.7±2.0	0.60±0.33
1.50	4.1±4.0	4.9±0.6	1.20±0.25

The presence of sodium chloride in reaction buffer (100 mM acetate, 3 mM DTT, 3 mM EDTA) decreased the affinity of cathepsin L for stefin A, indicating that the interactions are partially controlled by electrostatic interactions which are influenced by the presence of NaCl.

4.4.3 Kinetics of inactivated enzymes

The complex between papain and stefin B is formed even though the active site cysteine residue was carboxymethylated [89]. The interactions between chicken cystatins C and inactivated papain were also reported [150]. When papain was blocked with MMTS, the observed strength of interaction (0.7 pM) was 10 fold weaker than with the native papain (0.06 pM). When papain was carboxymethylated, the interaction (15 pM) was 2000 fold weaker. The decreasing affinity of cystatin for papains inactivated with groups of increasing size was shown to be consistent with the larger impairment of fit between the binding regions of the two molecules due to progressively higher dissociation rate constants. [150].

Because interactions between the inactivated cathepsin L and stefin A could not be measured using fluorogenic substrate, we used the surface plasmon resonance method. Results are summarized in Table 7.

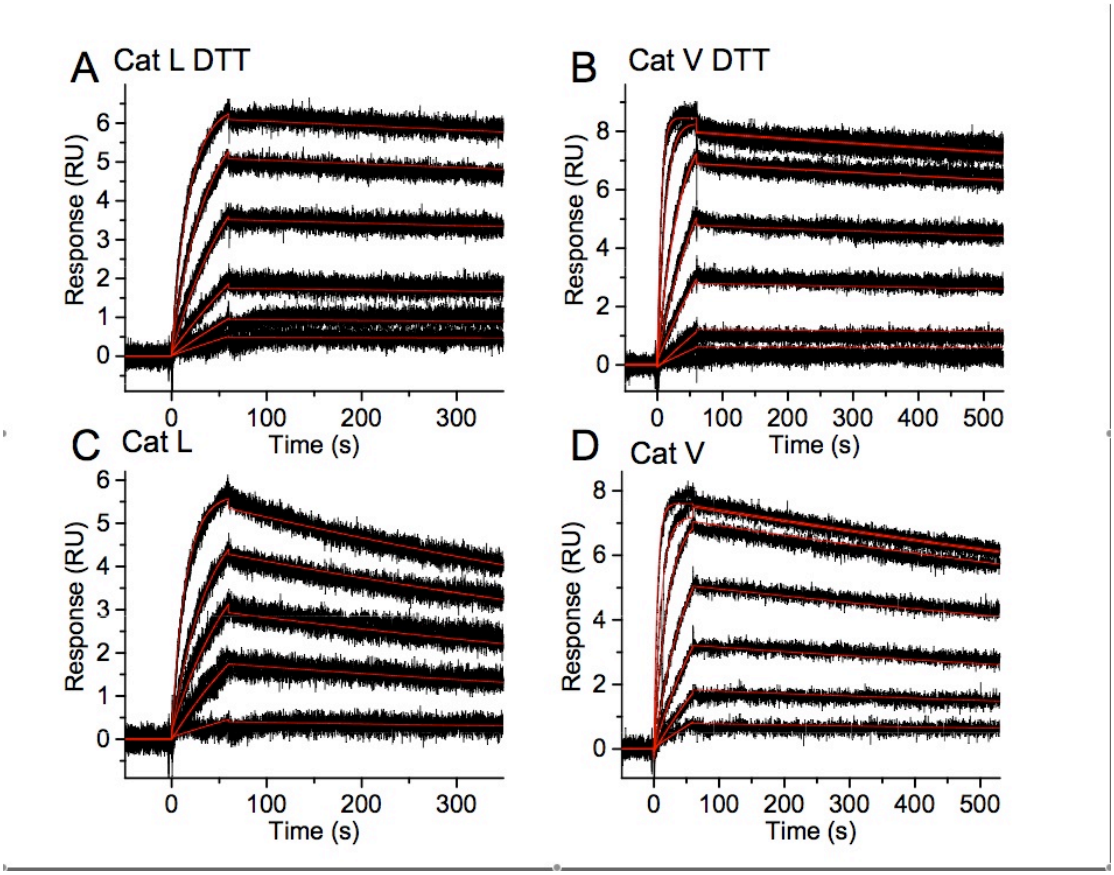


Figure 28: *SPR measurements*. A) reduced cathepsin L – stefin A, B) reduced cathepsin V – stefin A, C) unreduced cathepsin L – stefin A, D) unreduced cathepsin V – stefin A.

Table 7: *Kinetic constants of cathepsin inhibition by stefin A* (numbers in brackets represents the change from the native cathepsin).

	Side chain of Cys25	k_{ass} (10^6 $M^{-1}s^{-1}$)	k_{dis} (10^{-4} s^{-1})	K_i (nM)
Solution				
CatV	-CH_2-SH	2.4±0.2	2.9±0.7	1.24±0.31
CatL	-CH_2-SH	43.2±5.0	10.5±2.9	0.24±0.01
SPR				
CatV	-CH_2-SH	21.9±3.7	1.8±0.8	0.009±0.005
MMTS-CatV	-CH_2-S-CH_3	11.0±3.1 (0.5)	5.1±0.7 (2.8)	0.048±0.008 (5.3)
CatL	-CH_2-SH	2.9±0.2	2.1±0.4	0.074±0.015
MMTS-CatL	-CH_2-S-CH_3	1.0±0.2 (0.3)	14.2±4.6 (6.8)	1.51±0.07 (20.4)
CM-CatL	-CH_2-S-CH_2-COOH	0.056 (0.02)	63.7 (30)	109.5 (1500)

The interactions between MMTS blocked cathepsins L and V and stefin A is 5x and 20x weaker, when compared to the interactions with the native and reduced enzyme. The interaction of carboxymethylated cathepsin L is even weaker (110 nM or 1500 fold).

Contrary to the previously published data [150], when lower inhibition constants between chicken cystatins C and inactivated papains was a consequence of altered dissociation rate and not association rate, we observed similar effects on dissociation as well as on association rates (changes are shown in brackets in Table 7). The changed dissociation rate can be explained as before by the larger impairment of fit between the binding regions of the two molecules, whereas the changed association rate can be attributed to the repulsive interactions between additional atoms on cysteine residue and stefin A.

4.4.4 Interactions of the stefin B with the histones

Recently, the otherwise endosomal proteinase cathepsin L has been reported to be active in the nucleus. It cleaves the CUX1 transcription factor and as a result accelerates progression into the S phase of the cell

cycle [151]. In addition, cathepsin L was found to cleave histone H3.2 in the nucleus during mouse embryonic stem cell differentiation [152].

Interaction of stefin B with the Met-75 truncated form of cathepsin L in the nucleus was confirmed by fluorescence resonance energy transfer experiments in the living cells, as well as the interactions of stefin B with the histones H2A.Z, H2B, and H3 by immunoprecipitation with anti-stefin B antibodies [153].

First, we determined conditions under which inhibition of cathepsin L was in the linear range, i.e. up to 70% inhibition. Next, we tested if purified histones inhibit human cathepsin L in vitro and found that histones did not inhibit cathepsin L activity even at a 100-fold molar excess (2.15 μM histone concentration) (Figure 29).

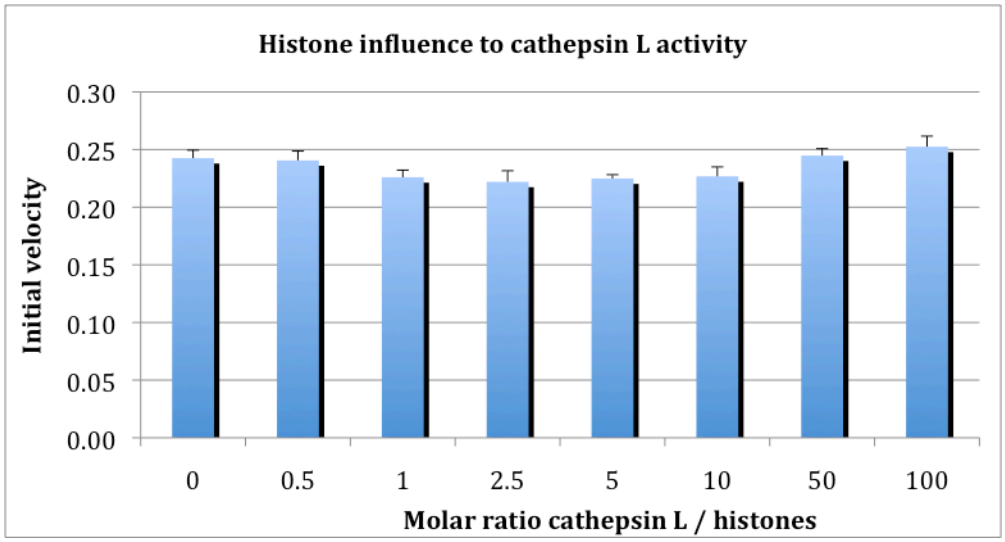


Figure 29: *Histones does not influence the cathepsin L activity*. Cathepsin L (21.5 nM) was incubated with increasing molar concentrations of histones. The data represent the mean of at least three independent experiments.

To further investigate the role of histones in the inhibition of cathepsin L by stefin B, we preincubated stefin B with increased molar concentrations of histones and measured cathepsin L activity, as described. We found that histone binding to stefin B did not affect the inhibition of cathepsin L by stefin B at lower histone concentrations (10.8 –210 nM) (Fig. 6B), whereas at higher histone concentrations (1.05–2.15 μM) we observed increased inhibition of cathepsin L by stefin B [153] (Figure 30).

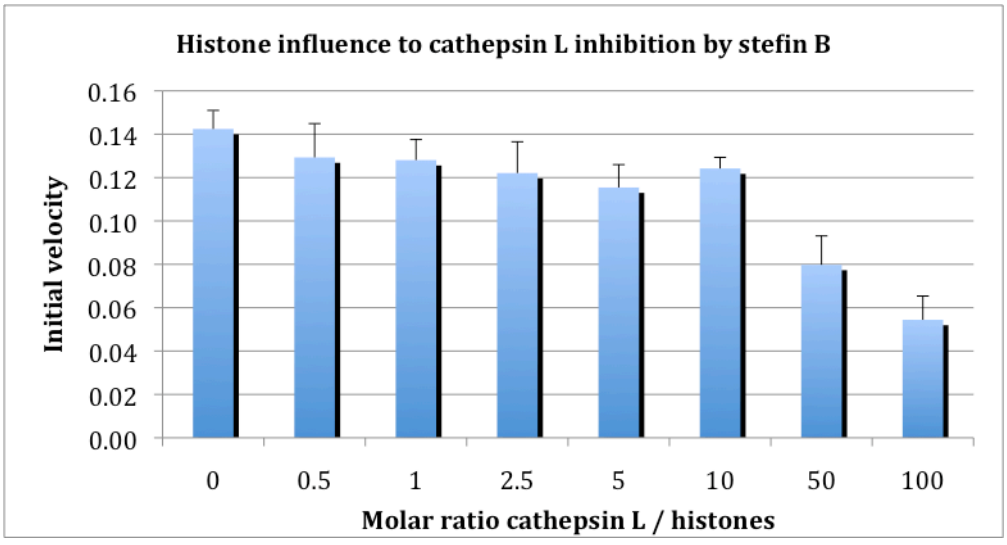


Figure 30: *Addition of histones increased the inhibitory activity of stefin B toward cathepsin L.* Stefin B (15 nM) was preincubated with increasing molar concentrations of histones (10.8 nM to 2.15 μM), before the addition of cathepsin L (21.5 nM). The data represent the mean of at least three independent experiments.

Interaction with histones represents yet another mechanism. In contrast to spacers which increase the separation between the bound molecules of cathepsins and their inhibitors and thereby decrease the strength of interactions histones have the opposite effect. We established that stefin B binds to histones and potentiate the inhibition of cathepsin L by stefin B, however, the mechanism of this interaction is not known. The crystal structure yet awats determination.

4.5 Overall structure of complexes of stefin A and cysteine proteases

As mentioned before, the structural mechanism of inhibition of cathepsins by stefin A was explained by the crystal structure of stefin B in complex with carboxymethylated papain [89] and the crystal structure of cathepsin H – stefin A complex [24]. These two structures have a few limitations. Papain is a plant protease and cathepsin H is an exopeptidase with a mini-chain which alters the genuine binding of stefin. We have therefore determined the first crystal structures of endopeptidases cathepsins with stefin A to gain insight into these type of interactions.

4.5.1 Cathepsin L – stefin A complex

The crystals of stefin A and cathepsin L complex contain complete protein sequences. The catalytic site of cathepsin L was blocked with the methyl methanethiosulfonate, leaving the $–S-CH_3$ on the active site cysteine residue. This form of cathepsin L is much more stable and resulted in diffracting crystals, whereas the reduced form of cathepsin L produced only crystalline precipitation. The crystals have P1 space group and contain three pairs of molecules per asymmetric unit. The three pairs od structures of cathepsin L and stefin A are almost identical. They exhibit RMSD over equivalent CA atoms in the ranges of 0.12 Å to 0.15 Å and 0.06 Å to 0.12 Å for cathepsin L and stefin A, respectively. All residues are well resolved, except Ala1 in all three cathepsin L molecules, Pro2 in the second cathepsin L and a few side chains. 36 residues were modelled in alternative conformation, 33 of them in cathepsin L molecules.

4.5.2 Cathepsin V – stefin A complex

The crystals of the complex of stefin A and cathepsin V contain complete protein sequences. The catalytic site of cathepsin V was blocked with the methyl methanethiosulfonate, leaving the $–S-CH_3$ on the active site cysteine residue. This form of cathepsin V is much more stable and resulted in faster growing crystals. The crystals have $P2_12_12_1$ space group and contain two pairs of molecules per asymmetric unit. The two pairs of structures of cathepsin V and stefin A are closely related. They exhibit RMSD of 0.29 Å and 0.56 Å for cathepsin V and stefin A, respectively. Most of the protein chain is unambiguously resolved by the electron density maps. The exceptions are the ends of a few side chains, short stretches Ala176-Asn180 in the second molecule of cathepsin V and Ala59-Asp61 in the second molecule of stefin A.

4.5.3 Reduced cathepsin V – stefin A complex

The crystals of the complex of stefin A and reduced cathepsin V contain complete protein sequences. The crystals have $P2_12_12_1$ space group and contain two pairs of molecules per asymmetric unit. The two pairs of structures of cathepsin V and stefin A are closely related. They exhibit RMSD of 0.27 Å and 0.42 Å for cathepsin V and stefin A respectively. The electron density map unambiguously reveals the structure of almost all residues, except for a few residues (Asn177-Ser178 in the first molecule of cathepsin V, Gly60, Asp61 and Asn77 in the first, and Asn32 in the second molecule of stefin A). 31 residues were modelled in alternative conformation, 24 of them in cathepsin V molecules.

4.5.4 Cathepsin B – stefin A complex

The crystals of the complex of stefin A and cathepsin B contain complete protein sequences. The positioning of nearly all residues is clearly revealed by the electron density maps. The exceptions are the residues Glu95 and the stretch of four occluding loop residues from Val112 to Ser115 in first molecule of cathepsin B, Gly75 and Gln76 in molecule A of stefin A, and Met1 in Glu78 in molecule B of stefin A. Additionally, 11 side chains lack adequate electron density. The RMSD distance between all pairs of superimposed CA atoms of cathepsin B molecules with the occluding loop residues 105 to 125 excluded is 0.34A, whereas the superimposed CA atoms of the pair of stefin A molecules exhibit somewhat larger RMSD (0.88A). This comparison shows that differences between two molecules of cathepsin B lie in the

occluding loop region, while the differences between the two stefin A molecules are spread out through the entire structure with a slightly increased variability in the Ser72 – Asp79 region forming the second binding loop.

4.5.4.1 Occluding loop

The occluding loop is the structural feature of cathepsin B responsible for its exopeptidase activity. The 22-residue insertion between Ile105 and Pro126 has no equivalent in other papain-like enzymes, where short connections bridge the gap. The occluding loop is cross-linked by the disulfide bridge between Cys108 and Cys119. Two salt bridges, His110 – Asp22 and Arg116 – Asp224 additionally stabilize the attachment of the loop to the body of the enzyme. To weaken the occluding loop embedding in the active site cleft, the first interaction has been disrupted in the chagasin cathepsin B complex by the H110A mutant [58] (PDB code 3CBJ) and in earlier studies [154], whereas in the here presented structure of the stefin A cathepsin B complex and in the procathepsin B structures [155, 156] the native sequences have been preserved.

The four structures shown in Figure 31 demonstrate that the occluding loop can adopt a variety of positions - the moving parts are residues between Glu109 and Asp124. The extent of the occluding loop shift from the position in the native enzyme (PDB code 1HUC) is shown in the series of structures starting with the proenzyme form (PDB code 3PBH), complex with stefin A, and chagasin [58] (PDB code 3CBJ) (Figure 4).

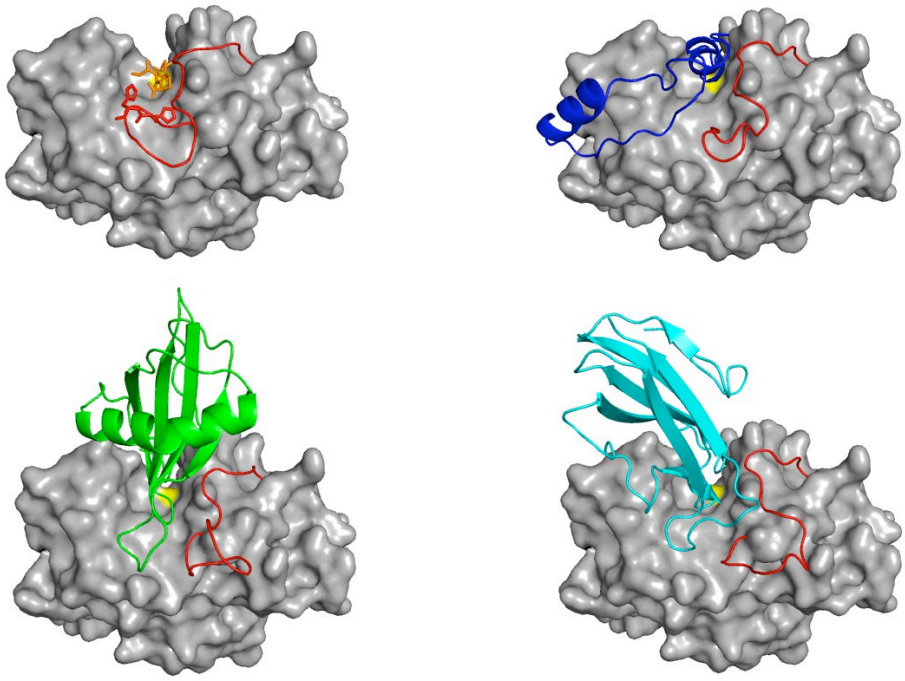


Figure 31: *Occluding loop (red) displacement by small synthetic inhibitor (orange), propeptide (blue), stefin A (green) and chagasin (cyan).*

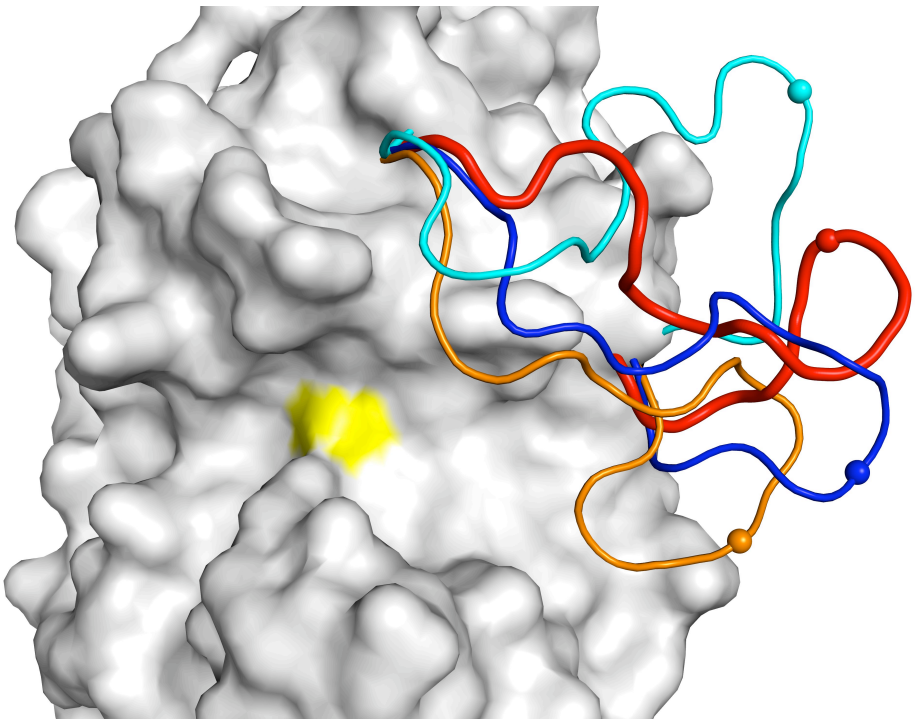


Figure 32: *Position of occluding loop in cathepsin B*. Yellow active site cysteine residue is shown. Cathepsin B core is shown in gray and occluding loop in orange (displaced by synthetic inhibitor), blue (displaced by propeptide), red (displaced by stefin A) and cyan (displaced by chagasin).

The CA atom positions of Asn113 are marked in the Figure 32 to indicate the shifted positions are 7 Å, 16 Å, and 22.5 Å away from the position in the native mature enzyme form. The conclusion here is that the size matters. The larger and the wider features have the ligands for competing with the occluding loop for the active site binding, further away the occluding loop residues are shifted. These structures thus demonstrate that the occluding loop residues can adopt a variety of conformations in contrast to the rest of the structure of cathepsin B, which appears rigid. The comparison of the interaction constants of chagasin and stefins to cathepsin B indicates that the extent of the shift does not effect the inhibition constants even though the interaction surface of chagasin with the occluding loop (160 $Å^2$) is slightly larger than that of stefin A (100 $Å^2$). This indicates that the energy cost is related to the removal of the occluding loop from the active site cleft, and that the adaption of the loop structure to fit to other ligands likely does not play a role, inhibitor or substrate – whatever comes across.

4.5.5 Complex description

Stefin A binds with its narrow edge into the V-shaped active site cleft of cathepsins. The wedged shape structure fills the active site cleft along the whole length.

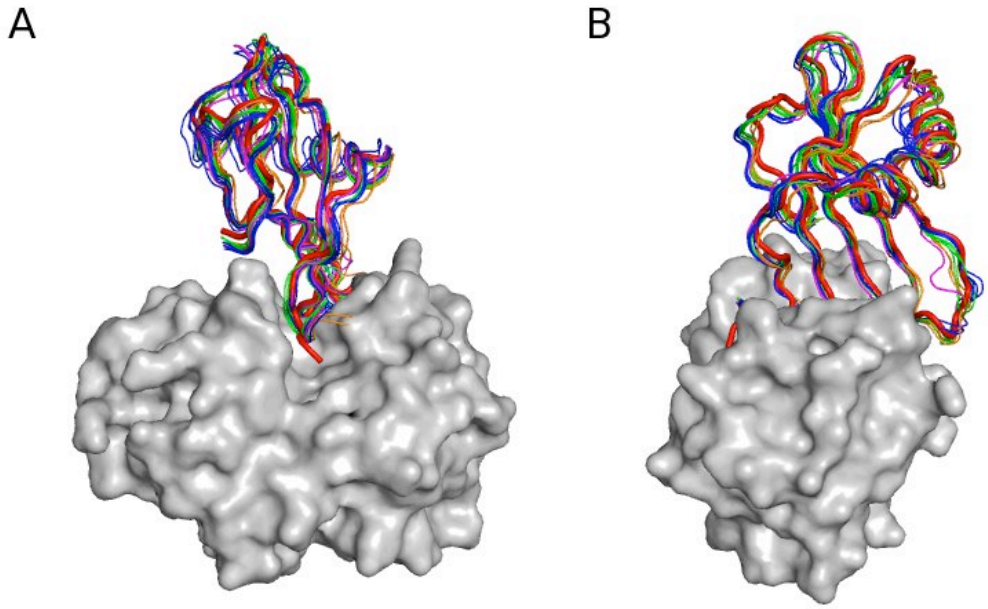


Figure 33: *Comparison of stefin binding to papain-like cysteine proteases.* A) View along the active site cleft and B) view perpendicular to the active site cleft. Papain is shown as gray surface, stefin B in papain – stefin B complex in red, stefin A in cathepsin B – stefin A complex in orange, stefin A in cathepsin L – stefin A complex in magenta, stefin A in reduced cathepsin V – stefin A complex in cyan and stefin A in cathepsin V – stefin A complex in blue.

The three binding loops are almost perfectly aligned, however, there are small differences in the angle, at which stefin A molecules binds to different cysteine proteases (Figure 33). To quantify the position of the adjacent parts of stefin A, we have arbitrary selected the CA atom of Asp61 in stefin A which lies in the loop positioned at the opposite site from the binding site of cysteine proteases. The maximum distance between the two equivalent CA atoms in all molecules of stefins in the complexes with papain-like cysteine proteases is 3.7 Å, however, it was observed between two stefins A in asymmetric unit of cathepsin V – stefin A crystals, therefore the movements of adjacent parts of stefin A are most likely the consequences of crystal packing rather than differences in binding to different cysteine proteases.

The first loop binds into the non-primed (S3-S1), the second and third into the primed substrate binding sites (S1'-S3'). They occlude the catalytic cysteine in the middle and thereby prevent the approach of substrate molecules.

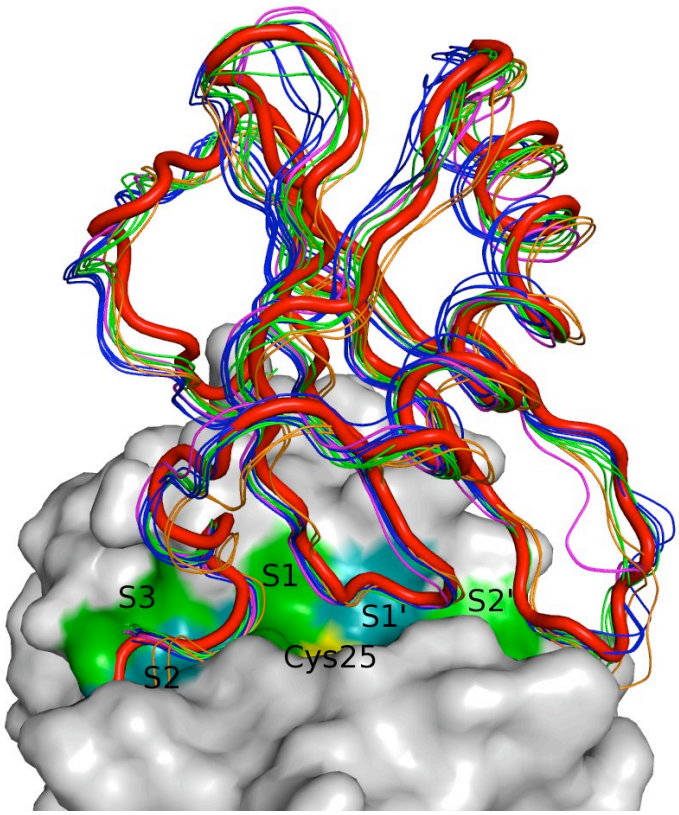


Figure 34: *Binding of N-terminal trunk and two loops of stefin A to cathepsins*. Cathepsin V is shown as gray surface, stefin B in papain – stefin B complex in red, stefin A in cathepsin B – stefin A complex in orange, stefin A in cathepsin L – stefin A complex in magenta, stefin A in reduced cathepsin V – stefin A complex in cyan and stefin A in cathepsin V – stefin A complex in blue. Active site cysteine residue is shown in yellow.

4.5.6 B-factor analysis

The analysis of the B factors of the main chain CA atoms shows that the highest values in all cathepsins are in the short turn (residues 175-180). This is not surprising, as this turn is the place, where native cathepsins are cleaved in light and heavy chain. The flexible loop is pointing away from the cathepsin core, where other proteases in lysosomes can cleave it. The highest B factors in stefin A are in the loops 33-39 and 59-63 exposed to the solvent.

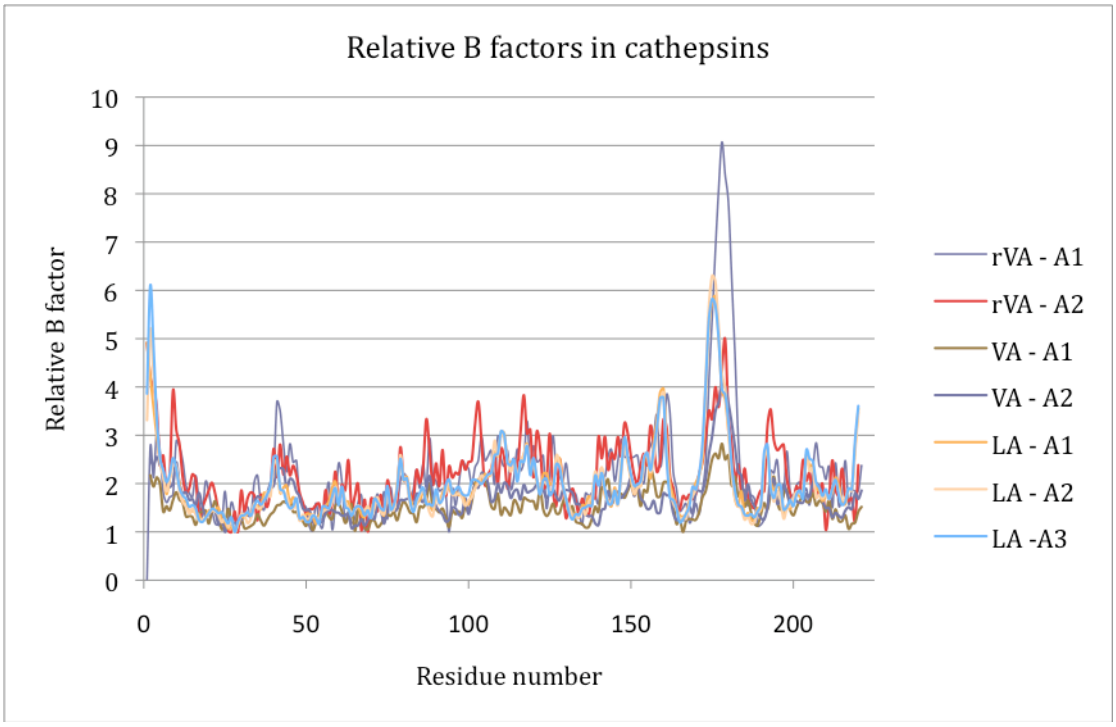


Figure 35: *Relative B factors of CA atom in cathepsins.* B factors were normalised with the smallest B-factor in each molecule.

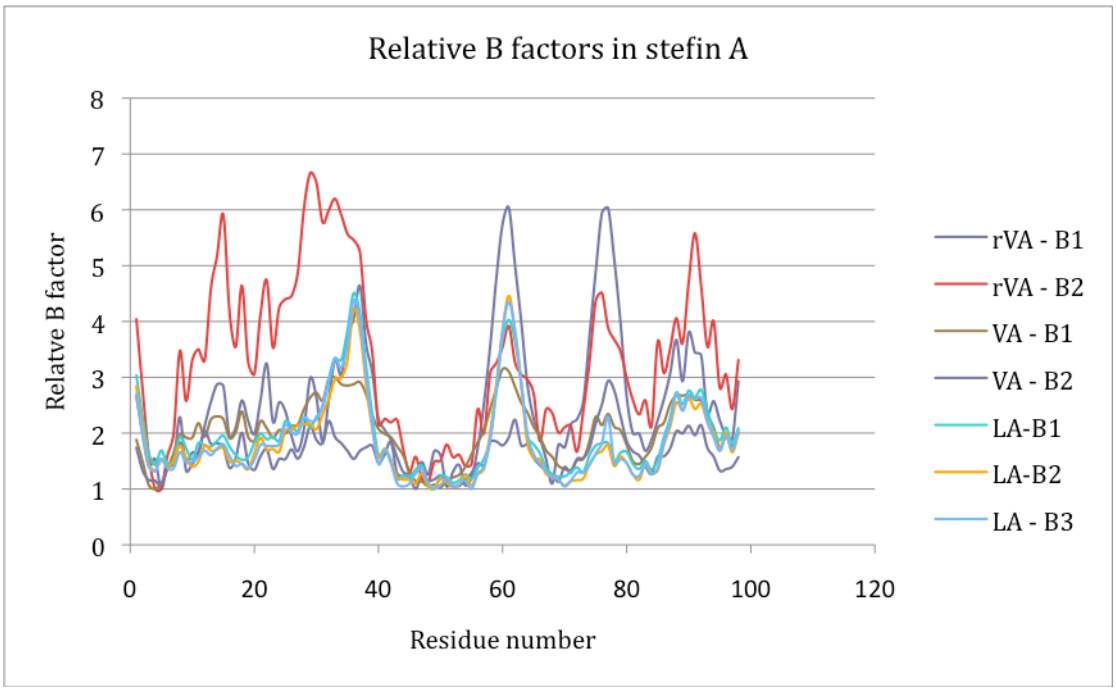


Figure 36: *Relative B factors of CA atom in stefin A.* B factors were normalised with the smallest B-factor in each molecule.

4.6 Structural properties of interaction area between stefin A and cathepsins

4.6.1 Structure of the N-terminal trunk

The N-terminal trunk of stefin A occupies the S3 binding site in cathepsin V with Ile 2 and S2 site with Pro3 (Ser3 in stefin B). The chain continues through the S1 binding site, which is occupied by Gly4.

The N-terminal trunk of stefin A in the complexes with endopeptidases (cathepsins L and V) shows little flexibility (Figure 37A), whereas the N-terminal trunk of the stefin A in complex with cathepsin B shows certain flexibility (Figure 37B). This flexibility of the N-terminal trunk in stefin A – cathepsin B complex is most likely the consequence of weaker binding of stefin A.

Two hydrogen bonds fasten the N-terminal trunk within the active site cleft. The first is formed between the amide proton of Gly4 in stefin A and the carbonyl oxygen of Asp162 (Asp162 in cathepsin L, Glu198 in cathepsin B, Asp158 in papain). The second hydrogen bond is formed with the same peptide bond: the carbonyl oxygen of Pro3 (Ser3 in stefin B) binds to the amide proton of the Gly68 of cathepsin V (Gly68 in cathepsin L, Gly74 in cathepsin B, Gly66 in papain).

The residues Gly3-Pro4 form antiparallel hydrogen bond arrangements with the Gly68 of cathepsin V (Gly68 in cathepsin L, Gly74 in cathepsin B, Gly66 in papain) and Asp168 (Asp162 in cathepsin L, Glu198 in cathepsin B, Asp158 in papain) in a substrate-like manner, firmly anchoring the N-terminal trunk between the two domains of the proteases (Figure 37C).

Ser3 of stefin B makes an additional hydrogen bond (Figure 37D) between its amide proton and carbonyl oxygen of Gly66 (Gly68 in cathepsins L and V, Gly74 in cathepsin B), however, Ser3 is replaced by Pro3 in stefin A. Proline residue lacks the amide proton, therefore the formation of this additional hydrogen bond is not possible in complexes with stefin A.

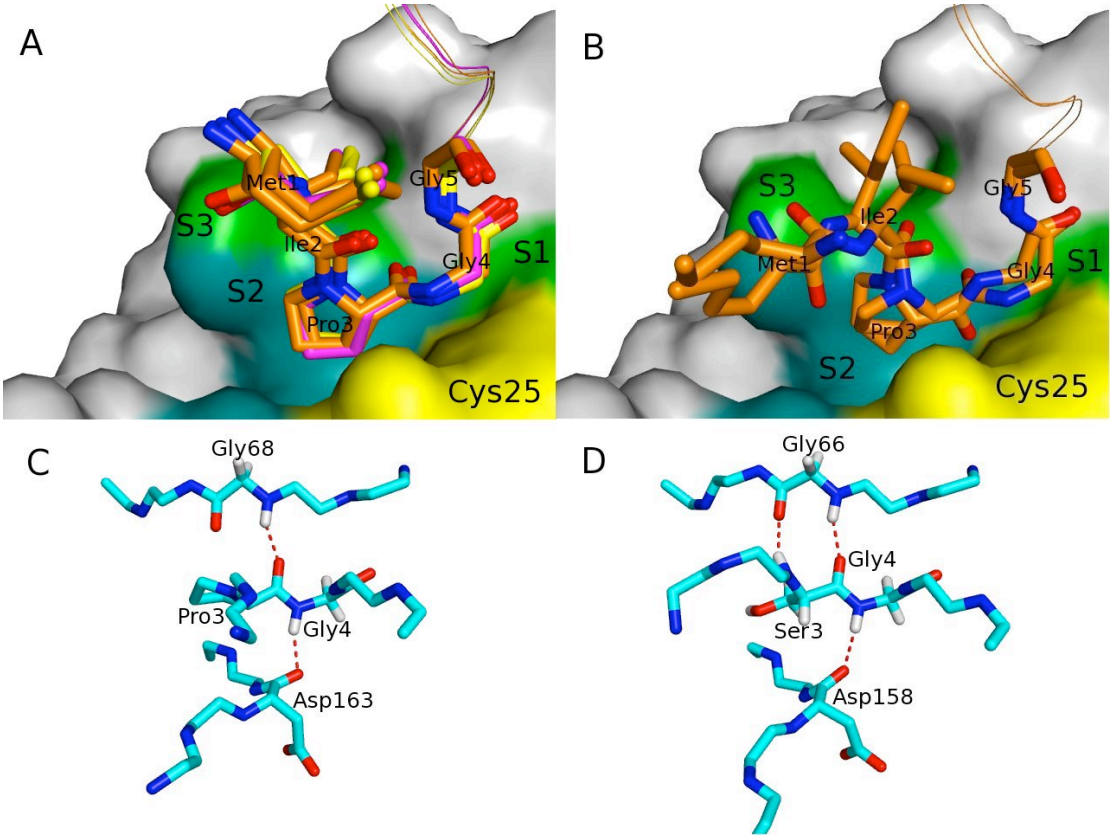


Figure 37: *N-terminal trunk of stefin A in complexes with cathepsin L and V (figure A) and cathepsin B (figure B). Hydrogen bonding of stefin A with cathepsin V (figure C) and stefin B with papain (figure D).* Nitrogen atoms are shown in blue, oxygen in red and hydrogen in white. Figure A and B: orange - stefin A molecules in cathepsin V – stefin A complex, magenta - stefin A molecules in reduced cathepsin V – stefin A complex, yellow - stefin A molecules in cathepsin L – stefin A complex and orange - stefin A molecules in cathepsin B – stefin A complex

Table 8: *Overview of the hydrogen bonds between N-terminal trunk and cysteine proteases and their distances between hydrogen and hydrogen bond acceptor.* First atoms are from stefins and second from cysteine proteases.

	Hydrogen bond 1	Hydrogen bond 2	Hydrogen bond 3
LA	O Pro3 – N Gly68 2.12±0.02 Å	N Gly4 – O Asp162 2.11± 0.04 Å	/
VA	O Pro3 – N Gly68 2.09±0.01 Å	N Gly4 – O Asp163 2.06±0.07 Å	/
rVA	O Pro3 – N Gly68 2.05±0.02 Å	N Gly4 – O Asp163 1.94±0.01 Å	/
BA	O Pro3 – N Gly74 2.25 Å	N Gly4 – O Glu198 2.36±0.40 Å	/
papain	O Ser3 – N Gly66 1.99 Å	N Gly4 – O Asp158 2.28 Å	N Ser3 – O Gly66 2.12 Å

The hydrogen bond length analysis indicates a correlation between the interaction strength and the bond distance. Hydrogen bonds are shorter in the reduced cathepsin V than in the unreduced, which is in agreement with the kinetic data.

It was shown that the sequential deletion of N-terminal residues in stefin A reduces the affinity towards cathepsins and papain [90]. The deletion of the N-terminal methionine in stefin A did not have any effect on the inhibition of papain and cathepsins L and B. These findings can now be confirmed by the crystal structures, where it is clearly seen, that the N-terminal methionine residue does not interact directly with the protease.

The deletion of Met1-Ile2 resulted in 900x, 3x and 200x fold reductions in affinities for papain, cathepsin L and cathepsin B, respectively. The deletion of Ile2 had the highest effect in endopeptidases papain in cathepsin L, but not cathepsin B. As seen in the structures, Ile2 fills the S3 pocket of endopeptidases. The deletion of this residue therefore disturbed the native binding. The small effect in binding to cathepsin B can be explained by the observed flexibility of the N-terminal trunk of stefin A when binding to cathepsin B.

The deletion of further residue (Pro3) resulted in additionally decreased affinities (2000x, 20 000x and 400x fold). The deletion of Pro3 removed the possibility of the hydrogen bond formation thereby strongly affecting the binding affinity.

The important role of Gly4 was confirmed with the point mutation in the native full-length stefin A, where even the smallest replacement, by Ala, resulted in 1000-, 10- and 6000-fold decreased affinities for papain, cathepsin L, and cathepsin B, respectively. Further mutations with larger amino acids additionally reduced the affinities. The highest reduction ($>10^5$x) was observed, when Gly4 was replaced by Arg or Glu [91]. These results can be explained by the flexibility of glycine residue. Glycine can adopt orientations which are unavailable to other amino acids. Suboptimal binding of the larger amino acids most likely alters or disrupts the hydrogen bonding pattern, which then results in the reduced binding affinities.

4.6.2 Structure of the first loop

The first loop with the sequence QVVAG, which is conserved in cystatins [24], docks into the active site cleft near the active site histidine and cysteine residues and leaves it at S2' site. Val47 lies over the catalytic residues; Val48 fills the S1' binding site and Ala49 partially occupies the S2' site. There are three hydrogen bonds which anchor the loop to cathepsins. One is formed between the amide hydrogen of Ala49 in stefin A and the carbonyl oxygen of Lys20 in cathepsin V (Gly20 in cathepsins L and B, Lys20 in papain) and attaches the alanine residue in S2' binding site. The other two hydrogen bonds attach the loop to the L-domain of cathepsins and are positioned on the upper edge of the active site cleft. They are both formed between the side chains of stefin A and the main chain of cathepsins. The first hydrogen bond is formed between the amine hydrogen in Glu46 and the carbonyl oxygen of Cys65 (cathepsins L, V and papain) or

Cys61 (in cathepsin B), whereas the second one is formed between the amine hydrogen in Asn52 in stefin A and Gln21 (in cathepsins L, V and papain) or Ser25 (in cathepsin B).

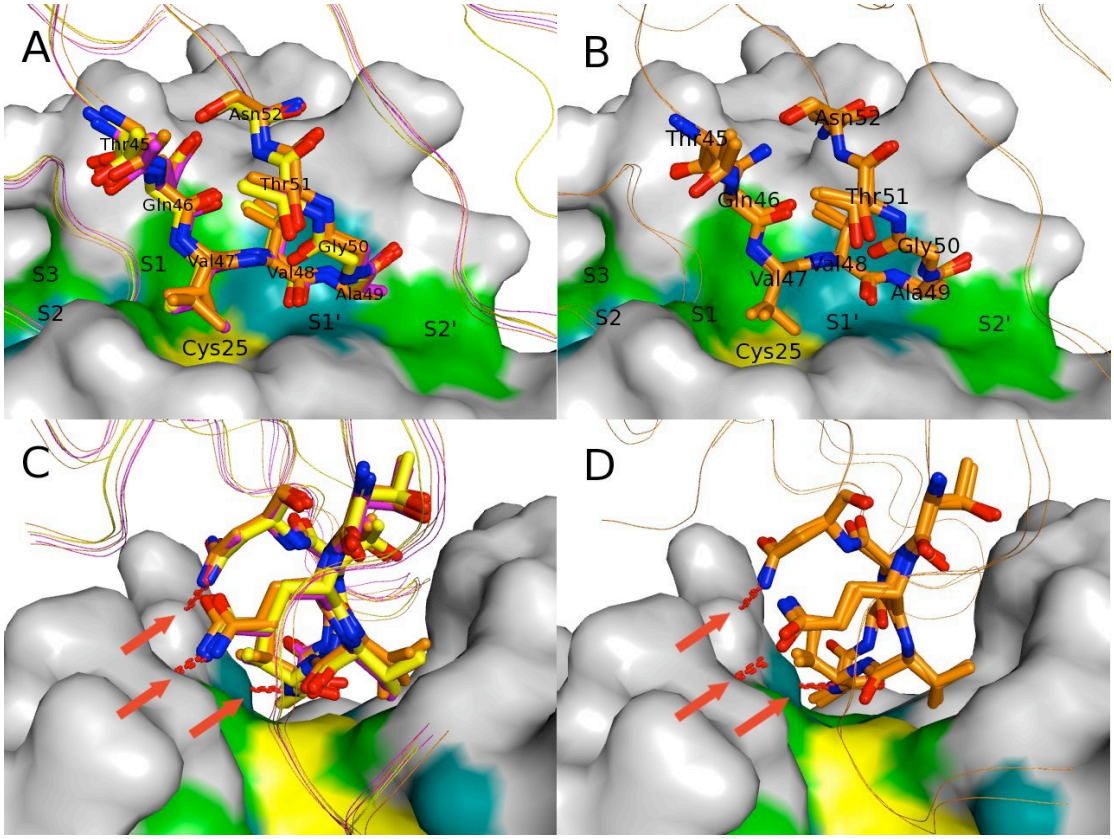


Figure 38: *First loop of stefin A in complexes with cathepsin L and V (figure A and C) and cathepsin B (figure B in D).* Nitrogen atoms are shown in blue, oxygen in red and hydrogen in white. Figure A and B: orange - stefin A molecules in cathepsin V – stefin A complex, magenta - stefin A molecules in reduced cathepsin V – stefin A complex, yellow - stefin A molecules in cathepsin L – stefin A complex and orange - stefin A molecules in cathepsin B – stefin A complex. Red arrows in figure C and D show hydrogen bonds (red dashed lines).

Table 9: *Overview of the hydrogen bonds between first loop and cysteine proteases and their distances between hydrogen and hydrogen bond acceptor.* First atoms are from stefins and second from cysteine proteases.

	Hydrogen bond 1	Hydrogen bond 2	Hydrogen bond 3
LA	NE2 Gln46 – O Cys65 2.11±0.03 Å	N Ala49 – O Gly20 2.05±0.03 Å	ND2 Asn52 – O Gln21 2.16±0.01 Å
VA	NE2 Gln46 – O Cys65 2.08±0.08 Å	N Ala49 – O Lys20 2.11±0.06 Å	ND2 Asn52 – O Gln21 2.34±0.42 Å
rVA	NE2 Gln46 – O Cys65 2.16±0.15	N Ala49 – O Lys20 2.05±0.01 Å	ND2 Asn52 – O Gln21 2.10±0.02 Å
BA	NE2 Gln46 – O Cys71 2.84±0.28 Å	N Ala49 – O Gly20 2.40±0.11 Å	ND2 Asn52 – O Ser25 1.77±0.04 Å
papain	NE2 Gln46 – O Cys65 2.57 Å	N Ala49 – O Lys20 1.83 Å	ND2 Asn52 – O Gln21 1.99 Å

The hydrogen bond length analysis showed no correlation between the interaction strength and the bond distances for the hydrogen bonds, which attach stefin A to the edge of the active site cleft. The correlation is more informative for the hydrogen bond between Ala49 and Lys20 in cathepsin V (Gly20 in cathepsins L and B, Lys20 in papain), This hydrogen bond is the shortest in papain, followed by cathepsin L and the reduced form of cathepsin V. The bond is again the longest in the cathepsin B – stefin A complex.

4.6.3 Structure of the second binding loop

The second binding loop of stefins comes down the S2' binding site and partially fills it with Leu73. It continues through S3' and S4' broad binding areas and leaves the active site cleft.

The N-terminal trunk and the first loop are held in position by hydrogen bonds, whereas in the stabilisation of the lower part of the second loop hydrogen bonds are not present. Therefore it can adopt numerous conformations (shown in Figure 33). This flexibility is a consequence of the wider active site cleft and less defined S3' and S4' binding sites. S2-S2' binding sites lie in a narrower part of the active site and are cleft, whereas the S3' and S4' binding sites are basically broad binding areas [157].

To evaluate the flexibility of different residues in stefin A, we have calculated the maximal distances between the equivalent CA atoms of stefin A molecules from all known complexes. The flexible region starts with Leu73 and Pro74 with the distance of 0.9 Å and 1.4 Å, respectively. The flexibility is the highest in the short stretch Gly75-Glu78 with a distance up to 5.0 Å for Gln78 (between CA atoms in HA and LA) and 5.0 Å for Gly75 (between CA atoms in HA and BA) and ends with Asp79-Leu80 (distance 1.7 Å and 1.1 Å).

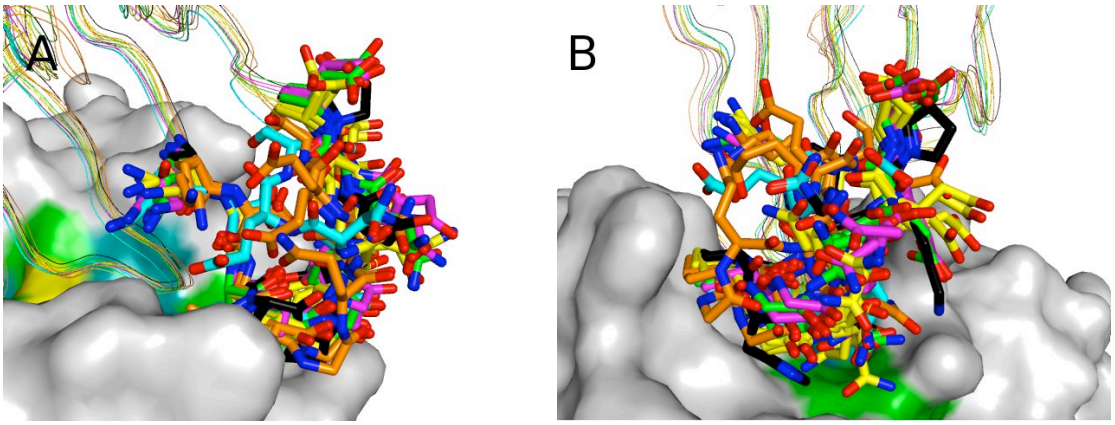


Figure 39: *Second loop of stefins in complexes with cysteine proteases.* Nitrogen atoms are shown in blue, oxygen in red and hydrogen in white. Freen - stefin A molecules in cathepsin V – stefin A complex, magenta - stefin A molecules in reduced cathepsin V – stefin A complex, cyan - stefin A molecules in cathepsin L – stefin A complex, orange - stefin A molecules in cathepsin B – stefin A complex, yellow - stefin A molecules in cathepsin H– stefin A complex and black - stefin B molecules in papain – stefin A complex.

Although the second binding loop is flexible when compared to the N-terminal trunk and the first loop, there is one hydrogen bond in cathepsin L – stefin A. The long and flexible side chain of Glu76 from stefin A provides the amine proton to the carboxyl oxygen atom in the flexible side chain of Glu192 in cathepsin L. This hydrogen bond is present only in two out of three complexes in the asymmetric unit. Their distances are 2.76 Å and 2.07 Å, whereas the distance between those two atoms in the third complex is longer than 4 Å.

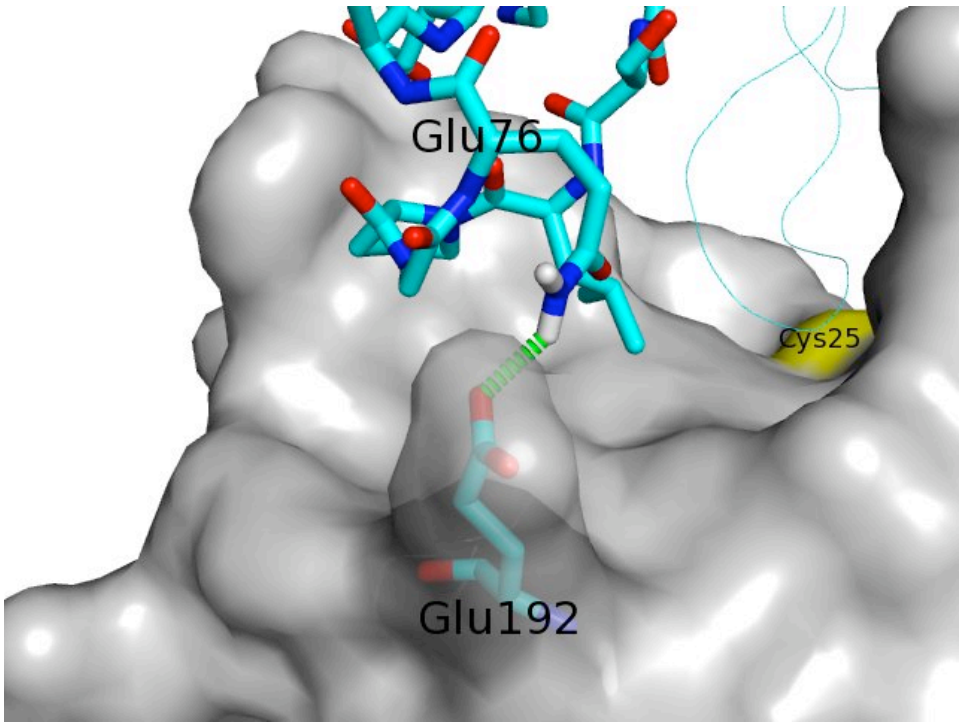


Figure 40: *Hydrogen bond in second loop in cathepsin L – stefin A complex.*

4.6.4 Distance determination

The comparison of stefin A – cathepsin H structure [24] with the structure of stefin B in complex with papain revealed that stefin A binds 0.7 Å deeper into the active site cleft than stefin B does in complex with carboxymethylated papain [89], because the modified active site cysteine residue prevents the native binding. Although cathepsin H was not modified, the presence of the mini chain might have interfered with the binding of stefin A and prevented the tight binding otherwise present in complexes of stefin A and endopeptidases members of the cathepsin family.

In addition, we have calculated the interatomic distances between all CA atoms in stefin A and the CA atoms of the active site cysteine and histidine residues of cathepsin moelcules and averaged them. The average distance between the reduced cathepsin L and stefin A was evaluated on the basis of the differences between the distances of the reduced and unreduced cathepsin V and stefin A. The complex of carboxymethylated cathepsin L and stefin A was modelled on the basis of the structure of carboxymethylated papain in complex with stefin B [89]. The calculated distances are summarized in Table 10.

Table 10: *Distances between stefins and cysteine proteases.* * Distance between unreduced cathepsin L and stefin A was evaluated on the basis of differences between distances of reduced and unreduced cathepsin V and stefin A. ** Complex of carboxymethylated cathepsin L and stefin A (cmLA) was modelled on the basis of the structure of carboxymethylated papain in complex with stefin B [89].

	Distance / Å	Side chain of Cys25
rLA*	22.9	$-CH_2-SH$
LA	23.17±0.03	$-CH_2-S-CH_3$
cmLA**	24.0	$-CH_2-S-CH_2-COOH$
rVA	23.21±0.11	$-CH_2-SH$
VA	23.44±0.20	$-CH_2-S-CH_3$
HA	23.36±0.23	$-CH_2-SH$
BA	23.34±0.15	$-CH_2-SH$
Papain – stefin B	23.93	$-CH_2-S-CH_2-COOH$

As expected, the distance analysis showed the correlation between the distances and the inhibition constant and disrupted binding, when additional features (like occluding loop and mini chain) on the

surface of the enzymes are present or active site cysteine residue is modified.

The tightest binding was found in cathepsin L and the reduced cathepsin V (23.2 Å). The binding was weaker with the unreduced form of cathepsin V, where stefin A bound 0.2 Å shallower, because the modified active site cysteine residues prevented the tighter binding. Stefin A approaches to cathepsins H and B at similar distances as to unreduced cathepsin V. The shallowest binding was observed in stefin B – papain complex [89], where the cysteine residue was modified with the largest fragment.

4.7 AFM measurements

4.7.1 Surface preparation

The used Si_3N_4 waffers were assessed by contact-mode AFM measurements, which revealed the surface roughness of app. 5 nm over hundreds of nanometers. The high roughness of the Si_3N_4 prevented the use of AFM to determine the surface density of the bound proteins, which are small compared to the roughness (stefin A ~2 nm and cathepsin L ~3 nm).

For the covalent attachment of proteins two approaches were used. Stefin A was attached to the surface using NHS-PEG-malimide linker, whereas NHS-PEG-aldehyde linker was used for cathepsin L. Figure 41 summarizes the immobilization process.

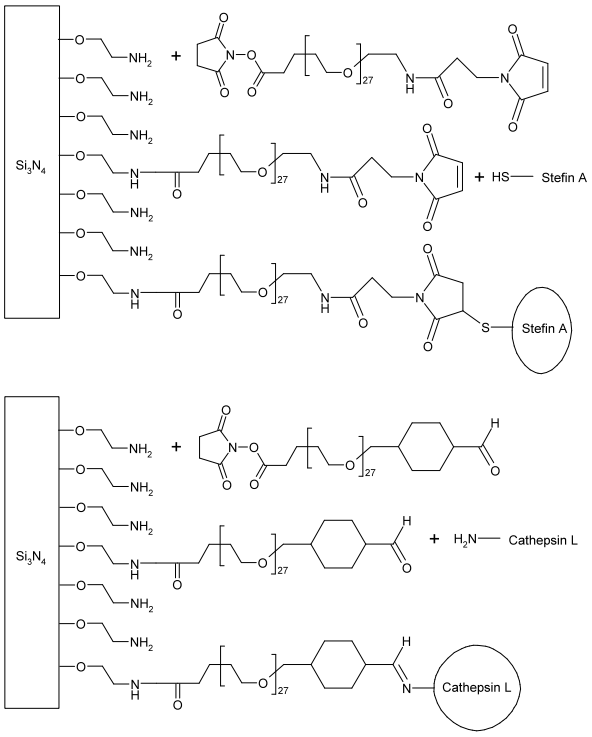


Figure 41: *Immobilization of cathepsin L and stefin A.*

4.7.2 The enzyme surface density determination

The surface density measurements were performed with three different chips to which cathepsin L was covalently attached via a flexible PEG linker.

The surface density of the bound enzymes was determined by the classical biochemical activity test after each step of washing. Cathepsin L modified chips were immersed into a cuvette with a fluorogenic substrate z-Phe-Arg-AMC. The change in fluorescence was followed for 5 minutes and the rate of reaction was compared with the calibration series in which known amounts of enzyme in solution were added to the reaction mixture. After 5 minutes, the chip was removed from the cuvette and measurements were continued for an additional minute. The washed amount of washed and immobilized enzyme was determined using the calibration series.

As a negative control a blank chip was used, where aldehyde groups were blocked with ethanoleamine prior to the cathepsin L addition.

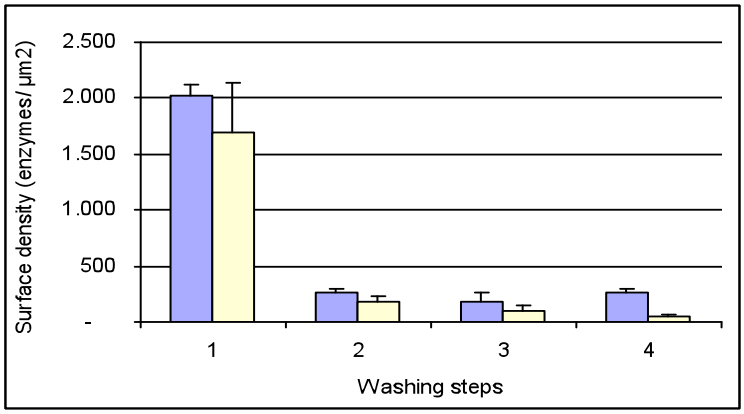


Figure 42: *Surface density determination.* Covalently attached cathepsin L surface density of the reactive (blue) and blocked surface (yellow). Before the first step of washing, the surface density was similar for all chips (app. 2000 enzymes per μm^2). The majority of the unspecific bound enzymes were washed away in the first step. After 4 steps of washing only covalently attached enzymes remained on the surface.

The relatively low surface density of cathepsin L can be partially explained with the pH induced denaturation of cathepsin L. Pure cathepsin L has a lifetime of a few minutes at pH=7.0. Insteed of the free enzyme, the complex of cathepsin L and stefin A was used to dimish the pH-induced denaturation. Cathepsin L in the complex with stefin A has a longer halftime (~ 1 h). The drawback of this approach is the presence of the additional lysines from stefin A molecules, which can react with the aldehyde linker on the surface.

4.7.3 Force measurements

A typical force curve recorded during the measurements is shown in Figure 43A. Identical curves were recorded on the different sites of the sample surface with different tips. The blue retrace curve displays the characteristic unbinding event. The observed nonlinear behaviour, which was used for a discrimination of a specific binding from the unspecific adhesion (shown in Figure 43B), is a consequence of a flexible PEG linker.

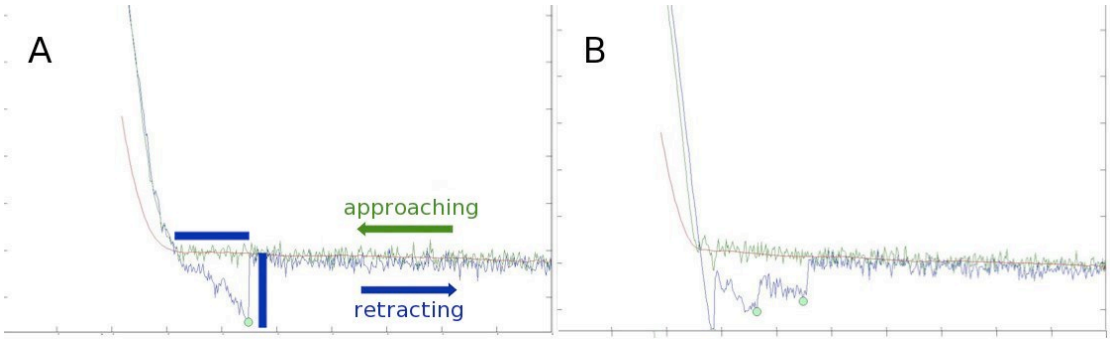


Figure 43: *Typical force curve with specific (A) and unspecific (B) binding.* Blue horizontal bar represents unbinding length, whereas the vertical bar represents unbinding force.

The measurements were performed in sets containing 1000 force curves, where each set was recorded at the same loading rate. The used cantilevers had a nominal spring constant of 0.01, 0.02 and 0,03 N/m. The retrace velocity ranged from 100 nm/s to 2000 nm/s, what resulted in loading rates from 1 nN/s to 50 nN/s.

The specificity of the unbinding event was verified with the blocking experiment. Cathepsin L was added in to the liquid cell solution to the final concentration of 90 µM. The solution was incubated for 10 minutes, so the stefin A molecules bound to the tip were blocked with cathepsin L and the measurements were repeated. Typically, the binding probability dropped from 5 to 10 fold.

Approximately 200 sets were measured and manually analyzed for the presence of specific unbinding events. Only curves with no or low unspecific adhesion were selected for the statistic. In the case of the multiple unbinding events only the last one was used. Approximately 40 sets showed binding probability

higher than 10%, additional 40 sets showed probability between 6 and 10%. The majority of sets showed no or very low binding, probably due to the presence of the free cathepsin L in liquid cell solution.

The unbinding force and lenght were measured from the pull off jump of the retrace curve. The statistical distribution of the force and lenght is shown in figure 42.

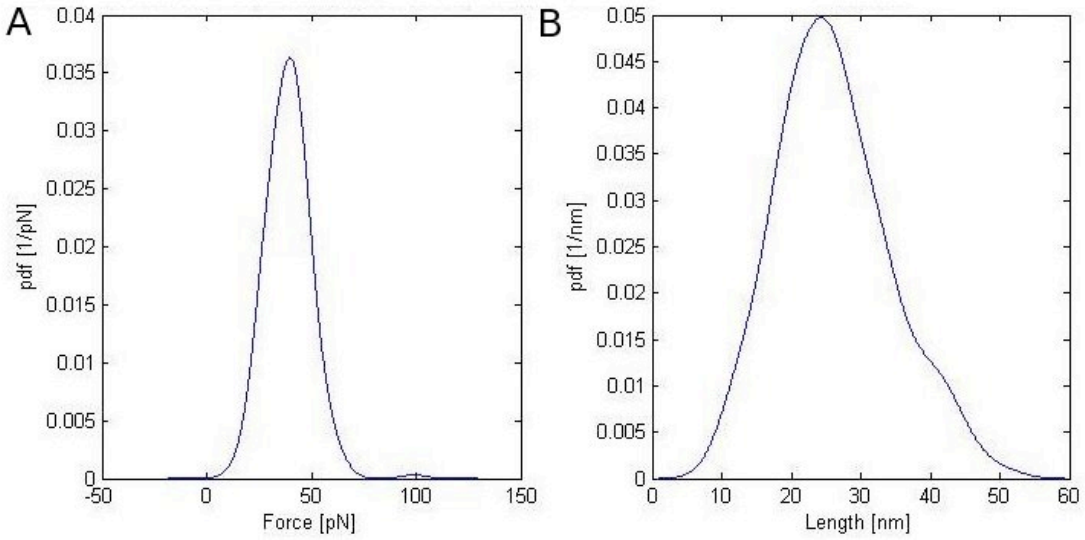


Figure 44: *Statistical distribution of unbinding forces (A) and length (B) for one set of measurements (1000 curves).* A) The most probable unbinding force, estimated from the force distribution, is 38.7 pN with the standard error of 0.8 pN. Binding probability in this set was 21.2%. B) Statistical distribution of unbinding lenght for the same set. Experimentally determined unbinding length is 25 nm, what is in agreement with the experimental setup and previously publish results with similar linkers [143].

4.7.4 Analysis of the data

The unbinding kinetics cannot be described with a classical biochemical equation. Commonly accepted is the Bell's model which explains the unbinding process. It treats it as an escape of a ligand from a potential well [141, 142]. When the most probable unbinding force F^* is plotted as a function of the loading rate v, one can determine the distance between the potential minimum and the transition state x_β and more important, the natural thermal off-rate for the dissociation at zero force $k_{off}(0)$.

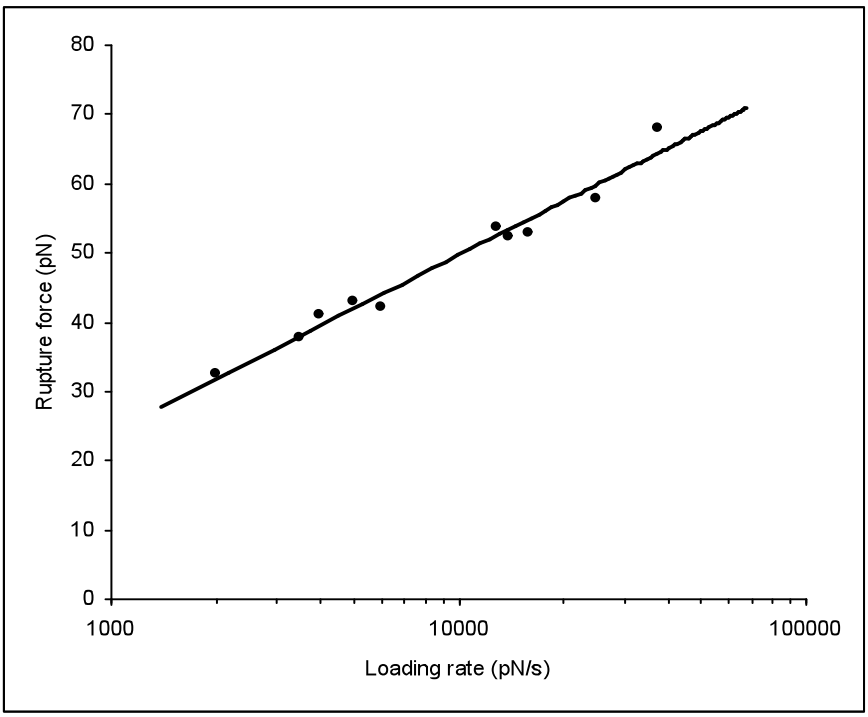


Figure 45: *Loading rate dependence of the unbinding force.* The solid line is the numerical fit of experimental data to the Bell model. Each experimental point is an average of at least 200 and at most 1000 unbinding events.

When fitted to equation 2, our experimental data show a good agreement with the Bell's model. The linear increase of the most probable unbinding force over the range of almost two orders or magnitude confirms the theoretical predictions.

The distance between the potential minimum and the transition state x_β was estimated to 8.5 Å. This value is in the range of previously published data, which ranges from 0.3 to 30 Å for biotin [158], 3.0 Å for N-cadherin dimers, 7.7 nm for E-cadherin [159] and 5.9 nm for cadherin – anti-cadherin [138].

The AFM determined natural thermal off-rate for the dissociation at zero force $k_{off}(0)$ is 0.50 s^{-1}, compared to the kinetic measurements in unstressed system, where koff is app. 10^{-3} s^{-1}. The difference is a consequence of measurements in different regimes. Similar results were obtained earlier, where off-rates for the avidin-biotin complex were 2,9 s^{-1} in AFM measurements with a similar loading rate range (10-60 nN/s), whereas the classical value is app. 10^{-7} s^{-1} [142].

The unbinding force at a loading rate of 10 nN/s is 45 pN for stefin A – cathepsin L complex. As reported earlier and summarized in table 19, values of other complexes are varying between 20 in 240 pN.

Table 11: *Unbinding forces for various complexes, reviewed in [138].*

Complex	Unbinding force (pN)
α5β1 intergrin ligand/GRGDSP peptide receptor	20±7
Cadherin/cadherin	35±16
Cathepsin L/stefin A	45±10
Selectin/sialyl Lewis X tetrasaccharide	50±15
Lysozyme/anti-lysozyme	52±18
Regulatory protein His6ExpG/DNA promoter regions	75±15
Biotin/avidin	80±15
AZ/C551	95±15
ICAM-1/anti-ICAM-1	100±50
P-selectin/ligand	115±40
Anti-Sendai-antibody/Sendai bacteriorhodopsin	126±15
Single-chain Fv fragment of fluorescein binding antibody/fluoroscein antigen	160±15
HSA/anti-HSA	244±22

4.8 Inhibition mechanism of mycocypins

4.8.1 Structure of macrocypin

Macrocypin 1 crystallized in $P3_12_1$ space group with one molecule in the asymmetric unit. Macrocypin crystals contain complete sequence of 168 residues, numbered from Gly1 to Glu168. Positioning of nearly all residues is clearly revealed by the electron density maps. The exceptions are the side chains of His114, Tyr140 and Lys21, which are only partially defined, and the stretch of residues Ser20, Lys21, Ile22, which is only loosely defined. Nine residues (Gly1, His17, Arg55, Ile75, Gln78, Ser80, Glu100, Gln110, Ile158) were modelled in alternative conformations [146].

4.8.2 Structure of clitocypin

Clitocypin isolated from a natural source contains a large number of isoforms in unknown ratio [116]. It crystallized in $P2_1$ space group with two molecules in the asymmetric unit. The positioning of nearly all residues is clearly revealed by the electron density maps. The exceptions are the loop Gln67-Tyr75 in molecule 1, Gly68-Asn70 and the side chain of Gln115 in molecule 2 and the first two N-terminal residues in both molecules. As default we have used the sequence of the clitocypin isoform used for the complex formation. However, when the electron density unambiguously showed disagreement with that sequence, we built an appropriate amino acid residue from an alternative sequence (H17S, Y62S, L82M, P84Q, I88M, A105T, T139N in both molecules, S46F, Q48R in molecule 1 and Q37K, N42S, A57S in molecule 2).

The structure of both clitocypin molecules in the asymmetric unit is basically the same (RMSD value of 0.28 Å) with one important exception. The peptide bond between Gly24 and Gly25 residues appears in two different orientations, both in glycine-preferred regions of the Ramachandran plot, which was clearly seen in the electron density. This suggests that this peptide bond is flexible and can exist in either orientation [146].

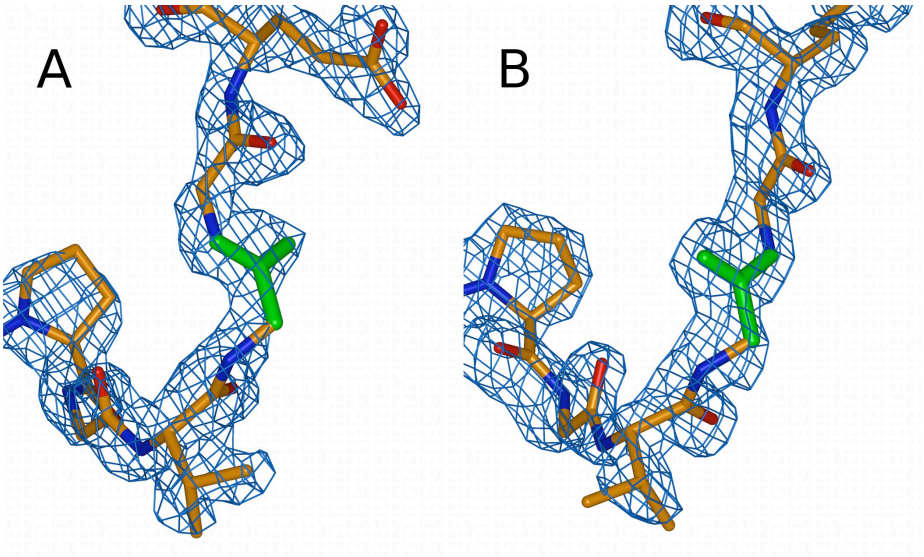


Figure 46: *Orientation of the Gly24-Gly25 peptide bond in both clitocypin molecules in the asymmetric unit.* Electron density is contoured at 1 σ. Glycine 24 is shown in green.

4.8.3 Description of mycocypins fold

The macrocypin and clitocypin structures share the same fold. In the projection used in Figure 47 the fold is a reminiscent of a tree with roots, a short and thick trunk and a crown with branches expanding far from the centre. The trunk part is an up-and-down β-barrel composed of six antiparallel β-strands (β1, β4, β5, β8, β9, β12). The strands are laid at an angle of more than 45 degrees to the axis of the barrel. The N- and C-termini are at the bottom. They form the roots of the tree together with the loops connecting strands β4-β5 and β8-β9. On the top, three long regions between the strands β1-β4, β5-β8 and β9-β12 constitute the tree crown. Each contains a pair of antiparallel β-strands. In this manner two additional loops are formed between the strands from the crown and the trunk, adding another layer of loops, which spread away from the trunk. Hence, the three-fold arrangement of loops and strands is preserved on four layers of the structure in the roots, trunk and lower and upper crown. The loop region before strand β8 in the lower crown layer of macrocypin is folded into a short three-turn α-helix, whereas in clitocypin the loop preceding the strand β4 contains a short helical region. Although macrocypin and clitocypin share the same fold, the RMSD deviation between 116 equipositioned CA atoms is 1.75 Å. The β-barrels are more similar, yielding RMSD of 0.67 Å between 31 equipositioned CA atoms. Macrocypin and clitocypin have pseudo-threefold symmetry (Figure 47) with the threefold rotational axis running through the six-stranded barrel.

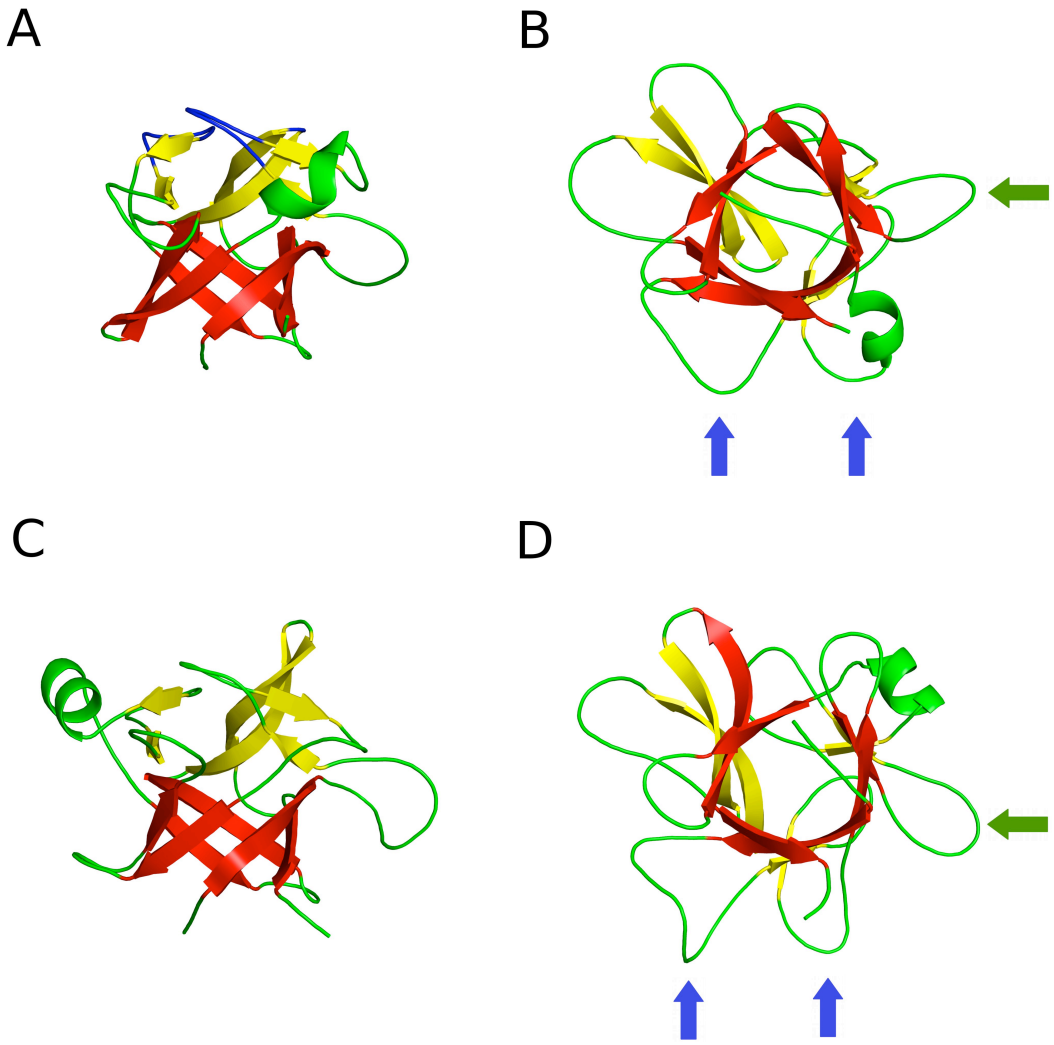


Figure 47: *Structures of clitocypin (A, B) and macrocypin (C, D).* Trunk strands are always shown in red and crown strands in yellow. A. The first layer of loops is shown in green and the second in blue. A) and C) in orientation of a tree, B) and D) the view along barrel. Binding loops for cathepsins are marked with blue arrows and the legumain loop with green arrows.

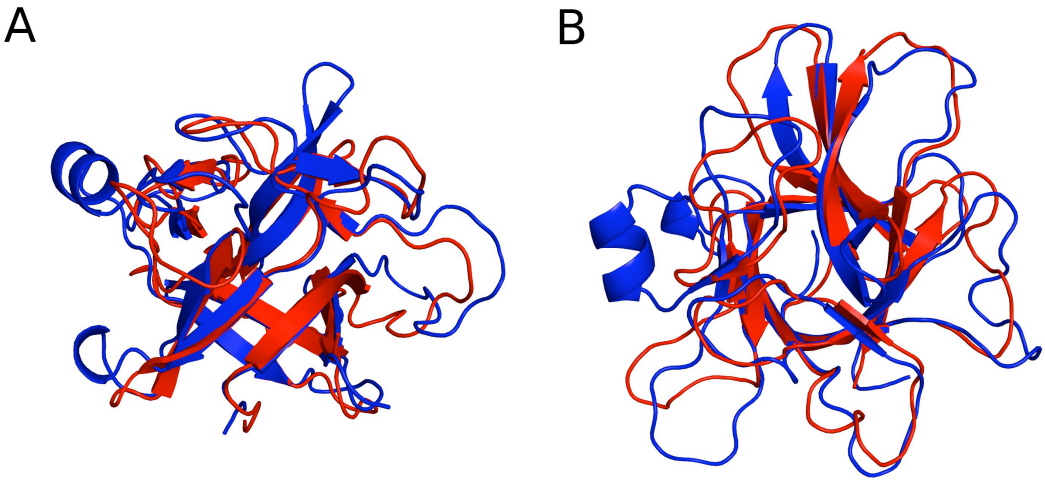


Figure 48: *Structural alignment of macrocypin (blue) and clitocypin (red) in two different orientations (A, B).*

4.8.4 Comparison of mycocypins with other β-trefoil proteins

When the macrocypin structure was submitted to the protein structure comparison service SSM [160] at European Bioinformatics Institute, the fold was identified as β-trefoil fold shared by proteins such as Kunitz type soybean trypsin inhibitor (STI) [74], inhibitor isolated from *Erythrina caffra* [75], interleukins-1α and 1β [76] and fibroblast growth factors [77]. The best structural alignment was found with human acidic fibroblast growth factor (PDB entry 1z4s), where 114 residues were aligned with RMSD of 1.94 Å [161].

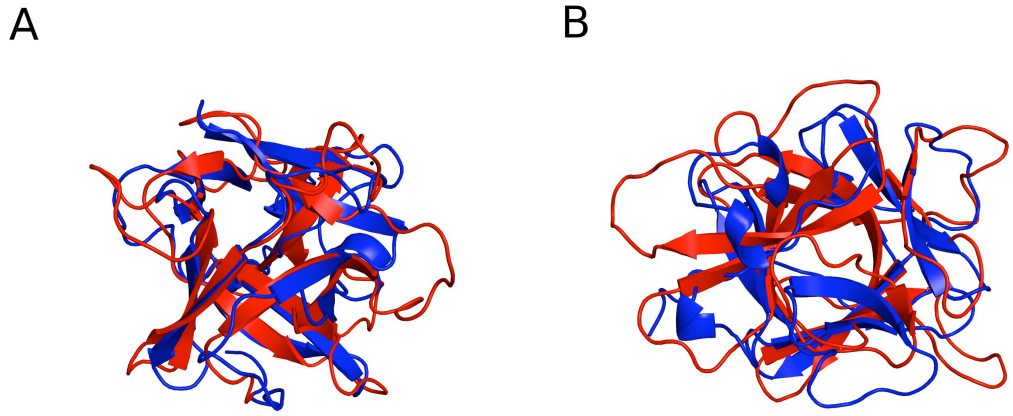


Figure 49: *Structural alignment of clitocypin (red) with human acidic fibroblast growth factor (PDB entry 1z4s) [161] in two different orientations (A, B).*

Sequence alignment shows low similarity of these proteins, in contrast to structural alignment (Figure 50), which reveals that the secondary structure patterns are quite similar. The number of β-strands in macrocypin (12 strands) and clitocypin is equal to those in STI, whereas their lengths differ significantly. STI has only 2 short strands (β6 and β7), whereas macrocypin and clitocypin have 4 (β2, β3, β6, β7). The highest structural as well as sequence similarity is in the regions building the beta barrel. Since these are rather short stretches of sequence this explains why homology searches based on sequence alignment have failed [146].

```
                10        20        30        40        50        60        70        80
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Clt     --ASL-------ED---GTYRLRAVTTHNPDPGV--GGEYAT -EG----ARQPVKAEPSTP-PFFERQIWQVTRNSD-------
Mcp     ---GF-------ED---GFYTILHLAEGQHPNSKIPGGMYASSKDG----KDVPVTAEP-LGPQS--KIRWWIARDPQAG----D
STI     --DFVLDNEGNPLEN-GGTYYILS ------IT---AFGGIRAAPTGNERCPLTVVQSR-NE-LDK-GIGTIISSPYRIRFIAEG
BbCI    HMSVILDTKGEPVSNAADAYYLVPV------S----GEGGLALAKVGNEAEPKAVVLDP-H---HRPGLTVRFETPLAIAIITES
ECI     ---VLLDGNGEVVQN-GGTYYLLPQVW---AQ-----GGGVQLAKTGEETCPLTVVQSP-NE-LSD-GKPIRIESRLRSAFIPDD
WBCI    -DDDLVDAEGNLVEN-GGTYYLLPH------IWA--HGGGIETAKTGNEPCPLTVVRSP-NE-VSK-GEPIRISSQFLSLFIPRG
                          ß1                  ß2        ß3              ß4

            90        100       110       120       130       140       150       160       180
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Clt     GQYTIKYQGLNA------PFEYGFSYD--QLE-QNAPVIAG--DP---------KEYILQLV-PS-TA-DVYIIRAPI------QR
McP     DMYTITEF-RID-----NSIPGQWSRSPVET---EVPVYLY--DRIKAEETGYTCAWRIQPADH--GADGVYHIVGNV------R-
STI     HPLSLKFD-SFAVIMLCVGIPTEWSVV--EDLPEGPAVKIG--ENKDA----MDGWFRLERV-SE-FN--NYKLVFCPQ----QD-
BbCI    FFLNIKFV-PSS-----SDSE-VWDVSKQ--YPIGLAVKVT--DTKS-----FVGPFRVEKE-G--E---GYKIVYYPDR---GQ-
ECI     DKVRIGFA-YAPK----CAPSPWWTVV--E----GLSVKLSEDESTQ-----FDYPFKFEQV-SDQLH--SYKLLYCEGK---HE-
WBCI    SLVALGFA-NPPS----CAASPWWTVVDS--P-QGPAVKLSQQKLPEK----DILVFKFEKV-SH-SNIHVYKLLYCQHDEEDVK-
              ß5                  ß6        ß7                  ß8        ß9

                190       200       210       220
            ....|....|....|....|....|....|....|....|
Clt         VG--VDVEVGVQ ------NNLVYKFFPVDGSGGDRP-AWRFTRE----
Mcp         --IGSTDWADLREEY-- EPQVYMK-PVPVIPNVYIP-RWFILGYEE--
STI         --K--CGDIGISIDHDDGTRRLVVS-K-------NKPLVVQFQKLD---
```

```
BbCI        --T--GLDIGLVHRN--DKYYLAAT-E-------GEPFVFKIRKAT---
ECI         --K--CASIGINRDQK-GYRRLVVT-E-------DYPLTVVLKKDE---
WBCI        --C--DQYIGIHRDRN-GNRRLVVT-E-------ENPLELVLLKAKSET
                   ß10          ß11                ß12
```

Figure 50: *Structural alignment of clitocypin, macrocypin and selected β-trefoil serine protease inhibitors.* Strands are shown with white text on black background, binding loops with white text on gray background and helixes with black text on gray background. STI – soybean trypsin inhibitor [74], BbCI - *Bauhinia bauhinioides* cruzipain inhibitor[80], ECI - inhibitor isolated from *Erythrina caffra* [75], WBCI-winged bean chymotrypsin inhibitor [162]. Alignment was done with Protein structure comparison service SSM at European Bioinformatics Institute [160].

4.8.5 Structures of the cathepsin V – clitocypin complex

Crystals of the complex of clitocypin with cathepsin V contain the complete protein sequences. The catalytic site of cathepsin V was blocked with a methyl methanethiosulfonate, leaving the $–S-CH_3$ on the active site cysteine residue. This form of cathepsin V is much more stable and resulted in better diffracting crystals. The crystals have $P2_12_12$ space group and contain four pairs of molecules per asymmetric unit. The four structures of cathepsin V and clitocypin are closely related. They exhibit RMSD over equivalent CA atoms in the ranges of 0.18 Å to 0.25 Å for cathepsin V and 0.33 Å to 0.50 Å for clitocypin. While the cathepsin V structures, apart from the ends of a few side chains, are unambiguously resolved by the electron density maps, three N-terminal residues and two loop regions (Gln67 – Asn70 in molecules 1, 3, 4 and Asp138 - Gly141 in molecules 2, 3, 4) of clitocypin lack adequate electron density or are only loosely defined [146].

Clitocypin binds into the active site of the target protease in the orientation of a fallen tree with trunk and roots pointing sideways and up (Figure 51). The wedge shaped structure fills the active site cleft along the whole length. The binding results in a area of 825 $Å^2$. The interaction surface of clitocypin is constituted basically by two broad loop regions positioned at the lower edge of the crown. The loop structure and the binding geometry are stabilized by numerous hydrogen bonds. Both loops originate from the first third of the clitocypin sequence – the first loop connects strands β1 and β2, and the second loop strands β3 and β4 [146].

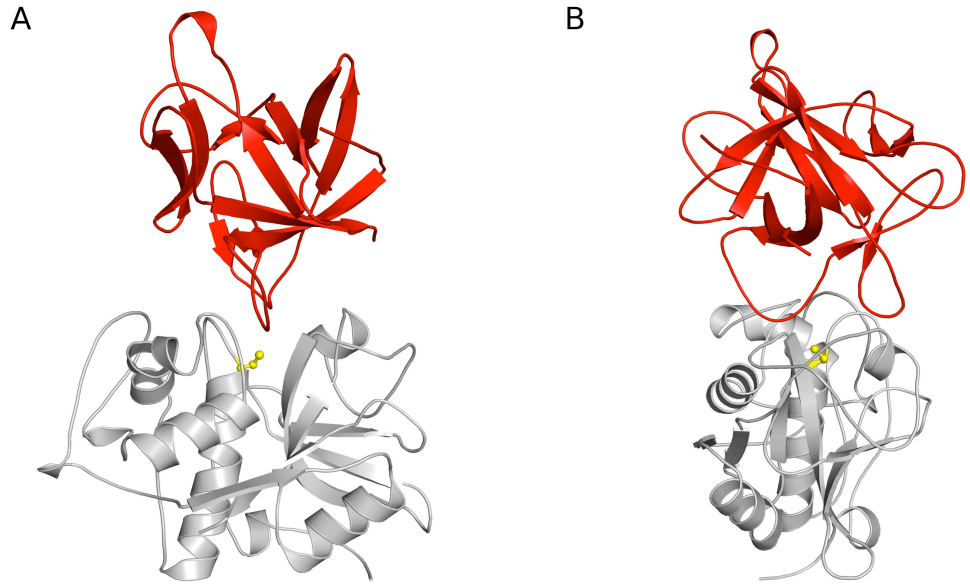


Figure 51: *Structure of the cathepsin V - clitocypin complex.* A. View along the active site cleft and B. view perpendicular to the active site cleft. Cathepsin V is shown in gray and clitocypin in red. The catalytic cysteine is shown in yellow.

The first loop is a broad lasso like structure cross-connected in the middle with a hydrogen bond between the Arg12 side chain and Gly22 carbonyl. The second loop region is narrower and contains a short

helix. The first loop binds into the non-primed and the second into the primed substrate binding sites (Fig. 4). They occlude the catalytic cysteine in the middle and thereby prevent the approach of substrate molecules. Because the reactive site Cys25 in the structure is modified we cannot exclude that interaction of the clitocypin with a naked cysteine cathepsin may deviate slightly from the observed one [146].

The first loop chain comes down the S3 binding area (Pro21, Gly22) of cathepsin V, occupies the S2 binding site with Val23 and continues through the S1 binding site upwards (away from the cathepsin V surface). A hydrogen bond from the side chain amide Asn18 attaches clitocypin to the Gln63 side chain of cathepsin V. A water molecule further mediates the contacts in the S3 binding area. Arg12 plays a dual role; it stabilizes the lasso as well as attaching it to the cathepsin V surface by the Asn66 main chain carbonyl – filling the interaction surface between S3 binding area and S1 binding site. Positioning of Val23 builds antiparallel hydrogen bond arrangements with the Gly68 of cathepsin V in a substrate like manner. The peptide bond Val23-Gly24 is additionally fixed by the interaction between its hydrogen and the carbonyl of Asp163 from cathepsin V. In the S1 binding site Gly23 of cathepsin V is involved in two hydrogen bonds formed between its carbonyl and amide hydrogen with the amide Gly25 and carboxylic oxygen atom of Glu26. The peptide bond between Gly24 and Gly25 is flipped when compared to the structure of the molecule 2 of clitocypin and macrocypin, in order to form a hydrogen bond with Gly23 of cathepsin V (Fig. 3). The other carboxylic oxygen atom of Glu26 interacts with the cathepsin V amide of the Asn66 side chain. However, the possibility that Glu26 is partially neutral and interacts with the carbonyl oxygen atom of Gln21 cannot be excluded. The binding of the first loop is further stabilized by the side chain amide group from Asn18, which forms a hydrogen bond with the peptide bond carbonyl atom of Gln21 of cathepsin V. It is notable that, of the listed hydrogen bond interactions between the enzyme and inhibitor, most are contributed by main chain atoms, at least on one side [146].

The second binding loop of clitocypin approaches the S1' and S2' binding sites of cathepsin V from the top. A single hydrogen bond between the carbonyl of Ser42 and the side chain amide of Gln145 of cathepsin V fastens the loop to the cathepsin V surface. In the first complex an additional hydrogen bond is formed between the carboxylic group of Glu48 and cathepsin V Ser142. A layer of solvent molecules mediates the contacts between the N-terminal bottom of the short helix and cathepsin V.

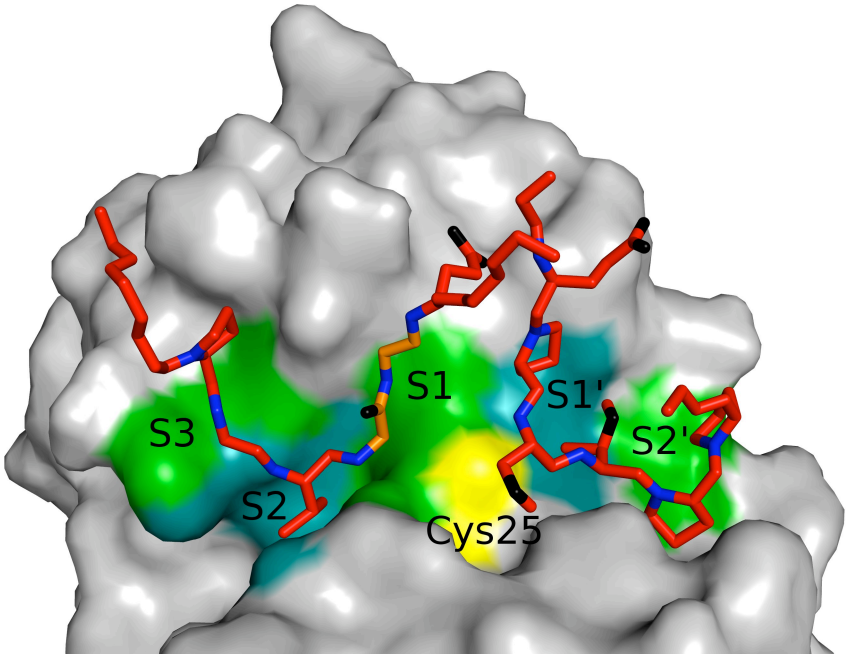


Figure 52: *Binding of clitocypin to cathepsin V.* Clitocypin loops are shown in red sticks and Gly24-Gly25 in orange. Cathepsin V is shown as a gray surface with catalytic cysteine yellow. S3, S2, S1, S1' and S2' binding site are shown in green and cyan.

When we modelled the complex between macrocypin and cathepsin V (by superimposing macrocypin on the clitocypin structure in the complex) it became evident that the binding loops do not fit into the active site.

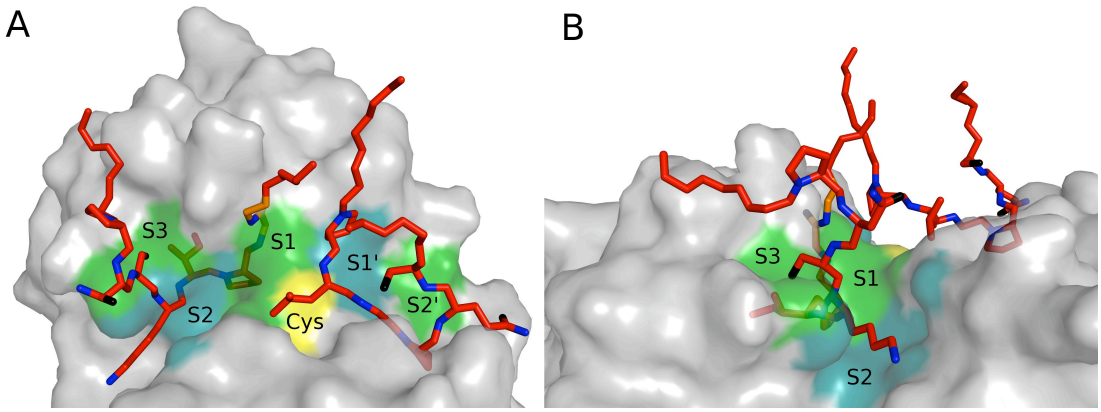


Figure 53: *Macrocypin binding to cathepsin V by superimposing macrocypin on the clitocypin structure in the complex.* Macrocypin loops are shown in red sticks and Gly24-Gly25 in orange. Cathepsin V is shown as a gray surface with catalytic cysteine yellow. S3, S2, S1, S1' and S2' binding site are shown in green and cyan.

To find out whether the binding loops are the same we have expressed four mutants, where Gly24 in clitocypin and Gly25 in macrocypin were either replaced by alanine or deleted. By these mutants we have also tried to assess the relevance of the Gly24-Gly25 peptide flip. We have assumed that the mutation to alanine as well as deletion of the residue will reduce the flexibility of the main chain.

The resulting clitocypin mutants yielded K_i values to cathepsin V that were 20 times higher than that of the native variant. It is notable that the major source of difference is the slower association, while dissociation was not significantly affected (Table 12). This suggests a mechanism, which is in agreement with the peptide bond flip occurring prior to or concurrently with the inhibitor docking. Since macrocypin mutants exhibit equivalent effects on their K_i constants, this confirms that the loops that bind into the active site of cysteine cathepsin are equivalent in clitocypin and macrocypin. This implies that the binding loops of macrocypin exhibit substantial conformational flexibility during binding into the active site of their target enzymes [146].

Table 12: *Inhibition constants of cathepsin V by macrocypin, clitocypin and their mutants.* Kinetic data for interaction of macrocypin 1 and cathepsin V was reported previously [85].

	10^{-6} k_{ass}	10^{4} k_{off}	K_i
	($M^{-1}s^{-1}$)	(s^{-1})	(nM)
Clt	1.26±0.07	1.61±0.65	0.08±0.03
Clt ΔG24	0.10±0.06	1.8±0.9	1.9±0.9
Clt G24A	0.08±0.01	1.2±0.6	1.6±0.7
Mcp1	1.48±0.01	10.3±0.7	0.69±0.06
Mcp1 G25A	0.13±0.02	16.0±5.0	12.5±5.2
Mcp1 ΔG25	0.15±0.02	12.4±2.4	8.5±2.1

4.8.6 Inhibition of AEP

AEP is inhibited by natural and recombinant clitocypin, natural macrocypin and some isoforms of expressed macrocypin (macrocypins 1 and 3) in the low nanomolar range (3-10 nM). Interestingly, macrocypin 4 does not inhibit AEP at all.

Inspection of the aligned sequences of these isoforms in their surface regions has focused attention on the ß5-ß6 loop (residues 71-76 containing sequence Ile-Asp-Asn-Ser-Ile). This part of the sequence is similar to the consensus sequence S/T-N-D/S-M/I found in three inhibitory cystatins C, E and F (Table 4), which bind to AEP in the nanomolar range [47]. Interestingly, in macrocypin 4 the residue at position 72 is Lys in contrast to equipositioned Asn in macrocypins 1 and 3.

Table 13: *Alignment of loops of macrocypin isoforms, clitocypin, cystatins V, E and F, STI and BbCI involved in legumain and trypsin inhibition*. Residues that bind to the P1 pocket are marked with asterisk. Mutated residues are shown in red.

*

Clt	56-YTIKYQGLNAPFEYG-75
Mcp1	61-TITEFRIDNSIPGQW-80
Mcp3	59-TITEIRDDNCIPGQW-78
Mcp4	59-TITEFRADKSIPGQW-78
CstC	26-VGEYNKASNDMYHSR-45
CstE	23-VASYNMGSNSIYYFR-42
CstF	33-VEKFNNCTNDMFLFK-52
BbCI	50-TVRFETPLAIAIITE-69
STI	50-GTIISSPYRIRFIAE-69

To verify the role of the residues in these regions, the different mutants were prepared. We exchanged the residues that are different in macrocypins 1 and 4. Ile72 and Asn74 in macrocypin 1 were replaced by equipositioned residues Ala and Lys in macrocypin 4, respectively, producing mutant Mcp1-I72A+N74K. Ala70 and Lys72 in macrocypin 4 were replaced by equipositioned residues Ile and Asn in macrocypin 1, respectively, producing mutant Mcp4-A70I+K72N. In addition, the corresponding clitocypin mutant Clt-N69K was also expressed, where equipositioned Lys residue was replaced by Asn.

As expected, the mutant Mcp1-I72A+N74K exhibited total loss of inhibition of AEP, whereas the mutant Mcp4-A70I+K72N became inhibitory (Table 5). Equivalently, Clt-N69K mutant lost its inhibitory properties against AEP. These studies thus confirm that Asn72 in macrocypins and Asn69 in clitocypins are the residues that bind into the S1 pocket of AEP [146].

Table 14: *Inhibition constants of legumain and cathepsin V by macrocypin, clitocypin and their mutants.* Kinetic data marked with asterisk were reported previously [85].

	K_i (nM)	K_i (nM)
	Legumain	Cathepsin V
Mcp 1*	3.38±1.44	0.69±0.06
Mcp1-I72A+N74K	n.i.	3.43±0.31
Mcp 4*	>1000	1.44±0.11
Mcp4-A70I+K72N	2.86±0.38	10.2±0.6
Mcp 4 K72R	n.i.	/
Clt*	21.5±2.81	0.084±0.03
Clt N70K	n.i.	0.26±0.09

4.8.7 Trypsin inhibition

The binding geometry of several families of protein inhibitors of serine proteases, including soybean Kunitz type inhibitor [74], are known to adopt a substrate-like conformation known as the “canonical” binding mode [72]. All Kunitz type serine protease inhibitors inhibit trypsin with a highly homologous loop from the root region that mimics the substrate and is positioned between the strands ß4-ß5. This loop contains either lysine or arginine, which binds into the S1 pocket of trypsin. From the sequence and structure alignments it is evident that the classical ß4-ß5 loop is missing in macrocypin and clitocypin. These proteins show no inhibition of the serine protease trypsin. Surprisingly, macrocypin 4 was found to exhibit weak inhibition of trypsin, with a Ki value in the micromolar range. The Ki values of the exchange mutants produced for AEP binding site identification (Table 5) show that the Lys74 residue of macrocypin 4 is mandatory for inhibition of trypsin. The Mcp4 mutant with Lys74 replaced by arginine (Mcp4-K74R) exhibited equivalent inhibitory properties (Table 5), thereby confirming the involvement of the loop ß5-ß6 positioned within the lower “crown” layer in binding to the trypsin active site. The binding loop of macrocypins and clitocypins is thus positioned differently from the serine protease-binding loop of known Kunitz-type inhibitors such as STI [146].

Table 15: *Inhibition constants of trypsin and cathepsin V by macrocypin, clitocypin and their mutants.* Kinetic data marked with asterisk were reported previously [85].

	K_i (μM)	K_i (nM)
	Trypsin	Cathepsin V
Mcp 1*	n.i.	0.69±0.06

Mcp1-I72A+N74K	0.18±0.02	3.43±0.31
Mcp 4*	0.16±0.01	1.44±0.11
Mcp4-A70I+K72N	n.i.	10.2±0.6
Mcp 4 K72R	0.13±0.02	/
Clt*	n.i.	0.084±0.03
Clt N70K	n.i.	0.26±0.09

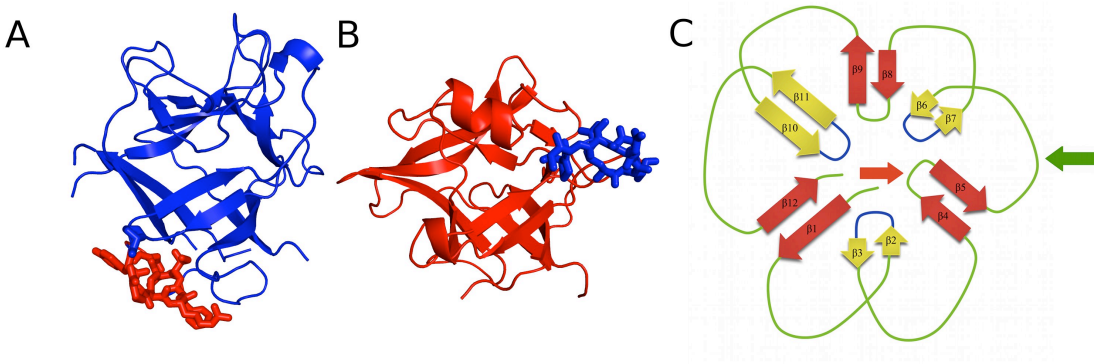


Figure 54: *Comparison of trypsin binding loop position in STI and macrocypin.* A) STI (blue) with trypsin binding loop in red sticks, B) macrocypin (red) with trypsin binding loops in blue sticks and C) schematic representation of β-trefoil fold. Green arrow indicates the position of trypsin binding loop in macrocypin and red arrow indicates the position of trypsin binding loop in Kunitz-type inhibitors.

5 Discussion

The crystal structures of complexes between stefin A and the naked cathepsins V and B reveal the geometry of binding in the state when the interacting molecules are packed together closest. We were not able to crystallize the complex between the naked cathepsin V and stefin A, while the complex between stefin A and the naked cathepsin H has been determined previously [24]. To increase the separation between the pairs of interacting molecules the reactive site cysteine residue was modified with a few small molecule reagents: methylmethanethiosulfonate (MMTS), which prolongs the side chain cysteine by a thiol methyl group (-S-CH_3) and iodoacetic acid, which results in carboxymethylation of the cysteine residue prolonging it with the -CH_2-COOH group. The crystal structures of cathepsins L and V blocked with MMTS reveal the geometry of binding with the slightly increased separation, with 0.1 to 0.2 Å longer separations when compared with the complex with the naked cathepsin V.

We have not succeeded to crystallize complexes between carboxymethylated enzymes and stefin A, so the only insight into the geometry of binding with the CH_3-COOH spacer is still provided by the papain - stefin B complex [89].

The structures of all these complexes reveal the same geometric mode of binding of inhibitors into the active sites of proteases. In the case of endopeptidases the structure of the enzymes remained unaltered compared to the structures of the free enzymes [129, 163] whereas in the complexes with exopeptidases inhibitors must displace the occluding loop residues of cathepsin B and the mini-chain residues of cathepsin H. These bindings are thus associated with structural changes, which obviously also affect the kinetics and the binding energy. Due to the binding induced structural changes the later systems are thus not useful for the studies of dependence of the binding energy upon the separation distance. They do, however, provide a valuable view into the geometry of binding of cysteine protease inhibitors to their targets and are therefore discussed below. Of similar importance are the structures of the clitocypin alone and in the complex with cathepsin V, and macrocypin, which provide a view into the binding mechanism and flexibility of the binding loops on the clitocypin side. Due to the uniqueness of these interactions they are discussed as last.

5.1 Cathepsin L and stefin A case study: Insight into intramolecular interactions

Atom force microscopy measurements provide direct insight into the separation of two molecules, at which they stop interacting. The behaviour of protein - protein interaction in single molecule force spectroscopy is described by the Bell's model which enables determination of the most probable unbinding force at a given loading rate, the distance between the potential minimum and the transition state x_β and the kinetic off-rates in the system under applied force.

These measurements confirmed the theoretical predictions made by the Bell's model, where the increase of a most probable unbinding force over the range of almost two orders or magnitude followed the theoretical predictions. The value for the distance between the potential minimum and the transition state xB was determined to be 8.5 A. This value is in the range of previously published data, which ranges from 0.3 to 30 Å for biotin [158], 3.0 Å for N-cadherin dimers, 7.7 Å for E-cadherin [159] and 5.9 Å for cadherin - anti-cadherin [138].

Atom force microscopy only provided insight into the two stages of separation, while the distance dependence of the interaction remains unapproachable. In order to gain insight into other stages the catalytic cysteine residue of cysteine cathepsins was modified with two functional groups of different sizes. By introduction of these modifications two additional complexes could be analysed at different separation distances using methods that rely on measurements of steady (equilibrium) states: crystal structure determination and binding constants. The cathepsins, blocked with MMTS, exhibited 0.2 Å shallower binding of inhibitors than the naked ones, whereas the distance 0.8 Å for the carboxymethylated cathepsins was estimated from the comparisons of the crystal structure of papain in complex with stefin B and cathepsin H in complex with stefin A determined previously [24, 89].

We have tried to determine the crystal structures of the carboxymethylated cathepsin V in complex with

stefin A, but the diffraction of the visually perfect crystals was limited to approximately 6 Å (data not shown), delivering data insufficient for determination of an atomic structure.

Even though we were only able to determine the distance for the MMTS-modified cathepsin L experimentally from the structure, two additional distances could be determined from the models and structures of cathepsin L. The distance between the native cathepsin L and stefin A was evaluated on the basis of the differences between the distances of the reduced and unreduced cathepsin V and stefin A, whereas the distance between carboxymethylated cathepsin L and stefin A was evaluated on the model prepared from the crystal structure of the carboxymethylated papain in the complex with stefin A.

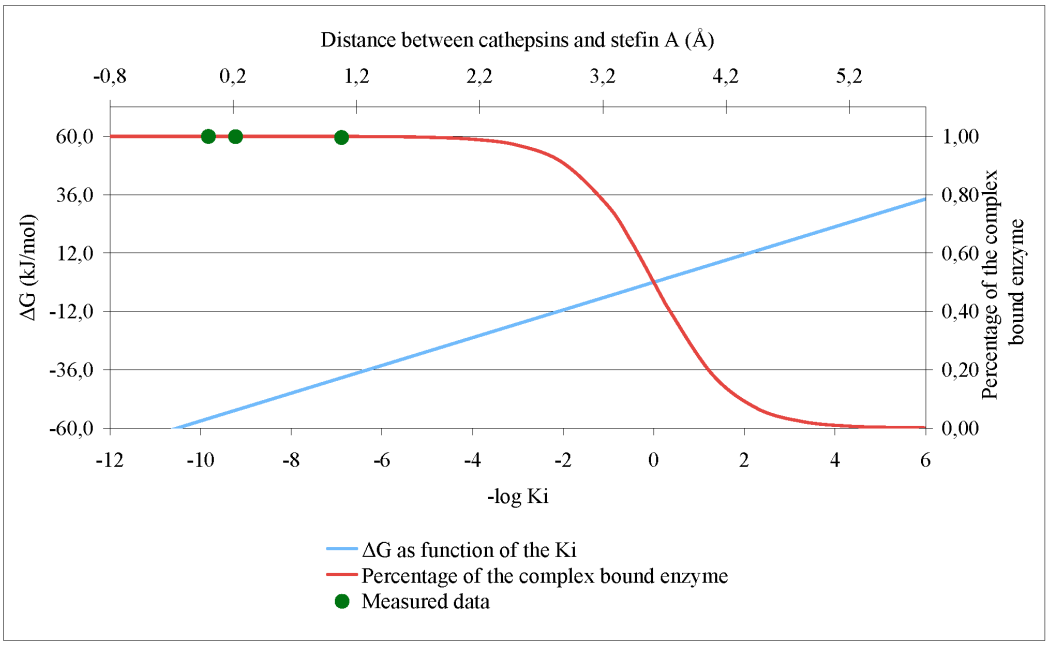


Figure 53: - *log Ki as function of the distance between different cathepsins.* Left point represents the reduced cathepsin L and stefin A distance (evaluated on the basis of differences between distances of reduced and unreduced cathepsin V and stefin A), second point represents the distance in unreduced cathepsin L - stefin A complex (experimentally determined) and third point is the distance between stefin A and carboxymethylated cathepsin L (modelled on the basis of structure of carboxymethylated papain in complex with stefin B). The inhibition constants for native L, MMTS modified and carboxymethylated cathepsin L were determined by surface plasmon resonance and plotted on the graph, showing the percentage of the complex-bound enzyme, and the secondary axis showing the corresponding distance was added.

The figure 53 summarizes these data. The equilibrium equation allowed us to calculate the percentage of the complex bound enzyme as a function of the inhibition constant. We assumed that the concentration of the enzyme and the inhibitor are the same at the beginning. If the inhibition constant is lower than 10^{-2} M, the majority of the enzyme is bound in the complex. The percentage of the complex-bound enzyme begins to lower when the dissociation constant increases and reaches the separated state, where the constant is larger than 10 M. It is evident from the graph that almost all cathepsins, modified or not, are bound in the complex with stefin A.

The graph shows linear dependence of the logarithmic value of the inhibition constant and the separation distance between the naked and modified cathepsins L and stefin A. The linear dependence in the range covering the three orders of magnitude for the inhibition constant which encompasses approximately 1 Å separation range between molecules of cathepsin L and stefin A. Upon binding approximately 25% of interaction energy is released within these range. Assuming the symmetrical shape and a corresponding linear relationship of the distance dependence of the interaction constant one can hypothesize that if the spacer had been about 7 Å long, then the two molecules would not form a complex at all. This distance is consistent with the interaction limit 8.5 Å of the two molecules obtained by the AFM measurement. Hence, as a crude approximation the relationship betwene the binding energy and separation distance is indeed linear. Furthermore, under the conditions of the measurement, the distance range of interactions between stefins and papain-like cysteine proteases is bellow 10 Å (assuming that they are properly oriented). At ranges beyond this separation, the molecules do not feel each other.

5.2 Inhibition mechanism of stefins: Rigid docking and flexibility of cathepsins

The insight into the mechanism of inhibition of cathepsins by stefins has been provided by the crystal structure of stefin B in complex with carboxymethylated papain [89] and later the crystal structure of cathepsin H - stefin A complex [24]. These two structures have limitations: papain is a plant protease, the observation of the genuine interaction with the active site is hindered by cysteine carboxymethylation, while cathepsin H is an exopeptidase with the mini-chain, which alters the binding of stefin A to the active site.

Stefin A binds into the active site cleft as expected, with the N-terminal trunk in nonprimed binding sites and the first and second loop in primed binding sites. The N-terminal trunk fills the S2 binding site with Pro3 and the S1 site with Gly4. Compared to papain - stefin B complex, the N-terminal trunk is held in position by only two hydrogen bonds, since Pro3 in stefin is lacking the amide hydrogen. The first loop with the sequence QVVAG, which is conserved in cystatins [24], approaches the active site cleft near the active site histidine and cysteine residue and leaves it at the S2 site. Val47 lies over the catalytic residues; Val48 fills the S1 site, and Ala49 partially occupies the S2 site. There are three hydrogen bonds, which anchor the loop to cathepsins. Whereas the N-terminal trunk and the first loop are well ordered in all complexes, the third loop is showing surprising flexibility. The positions of equivalent CA atoms in stefin A molecules from all known complexes differ for more than 5 A, indicating that the third loop has a low affect to overall affinity.

The crystal structures are crucial for understanding of the inhibitory mechanism; they do not provide a direct insight into the interaction strength. Therefore a classical biochemical determination of the inhibition constant using hydrolysis of the fluorogenic substrate was employed. By combining the two approaches we tried to find a correlation between the lengths of hydrogen bonds, the overall distance between stefin and cathepsin atoms and the kinetic parameters.

The hydrogen bond distance analysis showed a moderate correlation with the inhibition constants. It seems that the major factor in the complex stability are the three hydrogen bonds, which are formed by the main chain atom in stefin A (Pro3, Gly4, Ala49). These hydrogen bonds were generally shorter in complexes with the lower inhibition constant. Two additional hydrogen bonds between the long side chains in stefin A and the residues, forming the edge of the active site cleft in cathepsins, however, showed no correlation between the complex stability and intramolecular separation.

The comparison has also revealed that the inhibition constants of the native cathepsins V and L with stefin A determined by surface plasmon resonance are equivalent to the constants determined by the substrate degradation methods. The inhibition constants for the MMTS blocked cathepsins V and L are however, 5 and 20 fold higher, respectively. The carboxymethylated cathepsin L has a 1500 fold higher inhibition constant. Our findings are similar to the previously published data describing the interactions of inactivated papain with the chicken cystatins, where the modification of the active site cystein residue increased the inhibition constant 10 fold [150]. As seen from the Table 7, the difference in the interaction strength is a consequence of the changed on and off rates, whereas the work on inactivated papain suggested that the lower inhibition is due to the higher dissociation (off) rates.

The occluding loop is specific feature of cathepsin B, responsible for its exopeptidase activity. The 22-residue insertion between Ile105 and Pro126 has no equivalent in other endopeptidases, where only short a connection is used instead of occluding loop.

The occluding loop is stabilized by a salt bridge between His110-Asp22 and Arg116-Asp224 and blocks the active site beyond the S2 binding site [20, 164]. The occluding loop behaves quite independently from the rest of the structure and can adopt different conformations.

We have shown that the occluding loop can adopt different positions in the complex between cathespin B and stefin A. It is highly likely that the occluding loop is fully flexible and can move further away in solution, but its movement in the crystals of cathepsin B – stefin A complexes is limited because of the extremely tight crystal packing.

5.3 Inhibition mechanism of mycocypins: Flexibility of inhibitiors

Cathepsin V in the crystal structure of the complex with clitocypin is also methylated with MMTS. In accordance with the complexes of similarly modified cathepsins one must assume that the observed docking mode is genuine, however, the separation between the molecules may be larger than in the complex with the naked enzyme. In the absence of the binding constant with the blocked enzyme and absence of the structure with the naked enzyme this remains a hypothesis. (It must be said that we have

failed to crystallize the complex with the naked enzyme. Possible explanations are the instability of the complex during crystallization or unsuitable packing parameters.) Nevertheless, the mode of binding has been revealed. Together with the enzyme kinetics data we were able to conclude that the binding loops in clitocypins, macrocypins, and Kunitz type inhibitors are flexible and can adopt their conformation to fit into the active site cleft of papain-like cysteine proteases.

Like cystatins [24, 89], the p41 fragment [19] and chagasin [57], clitocypin and macrocypin bind to cysteine proteases along the whole active site cleft. These molecules have different folds yet, for docking to papain-like cysteine proteases, they utilize a similar architecture by which the activity of the target proteases is inhibited. Their constructs occlude the reactive site cysteine. On the non-primed substrate-binding site they utilize a single chain. The first binding region in clitocypin is the loop Asp19-Glu25. Its position is similar to those of the N-terminal region in stefin A and the first loops of the p41 fragment and chagasin. This region contains a residue that, in a substrate-like manner, fills the S2 binding pocket. In contrast, the loops covering the primed binding areas are much less similar. The second binding region in clitocypin is a single, broad loop (Glu39-Ile50) whereas cystatins, the p41 fragment and chagasin use two loop constructs. The two broad loops of mycocypins are stabilized by multiple hydrogen bonds and are much more rigid than the N-terminal trunk and two loops in cystatins [146]. This explains why cystatins are capable of competing for binding with the additional features of exopeptidase such as occluding loop and mini chain, whereas mycocypins exhibit lower affinity or no binding at all.

The mode of the binding of mycocypins to cysteine cathepsins differs markedly from the binding of Kunitz type of B-trefoil folded inhibitors to serine proteases. The two binding loops from the crown region bind into the non-primed and primed substrate binding regions of cysteine proteases and occlude the catalytic cysteine residue. In contrast, only one binding loop from the root region of Kunitz inhibitors docks into the active site of a serine protease, in a substrate-like manner. A possible explanation for the differences in the modes of "canonical" inhibition of cysteine cathepsins and trypsin-like serine proteases may lie in the features of the active site cleft. Whereas in the trypsin-like proteases the S1 binding site is a pocket in the protein structure, in cysteine cathepsins the S1 binding site is positioned on the surface on one side of the active site cleft, shaped so that the P1 residue side chain points away from the protein core [15]. Furthermore, analysis of the structural data has revealed that papain-like cysteine proteases have only three clearly defined substrate binding sites (S2, S1, S1') and one conditional site (S2'), while the binding into regions beyond position 2 can only be considered substrate binding areas spread over the surface of the widening active site cleft [157]. Cysteine cathepsins thus appear to lack the binding surface to which the P1 and neighboring residues could be tightly anchored in a substrate-like manner and therefore can probably not be inhibited by the single loop construct.

The flexible peptide bond, which can flip on docking to protease, is a unique feature among the cysteine protease inhibitors [146]. Peptide flipping has already been observed in the mitogen-activated protein kinase p38α MAPK, where the flip of the Met109 and Gly110 peptide bond facilitates the higher specificity of certain inhibitors [165].

The trefoil fold supports 11 loops coming out of the six-stranded B-barrel. Nine are in the crown region (six are positioned at the lower level of the crown and three enclosing the top of the crown) and two in the roots. Therefore it is easier to comprehend that the six loops from the lower crown region can act in pairs, whereas the two loops from the root region lack that capability and must bind alone. In this respect the report of the binding site of *Bauhinia bauhiniodes* cruzipain inhibitor (BbCI) to cysteine cathepsins is intriguing, since the authors [70] suggest that the same alanine residue positioned within the root region, which is responsible for binding to neutrophil elastase, is also crucial for the binding to cathepsin L and cruzipain. (The partial cleavage of the serine protease interacting loop following incubation of BbCI with cruzipain was the key evidence supporting the hypothesis of the common interaction site.) The binding of BbCI to trypsin is consistent with current structural knowledge, since the loop in which Ala63 resides folds very similarly to the loop of STI, including the position and orientation of the Ala63 residue. However, the single inhibitory loop is not consistent with the canonical inhibition mechanism of cysteine cathepsins evidenced here. Superimposition of the structure of BbCI on clitocypin in complex with cathepsin V showed that two broad loops in the BbCI structure are equivalent to the clitocypin binding loops, and that the BbCI sequence contains two consecutive glycine residues, Gly28-Gly29, homologous to the peptide bond flip residues Gly24 and Gly25. Hence, these two loops are probably responsible for cathepsin L inhibition, and not the loop containing the trypsin cleavage site. (It should be noted though that, as in the case of macrocypin, the tips of the loops require a slight adjustment in order to fit into the active site of a cysteine cathepsin.) The absence of inhibition of cathepsin V by BbCI, given the similarity of cathepsins L and V, cannot be explained.

Clitocypins and macrocypins exhibit no sequence similarity to other known proteins, which was the basis for establishing the I48 and I85 families, supported by the sequence alignment score with an E value

less than 0.001 (FASTA search with default BLOSUM50 matrix used in MEROPS). However, their structure has revealed that the basic element of their fold is the six-stranded B-barrel, the hallmark of the B-trefoil fold shared by the members of the I3 MEROPS family that includes serine protease inhibitors of the Kunitz type. The sequence similarity, based on superimposition of the structures, is low even within the 6-stranded B-barrel part, thus questioning the common origin of these two groups of proteins. This confirms that mycocypins (families I48 and I85) are indeed distinct from members of I3 family, while the structural similarity between these families provides support that they belong to the same clan IC.

The B-trefoil fold is armed with potent interacting loops that differ in shape and composition, and are able to inhibit several classes of proteases including cathepsins, AEP, cruzipain, trypsin, chymotrypsin, elastase, subtilisin and amylases. Several loops are involved in inhibition, while the same inhibitory loop can target different proteases. For example, the crown region loops B1-B2 and B3-B4 are used for inhibiting the papain-like cysteine proteases by mycocypins and, most probably, cruzipain and cathepsin L by BbCI [70]. The root region loop B4-B5 is involved in inhibiting chymotrypsin (by the winged bean chymotrypsin inhibitor) [162], trypsin (by STI) [74], and porcine pancreatic elastase and human neutrophil elastase (by BbCI) [70], whereas the crown region loop B5-B6 is involved in inhibiting AEP and trypsin (by mycocypins) and the subtilisin savinase (by barley α-amylase/subtilisin inhibitor, BASI) [166]. The numerous crown region loops, B1-B2, B3-B4, B6-B7, B9-B10 and B11-B12, are responsible for the interaction of BASI with barley α-amylase [167]. This makes B-trefoil inhibitors, in particular mycocypins, promising candidates for transgenic trials for the purposes of crop protection, where inhibitors with selectivity against only one class of proteases have failed due to the compensation of proteolytic activity by induced expression of other proteases insensitive to the transgenic inhibitor [168-170].

6 Conclusions

We have provided insight into the distance dependence of energetics of interactions between proteases and inhibitors using cathepsins and stefins. Using atom force microscopy we have shown that the molecules start interacting at distances below 10 Å. Using enzyme kinetics and surface plasmon resonance we have shown that at closer separations of molecules energy of interaction depends on their separation. Combining this with the insights provided by X-ray crystallography, we have observed that the energy of interaction between the partners depends on of their separation approximately linealy. Extrapolation of this relationship to zero energy indicates that the molecules stop interacting with each other at separations below 10 Å which is consistent with the atom force microscopy measurements.

The crystal structures have provided also novel insight into the interaction mechanism of cathepsins and their inihibitors: stefins, macrocypins and clitocypins. The structures have shown that modification of the active site cysteine residue prevents the genuine binding of the inhibitor to the protease, however, the complex is still formed. The binding of the N-terminal trunk and the first loop of the stefin A to cathepsins is well defined and is very similar between different cathepsins, whereas the second loop in stefin A exhibits large flexibility. We have also shown that the occluding loop in the cathepsin B is flexibile and can adopt a variety of conformations, depending on the size and shape of the inhibitor.

Contrary to the three loops in stefin A, clitocypin utilizes only two loops, which binds to the active site cleft. The crystal structures of the complex of clitocypin with cathepsin V, and macrocypin and clitocypin alone have revealed 3-dimensional structures of these inhibitors and a yet another motif of binding to a papain like-cysteine protease. The binding is associated with a peptide-bond flip of glycine that occurs prior to or concurrently with the inhibitor docking. Mycocypins posse a ß-trefoil fold, the hallmark of Kunitz type inhibitors. It is a tree-like structure with 2 loops in the root region, a stem comprising a six-stranded ß-barrel, and two layers of loops (6+3) in the crown region. The two loops that bind to the active site of cysteine cathepsins belong to the lower layer of the crown loops, while a single loop from the crown region can inhibit trypsin or asparaginyl endopeptidase, as demonstrated by site directed mutagenesis.

7 Acknowledgements

The inhibition kinetics of cathepsins, legumain, and trypsin with macrocypin, clitocypin and its mutants were determined by dr. Jerica Sabotič and Tatjana Popovič (Department of Biotechnology, Jožef Stefan Institute).

The SPR measurements were performed by dr. Gregor Anderluh and Vesna Hodnik (Department of Biology, Biotechnical Faculty, University of Ljubljana).

This thesis would not have been possible without:

- my supervisor Prof. Dr. Dušan Turk
- the Examination Board (Prof. Dr. Janko Kos, Prof. Dr. Igor Muševič, Prof. Dr. Peter Hinterdorfer)
- my coworkers (Marko Mihelič, Igor Štern, Gregor Gunčar, Andreja Doberšek, Miha Andrejašič, Zoran Štefanič, Katja Galeša, Nataša Kopitar Jerala and others)
- other coworkers from our department
- Jerica Sabotič, Jože Brzin, Tatjana Popovič and Janko Kos from Department of biotechnology, Jožef Stefan Institute
- Peter Hinterdorfer and Linda Wildling from Institute for Biophysics, Johannes Kepler Universität Linz
- Gregor Anderluh and Vesna Hodnik from Biotechnical faculty, Universty of Ljubljana
- all others, who helped me in any way....

Hvala družini, Jerici in Simoni za spodbudo in razumevanje.

8 References

1. Radzicka, A., Wolfenden, R. (1995) A proficient enzyme. *Science* **267**, 90-3.
2. Southan, C. (2000) Assessing the protease and protease inhibitor content of the human genome. *J Pept Sci* **6**, 453-8.
3. Medintz, I.L., Clapp, A.R., Brunel, F.M., Tiefenbrunn, T., Uyeda, H.T., Chang, E.L., Deschamps, J.R., Dawson, P.E., Mattoussi, H. (2006) Proteolytic activity monitored by fluorescence resonance energy transfer through quantum-dot-peptide conjugates. *Nat Mater* **5**, 581-9.
4. Quesada, V., Ordonez, G.R., Sanchez, L.M., Puente, X.S., Lopez-Otin, C. (2009) The Degradome database: mammalian proteases and diseases of proteolysis. *Nucleic Acids Res* **37**, D239-43.
5. Whitcomb, D.C., Lowe, M.E. (2007) Human pancreatic digestive enzymes. *Dig Dis Sci* **52**, 1-17.
6. Macfarlane, R.G. (1964) An Enzyme Cascade in the Blood Clotting Mechanism, and Its Function as a Biochemical Amplifier. *Nature* **202**, 498-9.
7. Kerr, J.F., Wyllie, A.H., Currie, A.R. (1972) Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br J Cancer* **26**, 239-57.
8. Kumar, S. (1999) Regulation of caspase activation in apoptosis: implications in pathogenesis and treatment of disease. *Clin Exp Pharmacol Physiol* **26**, 295-303.
9. Turk, B., Turk, D., Salvesen, G.S. (2002) Regulating cysteine protease activity: essential role of protease inhibitors as guardians and regulators. *Curr Pharm Des* **8**, 1623-37.
10. Biggs, R., Douglas, A.S., Macfarlane, R.G., Dacie, J.V., Pitney, W.R., Merskey (1952) Christmas disease: a condition previously mistaken for haemophilia. *Br Med J* **2**, 1378-82.
11. Hart, P.S., Zhang, Y., Firatli, E., Uygur, C., Lotfazar, M., Michalec, M.D., Marks, J.J., Lu, X., Coates, B.J., Seow, W.K., Marshall, R., Williams, D., Reed, J.B., Wright, J.T., Hart, T.C. (2000) Identification of cathepsin C mutations in ethnically diverse papillon-Lefevre syndrome patients. *J Med Genet* **37**, 927-32.
12. Lopez-Otin, C., Matrisian, L.M. (2007) Emerging roles of proteases in tumour suppression. *Nat Rev Cancer* **7**, 800-8.
13. Rawlings, N.D., Morton, F.R., Kok, C.Y., Kong, J., Barrett, A.J. (2008) MEROPS: the peptidase database. *Nucleic Acids Res* **36**, D320-5.
14. Rawlings, N.D., Tolle, D.P., Barrett, A.J. (2004) Evolutionary families of peptidase inhibitors. *Biochem J* **378**, 705-16.
15. Turk, B., Turk, V., Turk, D. (1997) Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors. *Biol Chem* **378**, 141-50.
16. Drenth, J., Jansonius, J.N., Koekoek, R., Wolthers, B.G. (1971) The structure of papain. *Adv Protein Chem* **25**, 79-115.
17. Schechter, I., Berger, A. (1967) On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* **27**, 157-62.
18. Turk, D., Podobnik, M., Popovic, T., Katunuma, N., Bode, W., Huber, R., Turk, V. (1995) Crystal structure of cathepsin B inhibited with CA030 at 2.0-A resolution: A basis for the design of specific epoxysuccinyl inhibitors. *Biochemistry* **34**, 4791-7.
19. 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.
20. 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 A X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *EMBO J* **10**, 2321-30.
21. Guncar, G., Klemencic, I., Turk, B., Turk, V., Karaoglanovic-Carmona, A., Juliano, L., Turk, D. (2000) Crystal structure of cathepsin X: a flip-flop of the ring of His23 allows carboxy-monopeptidase and carboxy-dipeptidase activity of the protease. *Structure* **8**, 305-13.
22. Guncar, G., Podobnik, M., Pungercar, J., Strukelj, B., Turk, V., Turk, D. (1998) Crystal structure of porcine cathepsin H determined at 2.1 A resolution: location of the mini-chain C-terminal carboxyl group defines cathepsin H aminopeptidase function. *Structure* **6**, 51-61.

23. Turk, D., Janjic, V., Stern, I., Podobnik, M., Lamba, D., Dahl, S.W., Lauritzen, C., Pedersen, J., Turk, V., Turk, B. (2001) 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**, 6570-82.
24. 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**, 875-85.
25. Stern, I., Schaschke, N., Moroder, L., Turk, D. (2004) Crystal structure of NS-134 in complex with bovine cathepsin B: a two-headed epoxysuccinyl inhibitor extends along the entire active-site cleft. *Biochem J* **381**, 511-7.
26. 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**, 4516-21.
27. 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 Ii degradation and CD4 T cell selection in the thymus. *Science* **280**, 450-3.
28. Nakagawa, T.Y., Brissette, W.H., Lira, P.D., Griffiths, R.J., Petrushova, N., Stock, J., McNeish, J.D., Eastman, S.E., Howard, E.D., Clarke, S.R., Rosloniec, E.F., Elliott, E.A., Rudensky, A.Y. (1999) Impaired invariant chain degradation and antigen presentation and diminished collagen-induced arthritis in cathepsin S null mice. *Immunity* **10**, 207-17.
29. 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**, 8627-32.
30. 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 pertubation of hair follicle cycling. *FASEB J* **14**, 2075-86.
31. Saftig, P., Hunziker, E., Wehmeyer, O., Jones, S., Boyde, A., Rommerskirch, W., Moritz, J.D., Schu, P., von Figura, K. (1998) Impaired osteoclastic bone resorption leads to osteopetrosis in cathepsin-K-deficient mice. *Proc Natl Acad Sci U S A* **95**, 13453-8.
32. 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.
33. Chapman, H.A., Riese, R.J., Shi, G.P. (1997) Emerging roles for cysteine proteases in human biology. *Annu Rev Physiol* **59**, 63-88.
34. Berquin, I.M., Sloane, B.F. (1996) Cathepsin B expression in human tumors. *Adv Exp Med Biol* **389**, 281-94.
35. Podgorski, I., Sloane, B.F. (2003) Cathepsin B and its role(s) in cancer progression. *Biochem Soc Symp*, 263-76.
36. Sloane, B.F. (1990) Cathepsin B and cystatins: evidence for a role in cancer progression. *Semin Cancer Biol* **1**, 137-52.
37. Premzl, A., Zavasnik-Bergant, V., Turk, V., Kos, J. (2003) Intracellular and extracellular cathepsin B facilitate invasion of MCF-10A neoT cells through reconstituted extracellular matrix in vitro. *Exp Cell Res* **283**, 206-14.
38. Joyce, J.A., Baruch, A., Chehade, K., Meyer-Morse, N., Giraudo, E., Tsai, F.Y., Greenbaum, D.C., Hager, J.H., Bogyo, M., Hanahan, D. (2004) Cathepsin cysteine proteases are effectors of invasive growth and angiogenesis during multistage tumorigenesis. *Cancer Cell* **5**, 443-53.
39. Turk, V., Stoka, V., Turk, D. (2008) Cystatins: biochemical and structural properties, and medical relevance. *Front Biosci* **13**, 5406-20.
40. Turk, V., Bode, W. (1991) The cystatins: protein inhibitors of cysteine proteinases. *FEBS Lett* **285**, 213-9.
41. Kordis, D., Turk, V. (2009) Phylogenomic analysis of the cystatin superfamily in eukaryotes and prokaryotes. *BMC Evol Biol* **9**, 266.
42. Jarvinen, M., Rinne, A., Hopsu-Havu, V.K. (1987) Human cystatins in normal and diseased tissues--a review. *Acta Histochem* **82**, 5-18.
43. Kos, J., Krasovec, M., Cimerman, N., Nielsen, H.J., Christensen, I.J., Brunner, N. (2000) Cysteine proteinase inhibitors stefin A, stefin B, and cystatin C in sera from patients with colorectal cancer: relation to prognosis. *Clin Cancer Res* **6**, 505-11.
44. Barrett, A.J., Rawlings, N.D., Woessner, J.F. (1998) Handbook of proteplytic enzymes. *London, UK, Academic Press*, 1867.

45. Shlipak, M.G., Katz, R., Sarnak, M.J., Fried, L.F., Newman, A.B., Stehman-Breen, C., Seliger, S.L., Kestenbaum, B., Psaty, B., Tracy, R.P., Siscovick, D.S. (2006) Cystatin C and prognosis for cardiovascular and kidney outcomes in elderly persons without chronic kidney disease. *Ann Intern Med* **145**, 237-46.

46. Levy, E., Lopez-Otin, C., Ghiso, J., Geltner, D., Frangione, B. (1989) Stroke in Icelandic patients with hereditary amyloid angiopathy is related to a mutation in the cystatin C gene, an inhibitor of cysteine proteases. *J Exp Med* **169**, 1771-8.

47. Alvarez-Fernandez, M., Barrett, A.J., Gerhartz, B., Dando, P.M., Ni, J., Abrahamson, M. (1999) Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. *J Biol Chem* **274**, 19195-203.

48. Dubin, G. (2005) Proteinaceous cysteine protease inhibitors. *Cell Mol Life Sci* **62**, 653-69.

49. 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**, 429-34.

50. Habib, H., Fazili, K.M. (2007) Plant protease inhibitors: a defense strategy in plants. *Biotechnol. Mol. Biol. Rev.* **2**, 17.

51. Lenarcic, B., Bevec, T. (1998) Thyropins--new structurally related proteinase inhibitors. *Biol Chem* **379**, 105-11.

52. Lenarcic, B., Turk, V. (1999) Thyroglobulin type-1 domains in equistatin inhibit both papain-like cysteine proteinases and cathepsin D. *J Biol Chem* **274**, 563-6.

53. Fowlkes, J.L., Thrailkill, K.M., Serra, D.M., Nagase, H. (1997) Insulin-like growth factor binding protein (IGFBP) substrate zymography. A new tool to identify and characterize IGFBP-degrading proteinases. *Endocrine* **7**, 33-6.

54. Mihelic, M., Turk, D. (2007) Two decades of thyroglobulin type-1 domain research. *Biol Chem* **388**, 1123-30.

55. Santos, C.C., Sant'anna, C., Terres, A., Cunha-e-Silva, N.L., Scharfstein, J., de, A.L.A.P. (2005) Chagasin, the endogenous cysteine-protease inhibitor of Trypanosoma cruzi, modulates parasite differentiation and invasion of mammalian cells. *J Cell Sci* **118**, 901-15.

56. Redzynia, I., Ljunggren, A., Bujacz, A., Abrahamson, M., Jaskolski, M., Bujacz, G. (2009) Crystal structure of the parasite inhibitor chagasin in complex with papain allows identification of structural requirements for broad reactivity and specificity determinants for target proteases. *FEBS J* **276**, 793-806.

57. Ljunggren, A., Redzynia, I., Alvarez-Fernandez, M., Abrahamson, M., Mort, J.S., Krupa, J.C., Jaskolski, M., Bujacz, G. (2007) Crystal structure of the parasite protease inhibitor chagasin in complex with a host target cysteine protease. *J Mol Biol* **371**, 137-53.

58. Redzynia, I., Ljunggren, A., Abrahamson, M., Mort, J.S., Krupa, J.C., Jaskolski, M., Bujacz, G. (2008) Displacement of the occluding loop by the parasite protein, chagasin, results in efficient inhibition of human cathepsin B. *J Biol Chem* **283**, 22815-25.

59. Rzychon, M., Chmiel, D., Stec-Niemczyk, J. (2004) Modes of inhibition of cysteine proteases. *Acta Biochim Pol* **51**, 861-73.

60. Schick, C., Bromme, D., Bartuski, A.J., Uemura, Y., Schechter, N.M., Silverman, G.A. (1998) The reactive site loop of the serpin SCCA1 is essential for cysteine proteinase inhibition. *Proc Natl Acad Sci U S A* **95**, 13465-70.

61. 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**, 25172-7.

62. Yamamoto, Y., Kurata, M., Watabe, S., Murakami, R., Takahashi, S.Y. (2002) Novel cysteine proteinase inhibitors homologous to the proregions of cysteine proteinases. *Curr Protein Pept Sci* **3**, 231-8.

63. Kurata, M., Yamamoto, Y., Watabe, S., Makino, Y., Ogawa, K., Takahashi, S.Y. (2001) Bombyx cysteine proteinase inhibitor (BCPI) homologous to propeptide regions of cysteine proteinases is a strong, selective inhibitor of cathepsin L-like cysteine proteinases. *J Biochem* **130**, 857-63.

64. Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J.S., Cygler, M. (1996) Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J* **15**, 5492-503.

65. Rzychon, M., Sabat, A., Kosowska, K., Potempa, J., Dubin, A. (2003) Staphostatins: an expanding new group of proteinase inhibitors with a unique specificity for the regulation of staphopains, Staphylococcus spp. cysteine proteinases. *Mol Microbiol* **49**, 1051-66.

66. Rzychon, M., Filipek, R., Sabat, A., Kosowska, K., Dubin, A., Potempa, J., Bochtler, M. (2003)

Staphostatins resemble lipocalins, not cystatins in fold. *Protein Sci* **12**, 2252-6.

67. Filipek, R., Rzychon, M., Oleksy, A., Gruca, M., Dubin, A., Potempa, J., Bochtler, M. (2003) The Staphostatin-staphopain complex: a forward binding inhibitor in complex with its target cysteine protease. *J Biol Chem* **278**, 40959-66.
68. Brzin, J., Popovic, T., Drobnic-Kosorok, M., Kotnik, M., Turk, V. (1988) Inhibitors of cysteine proteinases from potato. *Biol Chem Hoppe Seyler* **369 Suppl**, 233-8.
69. Oliveira, A.S., Pereira, R.A., Lima, L.M., Morais, A.H.A., Melo, F.R., Franco, O.L., Bloch, C., Grossi-de-Sá, M.F., Sales, M.P. (2002) Activity toward Bruchid Pest of a Kunitz-Type Inhibitor from Seeds of the Algaroba Tree (Prosopis juliflora D.C.) *Pesticide Biochemistry and Physiology* **72**, 11.
70. Araujo, A.P., Hansen, D., Vieira, D.F., Oliveira, C., Santana, L.A., Beltramini, L.M., Sampaio, C.A., Sampaio, M.U., Oliva, M.L. (2005) Kunitz-type Bauhinia bauhinioides inhibitors devoid of disulfide bridges: isolation of the cDNAs, heterologous expression and structural studies. *Biol Chem* **386**, 561-8.
71. Laskowski, M., Jr., Kato, I. (1980) Protein inhibitors of proteinases. *Annu Rev Biochem* **49**, 593-626.
72. Bode, W., Huber, R. (1992) Natural protein proteinase inhibitors and their interaction with proteinases. *Eur J Biochem* **204**, 433-51.
73. Sweet, R.M., Wright, H.T., Janin, J., Chothia, C.H., Blow, D.M. (1974) Crystal structure of the complex of porcine trypsin with soybean trypsin inhibitor (Kunitz) at 2.6-A resolution. *Biochemistry* **13**, 4212-28.
74. Song, H.K., Suh, S.W. (1998) Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from Erythrina caffra and tissue-type plasminogen activator. *J Mol Biol* **275**, 347-63.
75. Onesti, S., Brick, P., Blow, D.M. (1991) Crystal structure of a Kunitz-type trypsin inhibitor from Erythrina caffra seeds. *J Mol Biol* **217**, 153-76.
76. Graves, B.J., Hatada, M.H., Hendrickson, W.A., Miller, J.K., Madison, V.S., Satow, Y. (1990) Structure of interleukin 1 alpha at 2.7-A resolution. *Biochemistry* **29**, 2679-84.
77. Ago, H., Kitagawa, Y., Fujishima, A., Matsuura, Y., Katsube, Y. (1991) Crystal structure of basic fibroblast growth factor at 1.6 A resolution. *J Biochem* **110**, 360-3.
78. Krizaj, I., Drobnic-Kosorok, M., Brzin, J., Jerala, R., Turk, V. (1993) The primary structure of inhibitor of cysteine proteinases from potato. *FEBS Lett* **333**, 15-20.
79. Franco, O.L., Grossi de Sa, M.F., Sales, M.P., Mello, L.V., Oliveira, A.S., Rigden, D.J. (2002) Overlapping binding sites for trypsin and papain on a Kunitz-type proteinase inhibitor from Prosopis juliflora. *Proteins* **49**, 335-41.
80. Hansen, D., Macedo-Ribeiro, S., Verissimo, P., Yoo Im, S., Sampaio, M.U., Oliva, M.L. (2007) Crystal structure of a novel cysteinless plant Kunitz-type protease inhibitor. *Biochem Biophys Res Commun* **360**, 735-40.
81. Brzin, J., Rogelj, B., Popovic, T., Strukelj, B., Ritonja, A. (2000) Clitocypin, a new type of cysteine proteinase inhibitor from fruit bodies of mushroom clitocybe nebularis. *J Biol Chem* **275**, 20104-9.
82. Galesa, K., Thomas, R.M., Kidric, M., Pain, R.H. (2004) Clitocypin, a new cysteine proteinase inhibitor, is monomeric: impact on the mechanism of folding. *Biochem Biophys Res Commun* **324**, 576-8.
83. Kidric, M., Fabian, H., Brzin, J., Popovic, T., Pain, R.H. (2002) Folding, stability, and secondary structure of a new dimeric cysteine proteinase inhibitor. *Biochem Biophys Res Commun* **297**, 962-7.
84. Sabotic, J., Galesa, K., Popovic, T., Leonardi, A., Brzin, J. (2007) Comparison of natural and recombinant clitocypins, the fungal cysteine protease inhibitors. *Protein Expr Purif* **53**, 104-11.
85. Sabotic, J., Popovic, T., Puizdar, V., Brzin, J. (2009) Macrocypins, a family of cysteine protease inhibitors from the basidiomycete Macrolepiota procera. *FEBS J* **276**, 11.
86. Pol, E., Olsson, S.L., Estrada, S., Prasthofer, T.W., Bjork, 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 (Pt 1)**, 275-82.
87. Drenth, J., Jansonius, J.N., Koekoek, R., Swen, H.M., Wolthers, B.G. (1968) Structure of papain. *Nature* **218**, 929-32.
88. Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J., Turk, V. (1988) The 2.0 A X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. *EMBO J* **7**, 2593-9.
89. Stubbs, M.T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B., Turk, V. (1990) The refined 2.4 A 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**, 1939-47.
90. Estrada, S., Pavlova, A., Bjork, 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, 7339-45.

91. Estrada, S., Nycander, M., Hill, N.J., Craven, C.J., Waltho, J.P., Bjork, 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**, 7551-60.
92. Pavlova, A., Bjork, 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**, 5649-58.
93. Heckman, K.L., Pease, L.R. (2007) Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat Protoc* **2**, 924-32.
94. Weiner, M.P., Costa, G.L. (1994) Rapid PCR site-directed mutagenesis. *PCR Methods Appl* **4**, S131-6.
95. Daly, R., Hearn, M.T. (2005) Expression of heterologous proteins in Pichia pastoris: a useful experimental tool in protein engineering and production. *J Mol Recognit* **18**, 119-38.
96. Cregg, J.M., Tolstorukov, I., Kusari, A., Sunga, J., Madden, K., Chappell, T. (2009) Expression in the yeast Pichia pastoris. *Methods Enzymol* **463**, 169-89.
97. Lin Cereghino, G.P., Sunga, A.J., Lin Cereghino, J., Cregg, J.M. (2001) Expression of foreign genes in the yeast Pichia pastoris. *Genet Eng (N Y)* **23**, 157-69.
98. Cereghino, J.L., Cregg, J.M. (2000) Heterologous protein expression in the methylotrophic yeast Pichia pastoris. *FEMS Microbiol Rev* **24**, 45-66.
99. Baneyx, F. (1999) Recombinant protein expression in Escherichia coli. *Curr Opin Biotechnol* **10**, 411-21.
100. Baneyx, F., Mujacic, M. (2004) Recombinant protein folding and misfolding in Escherichia coli. *Nat Biotechnol* **22**, 1399-408.
101. (2003) pET System Manual. *www.novagen.com*.
102. Mihelic, M., Teuscher, C., Turk, V., Turk, D. (2006) Mouse stefins A1 and A2 (Stfa1 and Stfa2) differentiate between papain-like endo- and exopeptidases. *FEBS Lett* **580**, 4195-9.
103. Thannhauser, T.W., Konishi, Y., Scheraga, H.A. (1984) Sensitive quantitative analysis of disulfide bonds in polypeptides and proteins. *Anal Biochem* **138**, 181-8.
104. Kuhelj, R., Dolinar, M., Pungercar, J., Turk, V. (1995) The preparation of catalytically active human cathepsin B from its precursor expressed in Escherichia coli in the form of inclusion bodies. *Eur J Biochem* **229**, 533-9.
105. Strauss, M., Stollwerk, J., Lenarcic, B., Turk, V., Jany, K.D., Gassen, H.G. (1988) Chemical synthesis of a gene for human stefin A and its expression in E. coli. *Biol Chem Hoppe Seyler* **369**, 1019-30.
106. 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**, 331-43.
107. Sabotic, J., Popovic, T., Puizdar, V., Brzin, J. (2009) Macrocypins, a family of cysteine protease inhibitors from the basidiomycete Macrolepiota procera. *FEBS J* **276**, 4334-45.
108. Studier, F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**, 207-34.
109. Chayen, N.E. (2005) Methods for separating nucleation and growth in protein crystallisation. *Prog Biophys Mol Biol* **88**, 329-37.
110. Asherie, N. (2004) Protein crystallization and phase diagrams. *Methods* **34**, 266-72.
111. Chayen, N.E. (1998) Comparative studies of protein crystallization by vapour-diffusion and microbatch techniques. *Acta Crystallogr D Biol Crystallogr* **54**, 8-15.
112. Forsythe, E.L., Maxwell, D.L., Pusey, M. (2002) Vapor diffusion, nucleation rates and the reservoir to crystallization volume ratio. *Acta Crystallogr D Biol Crystallogr* **58**, 1601-5.
113. McPherson, A. (1990) Current approaches to macromolecular crystallization. *Eur J Biochem* **189**, 1-23.
114. Govada, L., Chayen, E. (2009) Crystallization by Controlled Evaporation Leading to High Resolution Crystals of the C1 Domain of Cardiac Myosin Binding Protein-C (cMyBP-C). *Cryst. Growth Des.* **2009**, 3.
115. Galesa, K., Brzin, J., Sabotic, J., Turk, D. (2006) Crystallization and preliminary X-ray crystallographic analysis of the cysteine protease inhibitor clitocypin. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **62**, 10-2.
116. Sabotic, J., Gaser, D., Rogelj, B., Gruden, K., Strukelj, B., Brzin, J. (2006) Heterogeneity in the cysteine protease inhibitor clitocypin gene family. *Biol Chem* **387**, 1559-66.
117. Rhodes, G. (1993) Crystallography Made Crystal Clear. *Academic Press*.
118. Otwinowski, Z., Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.* **276**, 21.

119. Taylor, G. (2003) The phase problem. *Acta Crystallogr D Biol Crystallogr* **59**, 1881-90.
120. Uson, I., Sheldrick, G.M. (1999) Advances in direct methods for protein crystallography. *Curr Opin Struct Biol* **9**, 643-8.
121. Hauptman, H. (1997) Phasing methods for protein crystallography. *Curr Opin Struct Biol* **7**, 672-80.
122. Rossmann, M.G. (1990) The molecular replacement method. *Acta Crystallogr A* **46 (Pt 2)**, 73-82.
123. Bella, J., Rossmann, M.G. (1998) A general phasing algorithm for multiple MAD and MIR data. *Acta Crystallogr D Biol Crystallogr* **54**, 159-74.
124. Ogata, C.M. (1998) MAD phasing grows up. *Nat Struct Biol* **5 Suppl**, 638-40.
125. Pahler, A., Smith, J.L., Hendrickson, W.A. (1990) A probability representation for phase information from multiwavelength anomalous dispersion. *Acta Crystallogr A* **46 (Pt 7)**, 537-40.
126. Dauter, Z., Dauter, M., Dodson, E. (2002) Jolly SAD. *Acta Crystallogr D Biol Crystallogr* **58**, 494-506.
127. Subramaniam, S., Tcheng, D.K., Fenton, J.M. (1996) A knowledge-based method for protein structure refinement and prediction. *Proc Int Conf Intell Syst Mol Biol* **4**, 218-29.
128. Navaza, J., Saludjian, P. (1997) *Methods Enzymol.* , 581-594.
129. Somoza, J.R., Zhan, H., Bowman, K.K., Yu, L., Mortara, K.D., Palmer, J.T., Clark, J.M., McGrath, M.E. (2000) Crystal structure of human cathepsin V. *Biochemistry* **39**, 12543-51.
130. Murshudov, G.N., Vagin, A.A., Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-55.
131. Turk, D. (1992) Weiterentwicklung eines Programms fuer Molekuelgraphik und Elektrondichte-Manipulation and Seine Anwendung auf Verschiedene Protein-Strukturaufklerungen. *Ph.D. thesis, Technische Universitaet Muenchen, Germany*.
132. Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., Terwilliger, T.C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**, 1948-54.
133. Navaza, J., Saludjian, P. (1997) AMoRe: an automated molecular replacement program package. *Methods Enzymol.* **276**, 581-594.
134. Perrakis, A., Morris, R., Lamzin, V.S. (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* **6**, 458-63.
135. Binnig, G., Quate, C.F., Gerber, C. (1986) Atomic force microscope. *Phys Rev Lett* **56**, 930-933.
136. Hinterdorfer, P., Dufrene, Y.F. (2006) Detection and localization of single molecular recognition events using atomic force microscopy. *Nat Methods* **3**, 347-55.
137. Kokkoli, E., Ochsenhirt, S.E., Tirrell, M. (2004) Collective and single-molecule interactions of alpha5beta1 integrins. *Langmuir* **20**, 2397-404.
138. Baumgartner, W., Hinterdorfer, P., Ness, W., Raab, A., Vestweber, D., Schindler, H., Drenckhahn, D. (2000) Cadherin interaction probed by atomic force microscopy. *Proc Natl Acad Sci U S A* **97**, 4005-10.
139. Zhang, X., Bogorin, D.F., Moy, V.T. (2004) Molecular basis of the dynamic strength of the sialyl Lewis X--selectin interaction. *Chemphyschem* **5**, 175-82.
140. Bartels, F.W., Baumgarth, B., Anselmetti, D., Ros, R., Becker, A. (2003) Specific binding of the regulatory protein ExpG to promoter regions of the galactoglucan biosynthesis gene cluster of Sinorhizobium meliloti--a combined molecular biology and force spectroscopy investigation. *J Struct Biol* **143**, 145-52.
141. Evans, E., Ritchie, K. (1997) Dynamic strength of molecular adhesion bonds. *Biophys J* **72**, 1541-55.
142. Merkel, R., Nassoy, P., Leung, A., Ritchie, K., Evans, E. (1999) Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature* **397**, 50-3.
143. Bonanni, B., Kamruzzahan, A.S., Bizzarri, A.R., Rankl, C., Gruber, H.J., Hinterdorfer, P., Cannistraro, S. (2005) Single molecule recognition between cytochrome C 551 and gold-immobilized azurin by force spectroscopy. *Biophys J* **89**, 2783-91.
144. Morrison, J.F. (1982) *Trends Biochem. Sci.* **7**.
145. Henderson, P.J.F. (1972) *Biochem. J.* **127**, 321-333.
146. Renko, M., Sabotic, J., Mihelic, M., Brzin, J., Kos, J., Turk, D. Versatile loops in mycocypins inhibit three protease families. *J Biol Chem* **285**, 308-16.
147. Cheng, T., Hitomi, K., van Vlijmen-Willems, I.M., de Jongh, G.J., Yamamoto, K., Nishi, K., Watts, C., Reinheckel, T., Schalkwijk, J., Zeeuwen, P.L. (2006) Cystatin M/E is a high affinity inhibitor of cathepsin V and cathepsin L by a reactive site that is distinct from the legumain-binding site. A novel clue for the role of cystatin M/E in epidermal cornification. *J Biol Chem* **281**, 15893-9.
148. Johnson, G.D., Jiang, W. (2005) Characterization of cathepsin L secreted by Sf21 insect cells. *Arch Biochem Biophys* **444**, 7-14.
149. Ong, P.C., McGowan, S., Pearce, M.C., Irving, J.A., Kan, W.T., Grigoryev, S.A., Turk, B., Silverman,

G.A., Brix, K., Bottomley, S.P., Whisstock, J.C., Pike, R.N. (2007) DNA accelerates the inhibition of human cathepsin V by serpins. *J Biol Chem* **282**, 36980-6.

150. Bjork, I., Ylinenjarvi, K. (1989) Interaction of chicken cystatin with inactivated papains. *Biochem J* **260**, 61-8.
151. Goulet, B., Baruch, A., Moon, N.S., Poirier, M., Sansregret, L.L., Erickson, A., Bogyo, M., Nepveu, A. (2004) A cathepsin L isoform that is devoid of a signal peptide localizes to the nucleus in S phase and processes the CDP/Cux transcription factor. *Mol Cell* **14**, 207-19.
152. Duncan, E.M., Muratore-Schroeder, T.L., Cook, R.G., Garcia, B.A., Shabanowitz, J., Hunt, D.F., Allis, C.D. (2008) Cathepsin L proteolytically processes histone H3 during mouse embryonic stem cell differentiation. *Cell* **135**, 284-94.
153. Ceru, S., Konjar, S., Maher, K., Repnik, U., Krizaj, I., Bencina, M., Renko, M., Nepveu, A., Zerovnik, E., Turk, B., Kopitar-Jerala, N. Stefin B interacts with histones and cathepsin L in the nucleus. *J Biol Chem* **285**, 10078-86.
154. Nagler, D.K., Storer, A.C., Portaro, F.C., Carmona, E., Juliano, L., Menard, R. (1997) Major increase in endopeptidase activity of human cathepsin B upon removal of occluding loop contacts. *Biochemistry* **36**, 12608-15.
155. Turk, D., Podobnik, M., Kuhelj, R., Dolinar, M., Turk, V. (1996) Crystal structures of human procathepsin B at 3.2 and 3.3 Angstroms resolution reveal an interaction motif between a papain-like cysteine protease and its propeptide. *FEBS Lett* **384**, 211-4.
156. Podobnik, M., Kuhelj, R., Turk, V., Turk, D. (1997) Crystal structure of the wild-type human procathepsin B at 2.5 A resolution reveals the native active site of a papain-like cysteine protease zymogen. *J Mol Biol* **271**, 774-88.
157. Turk, D., Guncar, G., Podobnik, M., Turk, B. (1998) Revised definition of substrate binding sites of papain-like cysteine proteases. *Biol Chem* **379**, 137-47.
158. de Odrowaz Piramowicz, M., Czuba, P., Targosz, M., Burda, K., Szymonski, M. (2006) Dynamic force measurements of avidin-biotin and streptavdin-biotin interactions using AFM. *Acta Biochim Pol* **53**, 93-100.
159. Panorchan, P., Thompson, M.S., Davis, K.J., Tseng, Y., Konstantopoulos, K., Wirtz, D. (2006) Single-molecule analysis of cadherin-mediated cell-cell adhesion. *J Cell Sci* **119**, 66-74.
160. Krissinel, E., Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256-68.
161. Lee, J., Dubey, V.K., Somasundaram, T., Blaber, M. (2006) Conversion of type I 4:6 to 3:5 beta-turn types in human acidic fibroblast growth factor: effects upon structure, stability, folding, and mitogenic function. *Proteins* **62**, 686-97.
162. Ravichandran, S., Sen, U., Chakrabarti, C., Dattagupta, J.K. (1999) Cryocrystallography of a Kunitz-type serine protease inhibitor: the 90 K structure of winged bean chymotrypsin inhibitor (WCI) at 2.13 A resolution. *Acta Crystallogr D Biol Crystallogr* **55**, 1814-21.
163. Fujishima, A., Imai, Y., Nomura, T., Fujisawa, Y., Yamamoto, Y., Sugawara, T. (1997) The crystal structure of human cathepsin L complexed with E-64. *FEBS Lett* **407**, 47-50.
164. 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**, 1197-202.
165. Fitzgerald, C.E., Patel, S.B., Becker, J.W., Cameron, P.M., Zaller, D., Pikounis, V.B., O'Keefe, S.J., Scapin, G. (2003) Structural basis for p38alpha MAP kinase quinazolinone and pyridol-pyrimidine inhibitor specificity. *Nat Struct Biol* **10**, 764-9.
166. Micheelsen, P.O., Vevodova, J., De Maria, L., Ostergaard, P.R., Friis, E.P., Wilson, K., Skjot, M. (2008) Structural and mutational analyses of the interaction between the barley alpha-amylase/subtilisin inhibitor and the subtilisin savinase reveal a novel mode of inhibition. *J Mol Biol* **380**, 681-90.
167. Vallee, F., Kadziola, A., Bourne, Y., Juy, M., Rodenburg, K.W., Svensson, B., Haser, R. (1998) Barley alpha-amylase bound to its endogenous protein inhibitor BASI: crystal structure of the complex at 1.9 A resolution. *Structure* **6**, 649-59.
168. Jongsma, M.A., Bakker, P.L., Peters, J., Bosch, D., Stiekema, W.J. (1995) Adaptation of Spodoptera exigua larvae to plant proteinase inhibitors by induction of gut proteinase activity insensitive to inhibition. *Proc Natl Acad Sci U S A* **92**, 8041-5.
169. De Leo, F., Bonade-Bottino, M.A., Ceci, L.R., Gallerani, R., Jouanin, L. (1998) Opposite effects on spodoptera littoralis larvae of high expression level of a trypsin proteinase inhibitor in transgenic plants. *Plant Physiol* **118**, 997-1004.
170. Cloutier, C., Jean, C., Fournier, M., Yelle, S., Michaud, D. (2000) Adult Colorado potato beetles, Leptinotarsa decemlineata compensate for nutritional stress on oryzacystatin I-transgenic potato plants by hypertrophic behavior and over-production of insensitive proteases. *Arch Insect Biochem Physiol* **44**,

69-81.

Index of Figures

Index of Tables

Appendix

Published articles:
Miha Renko, Jerica Sabotič, Marko Mihelič, Jože Brzin, Janko Kos, and Dušan Turk: Versatile Loops in Mycocypins Inhibit Three Protease Families *J. Biol. Chem. 2010 285: 308-316. First Published on October 21, 2009.*

Slavko Čeru, Špela Konjar, Katarina Maher, Urška Repnik, Igor Križaj, Mojca Benčina, Miha Renko, Alain Nepveu, Eva Žerovnik, Boris Turk, and Nataša Kopitar-Jerala: Stefin B Interacts with Histones and Cathepsin L in the Nucleus *J. Biol. Chem. 2010 285: 10078-10086. First Published on January 14, 2010.*

Submitted articles:
Miha Renko, Urška Požgan, Dušana Majera, Dušan Turk: The extent of displacement of the occluding loop of cathepsin B corresponds to the size of a ligand *FEBS J.*

Article in preparation:
Miha Renko, Vesna Hodnik, Marko Mihelič, Gregor Anderluh, Peter Hinterdorfer, Dušan Turk: Interactions between endopeptidase members of cathepsin family and stefin A

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 285, NO. 1, pp. 308–316, January 1, 2010

Versatile Loops in Mycocypins Inhibit Three Protease Families

Received for publication, July 10, 2009, and in revised form, September 30, 2009 Published, JBC Papers in Press, October 21, 2009, DOI 10.1074/jbc.M109.043331

Miha Renko[‡1], Jerica Sabotič[§1], Marko Mihelič[‡], Jože Brzin[§], Janko Kos[§], and Dušan Turk[‡2]

From the Departments of [‡]Biochemistry and Molecular and Structural Biology and [§]Biotechnology, Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia

Mycocypins, clitocypins and macrocypins, are cysteine protease inhibitors isolated from the mushrooms *Clitocybe nebularis* and *Macrolepiota procera*. Lack of sequence homology to other families of protease inhibitors suggested that mycocypins inhibit their target cysteine protease by a unique mechanism and that a novel fold may be found. The crystal structures of the complex of clitocypin with the papain-like cysteine protease cathepsin V and of macrocypin and clitocypin alone have revealed yet another motif of binding to papain like-cysteine proteases, which in a yet unrevealed way occludes the catalytic residue. The binding is associated with a peptide-bond flip of glycine that occurs before or concurrently with the inhibitor docking. Mycocypins possess a β-trefoil fold, the hallmark of Kunitz-type inhibitors. It is a tree-like structure with two loops in the root region, a stem comprising a six-stranded β-barrel, and two layers of loops (6 + 3) in the crown region. The two loops that bind to cysteine cathepsins belong to the lower layer of the crown loops, whereas a single loop from the crown region can inhibit trypsin or asparaginyl endopeptidase, as demonstrated by site-directed mutagenesis. These loops present a versatile surface with the potential to bind to additional classes of proteases. When appropriately engineered, they could provide the basis for possible exploitation in crop protection.

Inhibition of foreign protease activity is a widespread defense mechanism in plants against their pests, pathogens, and parasites (1). Protein inhibitors of proteases are present in a variety of plant tissues. They can be deployed alone or together with a variety of small molecules (2). It has been known for a long time that the expression of protease inhibitors is increased in injured plant leaves (3) and that their expression can be induced as a response to attack by insects or pathogens (2).

Given the negative environmental effects of chemical pesticides used in crop protection, it is important to explore alternative approaches, such as the incorporation of genes encoding protease inhibitors into plants. Transgenic plants expressing various protease inhibitors have shown enhanced levels of insect resistance; however, the adaptive capacity of insect digestive proteases limits the use of single protease inhibitors (4, 5). The use of hybrid protease inhibitors with multiple inhibitory activity could, however, affect the functional properties of the fused inhibitors (6). Incorporation of genes encoding a range of protease inhibitors is to run the risk of deleterious modification of plants, but the use of a single protease inhibitor with versatile functionality could be the way forward.

With these in mind, we have undertaken structural and mechanistic studies of the cysteine protease inhibitors clitocypin (Clt)[3] from mushroom Basidiomycetes *Clitocybe nebularis* and macrocypins (Mcp) from *Macrolepiota procera*. Based on the lack of sequence similarity to other protease inhibitors, they form separate protease inhibitor families, I48 and I85, in the MEROPS classification (7) and were named mycocypins. They are ≈17-kDa proteins exhibiting high thermal and broad pH range stability, with completely reversible unfolding (8, 9). They all inhibit endopeptidases from the papain family, such as papain, cathepsins L, V, S, and K, in the low nanomolar range, and exopeptidases with higher inhibition constants. They exhibit different inhibitory specificities. The cysteine protease asparaginyl endopeptidase (AEP, legumain) is inhibited in the low nanomolar range by macrocypin 1 and macrocypin 3, whereas clitocypin inhibits AEP in the higher nM range. Macrocypin 4 does not inhibit AEP, but in contrast to the others, it inhibits serine protease trypsin in the micromolar range (8, 9).

In this respect mycocypins are similar to cystatins and thyropins. Cystatins inhibit cysteine papain-like proteases and AEP. The crystal structures have revealed that cystatins bind to papain-like proteases with a wedge composed of three regions, N-terminal trunk and two β-hairpin loops (10, 11), whereas their binding geometry to AEP is still unknown. Thyropins inhibit papain-like cysteine and aspartic proteases (12, 13). The crystal structure of the p41 fragment bound to cathepsin L (14) has revealed similarity of the three binding loops to those of cystatins. The three-loop mode of binding is shared also by chagasin, the parasite inhibitor of papain-like proteases from *Trypanosoma cruzi* (15). Common to them all is the fact that the binding loops bind into the non-primed and primed substrate binding regions and occlude the catalytic cysteine residue but do not interact directly with it. *Bauhinia bauhinioides* cruzipain inhibitor BbCI (16), a Kunitz-type inhibitor belonging to the I3 MEROPS family, inhibits papain-like endopeptidases

* This work was supported by Slovenian Research Agency Grants P1-0048 (to D. T.) and P4-0127 (to J. K.), by a Young Researcher fellowship (to M. R.), and by the European Commission Framework VI Program (CAMP project, LSHG-2006-018830).

The atomic coordinates and structure factors (codes 3H6R, 3H6Q, and 3H6S) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[1] Both authors contributed equally to this work.

[2] To whom correspondence should be addressed. Tel.: 38614773857; Fax: 38614773984; E-mail: dusan.turk@ijs.si.

[3] The abbreviations used are: Clt, clitocypin; Mcp, macrocypin; AEP, asparaginyl endopeptidase; BbCI, *B. bauhinioides* cruzipain inhibitor; STI, soybean trypsin inhibitor; WBCI, winged bean chymotrypsin inhibitor; ECI, inhibitor isolated from *E. caffra*; Cst, cystatin; Bistris propane, 1,3-bis-[tris(hydroxymethyl)methylamino]propane; MES, 4-morpholineethanesulfonic acid; r.m.s., root mean square.

cruzipain and cathepsin L and serine protease trypsin but not endopeptidase cathepsin V or exopeptidases cathepsins B and X. The structure of BbCI protease complexes is not known, although it has been suggested that the same reactive loop is involved in inhibition of cysteine and serine proteases (17).

To gain an insight into the inhibitory mechanism of mycocypins, we have determined the crystal structures of the complex of clitocypin with the papain-like cysteine protease cathepsin V and of macrocypin 1 and clitocypin alone. The study has revealed yet another motif for binding to papain-like cysteine proteases that occludes the catalytic residue. The fold of mycocypins is based on a six-stranded β-barrel that composes the core of the β-trefoil fold, providing a versatile surface capable of binding to various protease types. We have identified their binding regions to AEP and trypsin by site-directed mutagenesis and shown that the loops binding to papain-like proteases are different from those binding to AEP and trypsin.

EXPERIMENTAL PROCEDURES

Protein Expression and Isolation—Natural clitocypin was purified from the fruiting bodies of Basidiomycetes *C. nebularis* (9). Macrocypin, clitocypin, methionine-containing clitocypin mutant (Clt-L82M, Clt-I89M), and their mutants (Mcp-1/4 (Mcp1 with β5-β6 loop of Mcp4), Mcp-4/1 (Mcp4 with β5-β6 loop of Mcp1), Mcp-ΔG25, Mcp-G25A, Mcp4-N74R, and Clt-ΔG24, Clt-G24A, Clt-N69K) were expressed in *Escherichia coli* (8, 18). Mutants were produced by PCR site-directed mutagenesis using the appropriate pET vectors as templates followed by digestion with DpnI (Fermentas) and recovery of the vectors containing mutated inserts (19). The selenomethionine mutant was produced using minimum autoinduction media with the addition of selenomethionine in *E. coli* BL384 cells (20). Cathepsin V was expressed in *Pichia pastoris* (21).

Crystallization—Recently we have reported crystallization conditions and phasing attempts using clitocypin purified from natural source (22). 1-ml drops of 15 mg/ml solution of clitocypin in 15 mM MES buffer, pH 6.0, gave crystals when mixed with 1 ml of crystallization buffer (50 mM monopotassium dihydrogen phosphate, 20% (w/v) polyethylene glycol 8000, pH 3.76) using the vapor diffusion method. A number of data sets with a variety of resolutions were collected. The structure could not be solved because of the high heterogeneity of the natural clitocypin and unsuccessful derivatization of the crystals, whereas the absence of significant sequence similarity (23) to other proteins with known structures discouraged molecular replacement attempts.

The initial crystallization screening for clitocypin alone and in complex with the target proteases was performed with the recombinant double methionine (L82M,I89M) mutant of clitocypin, its selenomethionine mutants, and cathepsins L and V. In contrast to the natural clitocypin, the recombinant clitocypin alone gave no crystals. Of the complexes, only cathepsin V and the methionine mutant produced diffracting crystals. Cathepsin V and clitocypin were mixed in molar ratio 1:1.1 and concentrated to 50 mg/ml in 10 mM acetate, 100 mM NaCl, pH 5.5. Crystals of dimensions of $0.2 \times 0.4 \times 0.1$ mm^3 were obtained in 0.4 M Li_2SO_4, 12% polyethylene glycol 800, 20% glycerol after 4 months. The selenomethionine mutant of clitocypin in complex with cathepsin V gave better diffracting crystals in a much shorter time. The crystals were frozen in liquid nitrogen before data collection. Diffraction data were collected at Synchrotrone Elletra, Trieste, from a single crystal using wavelength 1.0 Å.

Macrocypin 1 was concentrated to 30 mg/ml in 10 mM acetate buffer, 200 mM NaCl, pH 5.5, and crystallized by the sitting drop method at 20 °C using commercial screens from Qiagen. Crystals grew overnight in various conditions (Bistris propane buffer, pH 6.5–7.5, 100–500 mM different sodium salts, 20% polyethylene glycol 3350 or 0.8–1.6 M sodium/potassium phosphate, pH 7.0–8.0). The best diffracting crystals were grown in Bistris propane buffer, pH 7.0, 200 mM NaI, 20% polyethylene glycol 3350. They were soaked in a saturated solution of NaI in the same buffer before flash-freezing. Diffraction data were collected on an in-house copper rotating anode Rigaku (RU 200). Another high resolution data set was collected at Synchrotrone Elletra, Trieste, Italy, from a crystal grown in Bistris propane buffer, pH 7.0, 200 mM sodium citrate, 20% polyethylene glycol 3350 and soaked in the saturated solution of sodium citrate in the same buffer. All crystals were larger than 0.5 mm in all three dimensions.

Structure Solution and Refinement—All data were processed using the HKL2000 package (24). The macrocypin structure was solved with single wavelength anomalous diffraction phasing from the data collected from the crystal soaked in the saturated solution of sodium iodide. The data set was collected to 2.2 Å on the in-house Rigaku rotating anode (RU 200) using Xenox mirrors. 615 images were collected from a single crystal with the linear R-merge of 13% and redundancy of 30. Single wavelength anomalous diffraction phasing was based on 15 iodine positions with occupancy ranging from 0.8 to 0.15 using automated SOLVE/RESOLVE scripts incorporated in the AutoSol module of the PHENIX suite (25). Automated model building and docking to the macrocypin sequence gave a solution with ~120 of 159 amino acids built (data not shown). Despite the good data quality, we were unable to refine the structure, presumably due to the multiple conformations of several loop regions induced by the binding of iodine ions, quite a few of them with low occupancy positions inside the protein core. Therefore, another data set was collected with the crystals, grown in sodium citrate. We phased this data set with the partial structure of macrocypin from the iodine-soaked crystal using molecular replacement program Amore (26). Cycles of manual and automated building with ARP/warp (27) and refinement with Refmac (28) and MAIN (29) were performed until all residues were built in the electron density. The final structure was refined using MAIN against 1.64 Å resolution data (29).

The crystal structure of cathepsin V-clitocypin complex was determined by molecular replacement with Amore using cathepsin V (PDB code 1fh0) (30) as the search model. Four molecules of cathepsin V were positioned into the asymmetric unit. The 4-fold electron density averaging in MAIN (29) produced maps that enabled us to build manually substantial parts of the clitocypin structure. Fragments of the clitocypin model enabled manual superimposition of the macrocypin structure using the similarity between the two models, acceler-

TABLE 1

Data collection and refinement statistics

Numbers in parentheses are for the highest resolution shell. Data sets from only one crystal were used for determination of each structure.

	Macrocypin	Clitocypin-Cathepsin V	Clitocypin
Data collection			
PDB ID	3H6Q	3H6S	3H6R
Space group	$P3_121$	$P2_12_12$	$P2_1$
Cell dimensions			
a, *b*, *c* (Å)	77.1, 77.1, 60.9	98.2, 177.8, 60.9	46.5, 58.0, 58.3
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 111.2, 90
Resolution (Å)	50-1.64	27.5-2.22	30-1.94
R_{merge} (%)	5.0 (14.2)	3.7 (19.4)	2.6 (9.8)
$I/\sigma I$	69.0 (13.1)	43.9 (6.1)	63.1 (16.8)
Completeness (%)	98.6 (86.5)	98.8 (76.8)	97.9 (85.0)
Redundancy	10.3 (8.3)	3.9 (1.8)	7.2 (6.9)
Refinement			
Resolution	27.7-1.64	27.5-2.22	30-1.948
No. of reflections (work/free)	24.490/1.282	76.800/4.054	20.034/1.040
R_{work}/R_{free}	16.2/19.3	18.3/23.4	18.6/24.1
B factors			
Protein	19.0	26.5	28.7
Water	35.4	37.3	40.1
No. of atoms			
Protein	1424	11500	2388
Water	321	856	278
r.m.s. deviation			
Bond lengths (Å)	0.019	0.018	0.022
Bond angles (°)	2.00	1.70	2.05

ating the model building. The positions of the two selenomethionine residues in the clitocypin sequence were helpful in the initial sequence assignment. The structure was refined using MAIN against 2.24 Å resolution data. Geometric parameters for S-CH_3 bound to the active site cysteine residue were obtained from PURY server (31). Data collection and refinement statistics are summarized in Table 1."

Kinetic Measurements—Kinetic and equilibrium constants for the inhibition of cathepsin V were determined under pseudo-first order conditions in continuous kinetic assays at 25 °C and calculated by nonlinear regression analysis according to Morrison (32) or Henderson (33) as described previously (8, 9).

RESULTS

Structures of Macrocypin and Clitocypin—Macrocypin 1 crystallized in $P3_121$ space group with one molecule in the asymmetric unit. The macrocypin crystals contain the complete sequence, numbered from Gly-1 to Glu-168. Positioning of nearly all the residues is clearly revealed by the electron density maps. The exceptions are the side chains of His-114, Tyr-140, and Lys-21, which are only partially defined, and the stretch of residues Ser-20, Lys-21, Ile-22, which is only loosely defined. Nine residues (Gly-1, His-17, Arg-55, Ile-75, Gln-78, Ser-80, Glu-100, Gln-110, Ile-158) were modeled in alternative conformations.

Clitocypin crystallized in the $P2_1$ space group with two molecules in the asymmetric unit. The positioning of nearly all the residues is clearly revealed by the electron density maps. The exceptions are the loop Gln-67—Tyr-75 in molecule 1, Gly-68—Asn-70 and the side chain of Gln-115 in molecule 2, and the first two N-terminal residues in both molecules. Because of the genetic heterogeneity, clitocypin isolated from the natural source contains a large number of isoforms in unknown ratios (23). As default we have, therefore, used the sequence of the clitocypin isoform used for the complex formation. However, when the electron density unambiguously showed disagreement with that sequence, we built an appropriate amino acid residue from an alternative sequence based on amino acid sequences deduced from clitocypin genetic data (H17S, Y62S, L82M, P84Q, I88M, A105T, T139N in both molecules, S46F and Q48R in molecule 1, and Q37K, N42S, and A57S in molecule 2).

The macrocypin and clitocypin structures have the same fold. In the projection used in Fig. 1*A* the fold is reminiscent of a tree with a short, thick trunk, and a crown with branches expanding far from the center. The trunk part is an up-and-down β-barrel composed of six antiparallel β-strands (β1, β4, β5, β8, β9, β12). The strands are laid at an angle of less than 45 degrees to the axis of the barrel. The N and C termini are at the bottom. They form the roots of the tree together with the loops connecting strands β4-β5 and β8-β9. On the top three long regions between the strands β1-β4, β5-β8, and β9-β12 constitute the tree crown. Each contains a pair of antiparallel β-strands. In this manner two additional loops are formed between the strands from the crown and the trunk, adding another layer of loops that spread away from the trunk. Hence, the 3-fold arrangement of loops and strands is preserved in four layers of the structure: in the roots, the trunk, and the lower and upper layers of crown (Fig. 1, *B–D*). The loop region before strand β8 in the lower crown layer of macrocypin is folded into a short three-turn α-helix, whereas in clitocypin the loop preceding the strand β4 contains a short helical region. Although macrocypin and clitocypin have the same fold, the r.m.s. deviation between 116 equipositioned Cα atoms is 1.75 Å. The β-barrels are more similar, yielding r.m.s. deviations of 0.67 Å between 31 equipositioned Cα atoms. Macrocypin and clitocypin have a pseudo-3-fold symmetry (Fig. 1*B–D*), with the 3-fold rotational axis running through the six-stranded barrel.

The structure of both clitocypin molecules in the asymmetric unit is basically the same (r.m.s. value of 0.28 Å), with one important exception. The peptide bond between Gly-24 and Gly-25 residues appears in two different orientations that are clearly seen in the electron density (Fig. 2, *A* and *B*), both in glycine-preferred regions of the Ramachandran plot. This suggests that this peptide bond is flexible and can appear in either orientation.

When the macrocypin structure was submitted to the protein structure comparison service SSM (34) at the European Bioinformatics Institute, the fold was identified as the β-trefoil fold present in proteins such as Kunitz-type soybean trypsin inhibitor (STI) (35), inhibitor isolated from *Erythrina caffra* (36), interleukins-1α and -1β (37), and fibroblast growth factors (38). The sequence alignment shows the low similarity of these proteins, in contrast to the structural alignment (Table 2), which shows that the secondary structure patterns are quite similar. The number of β-strands in macrocypin (12 strands) and clitocypin is the same as in STI, whereas their lengths differ significantly. STI has only two short strands (β6 and β7), whereas macrocypin and clitocypin have four (β2, β3, β6, β7). The highest structural as well as sequence similarity is in the regions composing the β-barrel. The fact that these are rather short stretches of sequence explains why homology searches based on sequence alignment have failed.

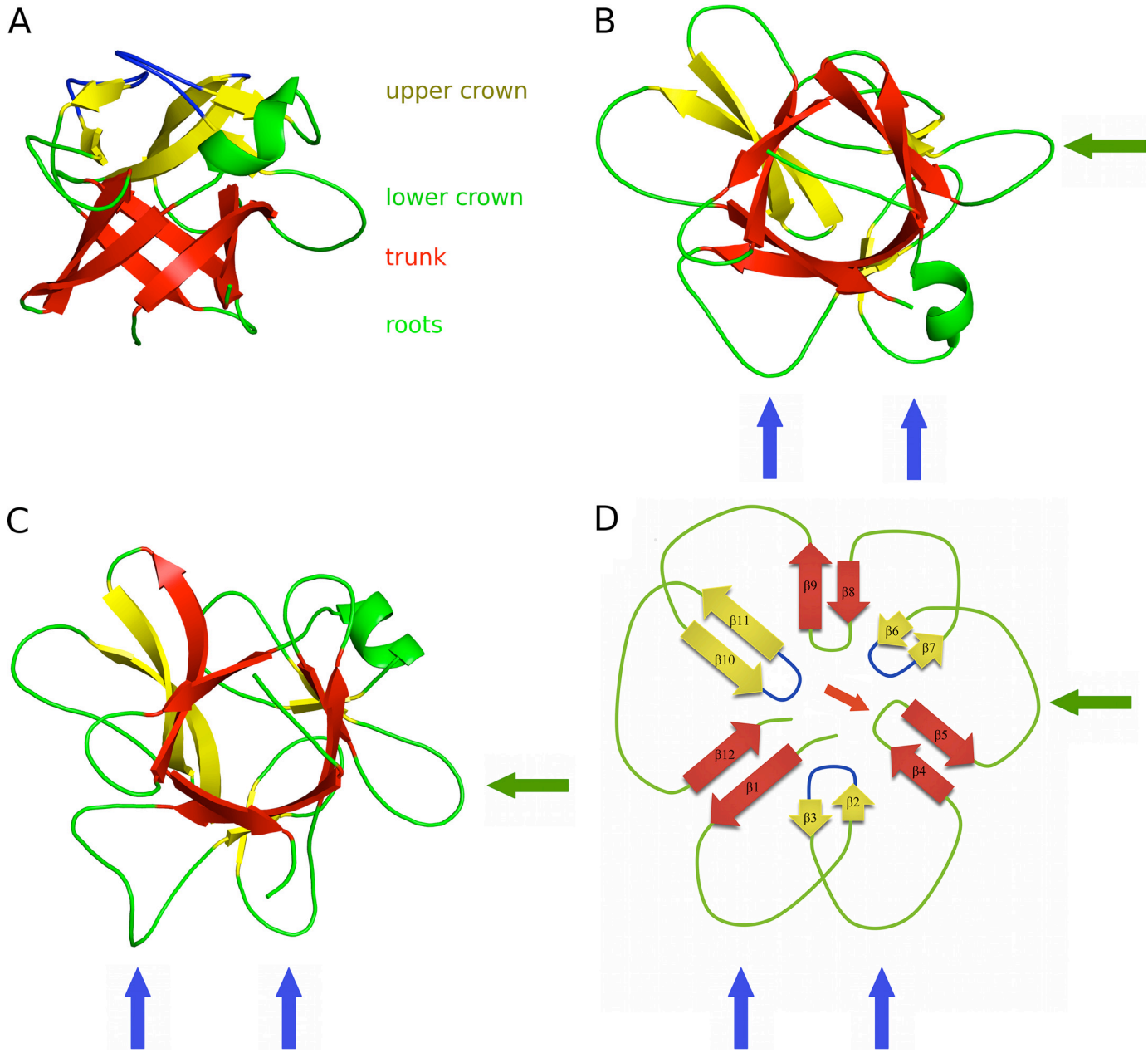


FIGURE 1. **Structures of clitocypin (*A* and *B*), macrocypin (*C*), and schematic representation of trefoil fold (*D*).** The trunk strands are always shown in *red*, and the crown strands are in *yellow*. *A*, the view from the side is shown. The structure resembles a tree structure, with two loops in the root region, a stem built of a six-stranded β-barrel, and two layers of loops (6 + 3) in the crown region. The first layer of the crown loops is shown in *green*, and the second is in *blue*. *B* and *C*, shown is the view along the barrel. The binding loops for cathepsins are marked with *blue arrows*, and the AEP loop is marked with *green arrows*. The two loops that bind to cysteine cathepsins belong to the lower layer of crown loops, whereas a single loop from the root region can inhibit trypsin or AEP.

Structures of the Cathepsin V-Clitocypin Complex—The complete protein sequences are seen in the crystal structure of the complex of clitocypin with cathepsin V. The catalytic site of cathepsin V was blocked with a methyl methanethiosulfonate, leaving the S-CH_3 group on the active site cysteine residue. This form of cathepsin V is much more stable and resulted in better diffracting crystals. The crystals have the $P2_12_12$ space group and contain four pairs of molecules per asymmetric unit. The four structures of cathepsin V and clitocypin are closely related. The r.m.s. deviations over equivalent Cα atoms range from 0.18 to 0.25 for cathepsin V and 0.33 to 0.50 for clitocypin. Although the cathepsin V structures, apart from the ends of a few side chains, are unambiguously resolved from the electron density maps, three N-terminal residues and two loop regions (Gln-67—Asn-70 in molecules 1, 3, and 4 and Asp-138—Gly-141 in molecules 2, 3, and 4) of clitocypin lack adequate electron density or are only loosely defined.

Clitocypin binds into the active site of the target protease in the orientation of a fallen tree, with trunk and roots pointing sideways and up (Fig. 3, *A* and *B*). The wedge-shaped structure fills the active site cleft along its whole length, resulting in a buried area of 825 $Å^2$. The interaction surface of clitocypin comprises basically two broad loop regions positioned at the lower edge of the crown. The loop structure and binding geom-

etry are stabilized by numerous hydrogen bonds. Both loops originate from the first third of the clitocypin sequence; the first loop connects strands β1 and β2, and the second loop strands β3 and β4. The first loop is a broad, lasso-like structure cross-connected in the middle with a hydrogen bond between the Arg-12 side chain and the Gly-22 carbonyl. The second loop region is narrower and contains a short helix. The first loop binds into the non-primed substrate, and the second loop binds into the primed substrate binding site (Fig. 4). They occlude the catalytic cysteine in the middle and thereby prevent the approach of substrate molecules. Because the reactive site Cys-25 in the structure is modified, we cannot exclude the possibility that the interaction of clitocypin with the naked cysteine cathepsin may deviate slightly from the observed one.

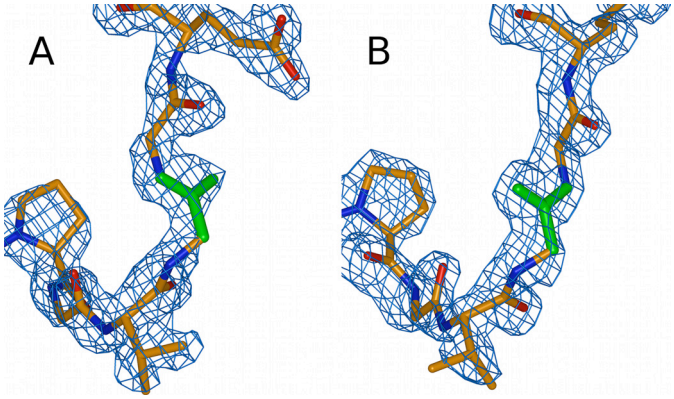


FIGURE 2. **Orientation of the Gly-24-Gly-25 peptide bond in the two clitocypin molecules.** The $2F_{obs} - F_{calc}$ electron density map is contoured at 1 σ. The bonds of glycine 24 are shown in *green*, whereas the rest of the chain is shown in *red* for oxygen, *blue* for nitrogen, and *orange* for carbon. The Gly-24–Gly-25 peptide bond is flexible and can exist in either orientation.

The chain from the first loop region comes down the S3 binding area (Pro-21, Gly-22) of cathepsin V, occupies the S2 binding site with Val-23, and continues through the S1 binding site, upwards and away from the cathepsin V surface. A hydrogen bond from the side chain amide of Asn-18 attaches clitocypin to the Gln-63 side chain of cathepsin V. A water molecule mediates additional contacts in the S3 binding area. Arg-12 plays a dual role; it stabilizes the lasso as well as attaching it to the cathepsin V surface by the Asn-66 main chain carbonyl, filling the interaction surface between the S3 binding area and the S1 binding site. Positioning of Val-23 builds antiparallel hydrogen bond arrangements with the Gly-68 of cathepsin V in a substrate-like manner. The peptide Val-23–Gly-24 bond is additionally fixed by the interaction between its hydrogen and the Asp-163 carbonyl from cathepsin V. In the S1 binding site, Gly-23 of cathepsin V is involved in hydrogen bonds between its carbonyl and the Gly-25 amide and between its amide hydrogen and the carboxylic oxygen atom of Glu-26. The peptide bond between Gly-24 and Gly-25 is flipped when compared with the structure of molecule 2 of clitocypin and macrocypin, presumably to form a hydrogen bond with Gly-23 of cathepsin V (Fig. 3). The other carboxylic oxygen atom of Glu-26 interacts with the amide of the cathepsin V Asn-66 side chain. However, the possibility that Glu-26 is partially neutral and interacts with the carbonyl oxygen atom of Gln-21 cannot be excluded. The

TABLE 2

Structural alignment of sequences of clitocypin, macrocypin, and selected β-trefoil serine protease inhibitors

Sequences were aligned by the Protein structure comparison service SSM at the European Bioinformatics Institute (33). The parts of the sequence belonging to β-strands are printed in white on a black background, the binding loops are in white on a gray background, and the helical parts are in black on a gray background.

```
                    10        20        30        40        50        60        70        80
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Clt       --ASL-------ED---GTYRLRAVTTHNPDPGV--GGEYAT -EG----ARQPVKAEPSTP-PFFERQIWQVTRNSD-------
Mcp       ---GF-------ED---GFYTILHLAEGQHPNSKIPGGMYASSKDG----KDVPVTAEP-LGPQS--KIRWWIARDPQAG----D
STI       --DFVLDNEGNPLEN-GGTYYILS ------IT---AFGGIRAAPTGNERCPLTVVQSR-NE-LDK-GIGTIISSPYRIRFIAEG
BbCI      HMSVILDTKGEPVSNAADAYYLVPV------S----GEGGLALAKVGNEAEPKAVVLDP-H---HRPGLTVRFETPLAIAIITES
ECI       ---VLLDGNGEVVQN-GGTYYLLPQVW---AQ-----GGGVQLAKTGEETCPLTVVQSP-NE-LSD-GKPIRIESRLRSAFIPDD
WBCI      -DDDLVDAEGNLVEN-GGTYYLLPH------IWA--HGGGIETAKTGNEPCPLTVVRSP-NE-VSK-GEPIRISSQFLSLFIPRG
                              ß1                  ß2              ß3              ß4

                    90       100       110       120       130       140       150       160       180
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Clt       GQYTIKYQGLNA------PFEYGFSYD--QLE-QNAPVIAG--DP---------KEYILQLV-PS-TA-DVYIIRAPI------QR
Mcp       DMYTITEF-RID-----NSIPGQWSRSPVET---EVPVYLY--DRIKAEETGYTCAWRIQPADH--GADGVYHIVGNV------R-
STI       HPLSLKFD-SFAVIMLCVGIPTEWSVV--EDLPEGPAVKIG--ENKDA----MDGWFRLERV-SE-FN--NYKLVFCPQ----QD-
BbCI      FFLNIKFV-PSS-----SDSE-VWDVSKQ--YPIGLAVKVT--DTKS-----FVGPFRVEKE-G--E---GYKIVYYPDR---GQ-
ECI       DKVRIGFA-YAPK----CAPSPWWTVV--E----GLSVKLSEDESTQ-----FDYPFKFEQV-SDQLH--SYKLLYCEGK---HE-
WBCI      SLVALGFA-NPPS----CAASPWWTVVDS--P-QGPAVKLSQQKLPEK----DILVFKFEKV-SH-SNIHVYKLLYCQHDEEDVK-
              ß5                  ß6        ß7                  ß8              ß9

                              190       200       210       220
                              ....|....|....|....|....|....|....|....|....|
Clt       VG--VDVEVGVQ ------NNLVYKFFPVDGSGGDRP-AWRFTRE----
Mcp       --IGSTDWADLREEY-- EPQVYMK-PVPVIPNVYIP-RWFILGYEE--
STI       --K--CGDIGISIDHDDGTRRLVVS-K-------NKPLVVQFQKLD---
BbCI      --T--GLDIGLVHRN--DKYYLAAT-E-------GEPFVFKIRKAT---
ECI       --K--CASIGINRDQK-GYRRLVVT-E-------DYPLTVVLKKDE---
WBCI      --C--DQYIGIHRDRN-GNRRLVVT-E-------ENPLELVLLKAKSET
                    ß10       ß11             ß12
```

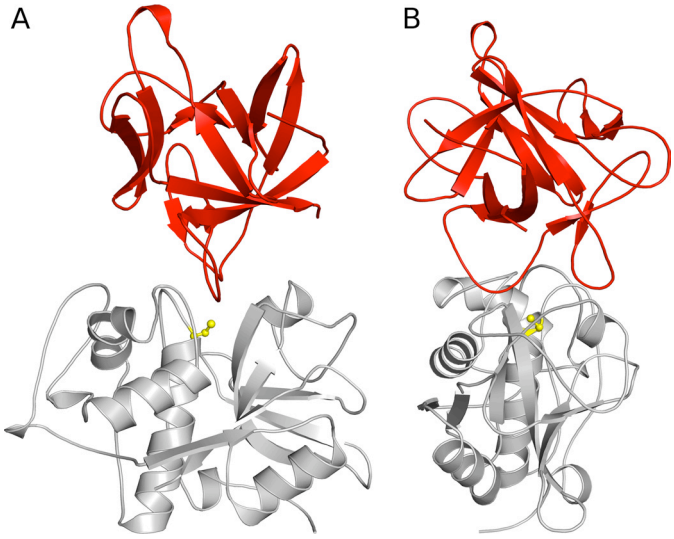


FIGURE 3. **The cathepsin V-clitocypin complex.** *A*, shown is the view along the active site cleft. *B*, shown is the view perpendicular to the active site cleft. The folds of cathepsin V and clitocypin are shown in *gray* and *red*. The catalytic cysteine is shown in *yellow*. Clitocypin binds into the active site of cathepsin V in the orientation of a fallen tree with the trunk and roots pointing sideways and up. The wedge shaped structure fills the active site cleft along its whole length.

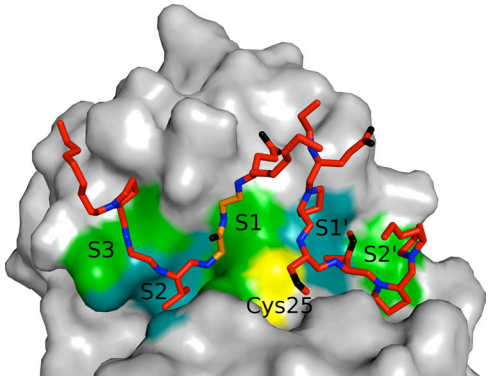


FIGURE 4. **Binding loops of clitocypin in its complex with cathepsin V.** Clitocypin loops are shown as *sticks*. Only main chain atoms, without the carbonyl oxygen atom and side chains, are shown. Nitrogen atoms are shown in *blue*, oxygen is in *black*, and *carbon* is in *red*, with the exception of the Gly-24-Gly-25 part, shown in *orange*. The surface of cathepsin V is shown in *gray*, apart from the catalytic cysteine shown in *yellow* and the S3, S2, S1, S1′ and S2′ binding site, shown in *green* and *cyan*. The chain of the first binding loop comes down the S3 binding area of cathepsin V, occupies the S2 binding site, and continues upward through the S1 binding site. The second binding loop of clitocypin approaches the S1′ and S2′ binding sites of cathepsin V from the *top*.

binding of the first loop is further stabilized by the side chain amide group of Asn-18, which forms a hydrogen bond with the peptide bond carbonyl atom of Gln-21 of cathepsin V. It is notable that, of the hydrogen bond interactions between the enzyme and inhibitor, most are contributed by main chain atoms, at least on one side.

The second binding loop of clitocypin approaches the S1′ and S2′ binding sites of cathepsin V from the top. A single hydrogen bond between the carbonyl of Ser-42 and the side chain amide of Gln-145 of cathepsin V fastens the loop to the cathepsin V surface. In the first complex, an additional hydrogen bond is formed between the carboxylic group of Glu-48 and cathepsin V Ser-142. A layer of solvent molecules mediates the

TABLE 3

Inhibition constants of cathepsin V by macrocypin, clitocypin, and their mutants

Kinetic data for interaction of macrocypin 1 and cathepsin V were reported previously (9).

	$10^{-6}\ k_{on}$	$10^{4}\ k_{off}$	K_i
	$M^{-1}s^{-1}$	s^{-1}	nM
Clt	1.26 ± 0.07	1.61 ± 0.65	0.08 ± 0.03
Clt ΔG24	0.10 ± 0.06	1.8 ± 0.9	1.9 ± 0.9
Clt G24A	0.08 ± 0.01	1.2 ± 0.6	1.6 ± 0.7
Mcp1	1.48 ± 0.01	10.3 ± 0.7	0.69 ± 0.06
Mcp1 G25A	0.13 ± 0.02	16.0 ± 5.0	12.5 ± 5.2
Mcp1 ΔG25	0.15 ± 0.02	12.4 ± 2.4	8.5 ± 2.1

contacts between the N-terminal bottom of the short helix and cathepsin V. When we modeled the complex between macrocypin and cathepsin V by superimposing macrocypin on the clitocypin structure in the complex, it became evident that the binding loops do not fit into the active site. To find out whether the binding loops are the same in clitocypin and macrocypin, we expressed four mutants in which Gly-24 in the S3 binding area of clitocypin and Gly-25 in macrocypin were either replaced by alanine or deleted. We have also used these mutants to assess the relevance of the Gly-24-Gly-25 peptide flip. We assumed that the mutation to alanine or its deletion will reduce the flexibility of the main chain. The resulting clitocypin mutants yielded K_i values to cathepsin V that were 20 times higher than that of the native variant. It is notable that the major source of this difference is the slower association, whereas dissociation was not significantly affected (Table 3). This suggests a mechanism in which the peptide bond flip occurs before or concurrently with the inhibitor docking. The macrocypin mutants exhibit equivalent effects on their K_i constants, indicating that the loops that bind into the active site of cysteine cathepsin are equivalent in clitocypin and macrocypin. This implies that the binding loops of macrocypin exhibit substantial conformational flexibility during binding into the active site of their target enzymes.

Inhibition of AEP—When mammalian asparaginyl endopeptidase was characterized, it was named according to its distinctive specificity (39). It suggests that AEP must have an S1 substrate binding site that is highly specific for asparagine. AEP is inhibited in the low nanomolar range (3–20 nM) by natural and recombinant clitocypin, natural macrocypin, and some isoforms of expressed macrocypin (macrocypins 1 and 3), whereas macrocypin 4 does not inhibit AEP at all. The availability of the mycocypins three-dimensional structure enabled the search for potential interacting areas to be narrowed down. Inspection of the aligned sequences of these isoforms in their surface regions focused attention on the β5-β6 loop, positioned in the lower crown region (residues 71–76 containing the sequence Ile-Asp-Asn-Ser-Ile). This part of the sequence is similar to the consensus sequence (S/T)N(D/S)(M/I) found in three inhibitory cystatins C, E, and F (Table 4) that bind to AEP in the nanomolar range (40). Interestingly, in macrocypin 4, the residue at position 72 is Lys, in contrast to the equipositioned Asn in macrocypins 1 and 3. To verify the role of the residues in these regions, like Alvarez-Fernandez *et al.* (40) in the case of cystatin C, we introduced mutations in the inhibitory sequence. The residues that differ between macrocypins 1 and 4 in the β5-β6 loop were

TABLE 4

Alignment of loops of macrocypin isoforms, clitocypin, cystatins V, E, and F, STI, and BbCI involved in AEP and trypsin inhibition

Residues that bind to the P1 pocket are marked with an asterisk. Mutated residues are shown bold. Cst, cystatin.

	*
Clt	56-YTIKYQGL**N**APFEYG-75
Mcp1	61-TITEFR**I**D**N**SIPGQW-80
Mcp3	59-TITEIRDDNCIPGQW-78
Mcp4	59-TITEFR**A**D**K**SIPGQW-78
CstC	26-VGEYNKASNDMYHSR-45
CstE	23-VASYNMGSNSIYYFR-42
CstF	33-VEKFNNCTNDMFLFK-52
BbCI	50-TVRFETPLAIAIITE-69
STI	50-GTIISSPYRIRFIAE-69

TABLE 5

Inhibition constants of AEP, cathepsin V and trypsin by macrocypin, clitocypin and their mutants

NI, no inhibition.

	K_i AEP	K_i Cathepsin V	K_i Trypsin
	nM	nM	μM
Mcp 1[a]	3.38 ± 1.44	0.69 ± 0.06	NI
Mcp 1/4	NI	3.43 ± 0.31	0.18 ± 0.02
Mcp 4[a]	>1000	1.44 ± 0.11	0.16 ± 0.01
Mcp 4/1	2.86 ± 0.38	10.2 ± 0.6	NI
Mcp 4 K72R	NI	6.9 ± 1.1	0.13 ± 0.02
Clt[a]	21.5 ± 2.81	0.084 ± 0.03	NI
Clt N70K	NI	0.26 ± 0.09	NI

[a] Data reported previously (9).

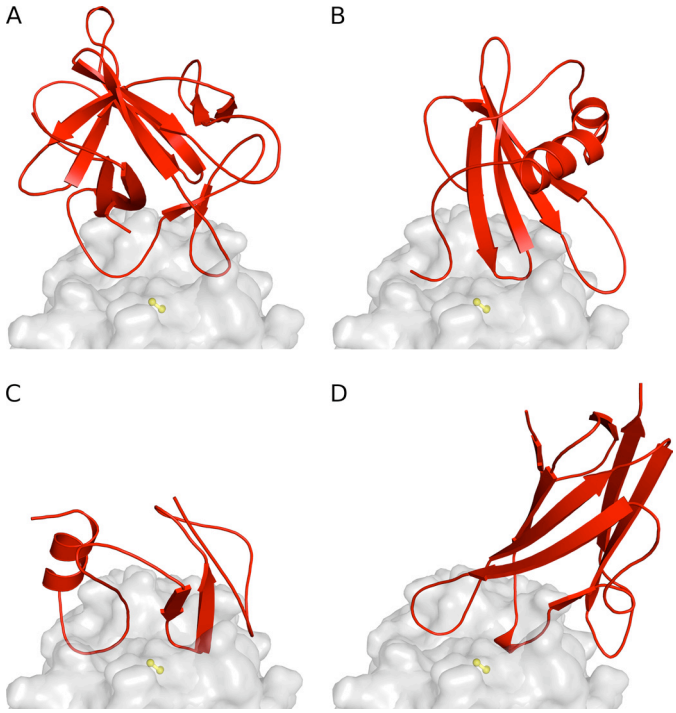


FIGURE 5. **Comparison of the binding modes of inhibitors.** The fold of inhibitors is shown in *red* (*A*, clitocypin; *B*, stefin A; *C*, p41 fragment; *D*, chagasin), whereas the endopeptidase cathepsin V is presented as a *gray surface*. The surface of the catalytic cysteine is shown in *yellow*. All inhibitors have different folds, yet for docking to the active site of papain-like cysteine proteases they utilize loops that fill the active site along its whole length and occlude the reactive site cysteine residue by two loops.

exchanged. In addition, the corresponding clitocypin mutant (Clt-N69K) was also expressed. As expected, the mutant Mcp-1/4 (Mcp1 with β5-β6 loop of Mcp4) exhibited no inhibition, whereas the mutant Mcp-4/1 (Mcp4 with β5-β6 loop of Mcp1) increased inhibition against AEP (Table 5). Equivalently, the Clt-N69K mutant did not inhibit AEP. Thus, Asn-72 in macrocypins and Asn-69 in clitocypins are confirmed to be the residues responsible for the inhibition of AEP. Mycocypins are, in this respect, similar to cystatin C, which has two different binding sites, one for papain-like proteases and another for AEP (40).

Trypsin Inhibition—The binding geometry of several families of protein inhibitors of serine proteases, including the soybean Kunitz-type inhibitor (35), are known to adopt a substrate-like conformation known as the "canonical" binding mode (41). All Kunitz-type serine protease inhibitors inhibit trypsin with a highly homologous loop from the root region that mimics the substrate and is positioned between strands β4 and β5. This loop contains either lysine or arginine, which binds into the S1 pocket of trypsin. From the sequence and structure alignments it is evident that the classical β4-β5 loop is missing in macrocypin and clitocypin, and these proteins do not inhibit trypsin. Surprisingly, macrocypin 4 was found to inhibit trypsin with a K_i value in the micromolar range. The K_i values of the exchange mutants produced for AEP binding site identification (Table 5) show that the Lys-74 residue of macrocypin 4 is mandatory for inhibition of trypsin. The Mcp4 mutant with Lys-74 replaced by arginine (Mcp4-K74R) was similarly inhibitory (Table 5), thereby confirming the involvement of the loop β5-β6 positioned within the lower crown layer in binding to the trypsin active site. These mutants have no significant effect on the inhibition of cathepsin V (Table 5). The binding loop of macrocypins and clitocypins is, thus, positioned differently from the serine protease binding loop of known Kunitz-type inhibitors such as STI.

DISCUSSION

Like cystatins (10, 11), the p41 fragment (14) and chagasin (15), clitocypins, and macrocypins bind to cysteine proteases along the whole active site cleft (Fig. 5). These molecules have different folds, yet for docking to papain-like cysteine proteases, they utilize a similar architecture by which the activity of the target proteases is inhibited. Their constructs occlude the reactive site cysteine. On the non-primed substrate binding site they utilize a single chain. The first binding region in clitocypin is the loop Asp-19—Glu-25. Its position is similar to those of the N-terminal region in stefin A and the first loops of the p41 fragment and chagasin. This region contains a residue that, in a substrate-like manner, fills the S2 binding pocket. In contrast, the loops covering the primed binding areas are much less similar. The second binding region in clitocypins is a single, broad loop (Glu-39—Ile-50), whereas cystatins, the p41 fragment, and chagasin use two loop constructs. The two broad loops of mycocypins are stabilized by multiple hydrogen bonds and are much more rigid than the N-terminal trunk and two loops in

cystatins. This explains why cystatins are capable of competing for binding with the additional features of exopeptidase such as occluding loop and mini chain, whereas mycocypins exhibit lower affinity or no binding at all.

The mode of the binding of mycocypins to cysteine cathepsins differs markedly from the binding of the Kunitz type of β-trefoil folded inhibitors to serine proteases. The two binding loops from the crown region bind into the non-primed and primed substrate binding regions of cysteine proteases and occlude the catalytic cysteine residue. In contrast, only one binding loop from the root region of Kunitz inhibitors docks into the active site of a serine protease, in a substrate-like manner. A possible explanation for the differences in the modes of canonical inhibition of cysteine cathepsins and trypsin-like serine proteases may lie in the features of the active site cleft. Whereas in the trypsin-like proteases the S1 binding site is a pocket in the protein structure, in cysteine cathepsins the S1 binding site is positioned on the surface on one side of the active site cleft, shaped so that the P1 residue side chain points away from the protein core (42). Furthermore, analysis of the structural data has revealed that papain-like cysteine proteases have only three clearly defined substrate binding sites (S2, S1, S1′) and one conditional site (S2′), whereas the binding into regions beyond position 2 can only be considered substrate binding areas spread over the surface of the widening active site cleft (43). Cysteine cathepsins, thus, appear to lack the binding surface to which the P1 and neighboring residues could be tightly anchored in a substrate-like manner and, therefore, can probably not be inhibited by the single loop construct.

The flexible peptide bond, which can flip on docking to protease, is a unique feature among the cysteine protease inhibitors. Peptide flipping has already been observed in the mitogen-activated protein kinase p38α MAPK, where the flip of the Met-109 and Gly-110 peptide bond facilitates the higher specificity of certain inhibitors (44).

The trefoil fold supports 11 loops coming out of the six-stranded β-barrel. Nine are in the crown region (six are positioned at the lower level of the crown, and three enclose the top of the crown), and two are in the roots. Therefore, it is easier to comprehend that the six loops from the lower crown region can act in pairs, whereas the two loops from the root region lack that capability and must bind alone. In this respect the report of the binding site of *B. bauhinioides* cruzipain inhibitor (BbCI) to cysteine cathepsins is intriguing, as the authors (17) suggest that the same alanine residue positioned within the root region, which is responsible for binding to neutrophil elastase, is also crucial for the binding to cathepsin L and cruzipain (the partial cleavage of the serine protease interacting loop after incubation of BbCI with cruzipain was the key evidence supporting the hypothesis of the common interaction site). The binding of BbCI to trypsin is consistent with current structural knowledge, as the loop in which Ala-63 resides folds very similarly to the loop of STI, including the position and orientation of the Ala-63 residue. However, the single inhibitory loop is not consistent with the canonical inhibition mechanism of cysteine cathepsins evidenced here. Superimposition of the structure of BbCI on clitocypin in complex with cathepsin V showed that two broad loops in the BbCI structure are equivalent to the clitocypin binding loops and that the BbCI sequence contains two consecutive glycine residues, Gly-28—Gly-29, homologous to the peptide bond flip residues Gly-24 and Gly-25. Hence, these two loops are probably responsible for cathepsin L inhibition and not the loop containing the trypsin cleavage site (it should be noted, however, that as in the case of macrocypin, the tips of the loops require a slight adjustment to fit into the active site of a cysteine cathepsin). The absence of inhibition of cathepsin V by BbCI, given the similarity of cathepsins L and V, cannot, however, be explained.

Clitocypins and macrocypins exhibit no sequence similarity to other known proteins, which was the basis for establishing the I48 and I85 families supported by the sequence alignment score with an E value less than 0.001 (FASTA search with default BLOSUM50 matrix used in MEROPS). However, their structure has revealed that the basic element of their fold is the six-stranded β-barrel, the hallmark of the β-trefoil fold shared by the members of the I3 MEROPS family that includes serine protease inhibitors of the Kunitz type. The sequence similarity, based on superimposition of the structures, is low even within the 6-stranded β-barrel part (Fig. 4), thus, questioning the common origin of these two groups of proteins. This confirms that mycocypins (families I48 and I85) are indeed distinct from members of I3 family, whereas the structural similarity between these families provides support that they belong to the same clan IC.

The β-trefoil fold is armed with potent interacting loops that differ in shape and composition and are able to inhibit several classes of proteases including cathepsins, AEP, cruzipain, trypsin, chymotrypsin, elastase, subtilisin, and amylases. Several loops are involved in inhibition, whereas the same inhibitory loop can target different proteases. For example, the crown region loops β1-β2 and β3-β4 are used for inhibiting the papain-like cysteine proteases by mycocypins and, most probably, cruzipain and cathepsin L by BbCI (17). The root region loop β4-β5 is involved in inhibiting chymotrypsin (by the winged bean chymotrypsin inhibitor) (45), trypsin (by STI) (35), and porcine pancreatic elastase and human neutrophil elastase (by BbCI) (17), whereas the crown region loop β5-β6 is involved in inhibiting AEP and trypsin (by mycocypins) and the subtilisin savinase (by barley α-amylase/subtilisin inhibitor) (46). The numerous crown region loops, β1-β2, β3-β4, β6-β7, β9-β10 and β11-β12, are responsible for the interaction of barley α-amylase/subtilisin inhibitor with barley α-amylase (47). This makes β-trefoil inhibitors, in particular mycocypins, promising candidates for transgenic trials for the purposes of crop protection, where inhibitors with selectivity against only one class of proteases have failed because of the compensation of proteolytic activity by induced expression of other proteases insensitive to the transgenic inhibitor (48–50).

Acknowledgments—We are very grateful to Dr. Roger Pain for critical reading of the manuscript and Dr. Tatjana Popovič for help with kinetics. Vida Puizdar is acknowledged for providing asparaginyl endopeptidase and Dr. Katja Galeša for cloning and test expression of double methionine clitocypin mutant.

REFERENCES

1. Ryan, C. A. (1990) *Annu. Rev. Phytopathol.* **28,** 425–449
2. Frost, C. J., Mescher, M. C., Carlson, J. E., and De Moraes, C. M. (2008) *Plant Physiol.* **146,** 818–824
3. Green, T. R., and Ryan, C. A. (1972) *Science* **175,** 776–777
4. Ussuf, K. K., Laxmi, N. H., and Mitra, R. (2001) *Curr. Sci.* **80,** 847–854
5. Christou, P., Capell, T., Kohli, A., Gatehouse, J. A., and Gatehouse, A. M. (2006) *Trends Plant Sci.* **11,** 302–308
6. Benchabane, M., Goulet, M. C., Dallaire, C., Côté, P. L., and Michaud, D. (2008) *Plant Physiol. Biochem.* **46,** 701–708
7. Rawlings, N. D., Morton, F. R., Kok, C. Y., Kong, J., and Barrett, A. J. (2008) *Nucleic Acids Res.* **36,** D320–D325
8. Sabotic, J., Popovic, T., Puizdar, V., and Brzin, J. (2009) *FEBS J.* **276,** 4334–4345
9. Brzin, J., Rogelj, B., Popovic, T., Strukelj, B., and Ritonja, A. (2000) *J. Biol. Chem.* **275,** 20104–20109
10. Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B., and Turk, V. (1990) *EMBO J.* **9,** 1939–1947
11. Jenko, S., Dolenc, I., Guncar, G., Dobersek, A., Podobnik, M., and Turk, D. (2003) *J. Mol. Biol.* **326,** 875–885
12. Lenarcic, B., and Turk, V. (1999) *J. Biol. Chem.* **274,** 563–566
13. Mihelic, M., and Turk, D. (2007) *Biol. Chem.* **388,** 1123–1130
14. Guncar, G., Pungercic, G., Klemencic, I., Turk, V., and Turk, D. (1999) *EMBO J.* **18,** 793–803
15. Ljunggren, A., Redzynia, I., Alvarez-Fernandez, M., Abrahamson, M., Mort, J. S., Krupa, J. C., Jaskolski, M., and Bujacz, G. (2007) *J. Mol. Biol.* **371,** 137–153
16. Hansen, D., Macedo-Ribeiro, S., Veríssimo, P., Yoo Im, S., Sampaio, M. U., and Oliva, M. L. (2007) *Biochem. Biophys. Res. Commun.* **360,** 735–740
17. Araújo, A. P., Hansen, D., Vieira, D. F., Oliveira, C., Santana, L. A., Beltramini, L. M., Sampaio, C. A., Sampaio, M. U., and Oliva, M. L. (2005) *Biol. Chem.* **386,** 561–568
18. Sabotic, J., Galesa, K., Popovic, T., Leonardi, A., and Brzin, J. (2007) *Protein Expr. Purif.* **53,** 104–111
19. Weiner, M. P., and Costa, G. L. (1994) *PCR Methods Appl.* **4,** S131–S136
20. Studier, F. W. (2005) *Protein Expr. Purif.* **41,** 207–234
21. Mihelic, M., Teuscher, C., Turk, V., and Turk, D. (2006) *FEBS Lett.* **580,** 4195–4199
22. Galesa, K., Brzin, J., Sabotic, J., and Turk, D. (2006) *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **62,** 10–12
23. Sabotic, J., Gaser, D., Rogelj, B., Gruden, K., Strukelj, B., and Brzin, J. (2006) *Biol. Chem.* **387,** 1559–1566
24. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
25. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) *Acta Crystallogr. D Biol. Crystallogr.* **58,** 1948–1954
26. Navaza, J. (1994) *Acta Cryst.* **50,** 157–163
27. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat. Struct. Biol.* **6,** 458–463
28. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255
29. Turk, D. (1992) *Weiterentwicklung eines Programms fuer Molekülgraphik und Elektrondichte-Manipulation und Seine Anwendung auf Verschiedene Protein-Strukturaufklërungen*. Ph.D. thesis, Technische Universitaet Muenchen, Germany
30. Somoza, J. R., Zhan, H., Bowman, K. K., Yu, L., Mortara, K. D., Palmer, J. T., Clark, J. M., and McGrath, M. E. (2000) *Biochemistry* **39,** 12543–12551
31. Andrejasic, M., Praaenikar, J., and Turk, D. (2008) *Acta Crystallogr. D. Biol. Crystallogr.* **64,** 1093–1109
32. Morrison, J. F. (1982) *Trends Biochem. Sci.* **7,** 102–105
33. Henderson, P. J. F. (1972) *Biochem. J.* **127,** 321–333
34. Krissinel, E., and Henrick, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2256–2268
35. Song, H. K., and Suh, S. W. (1998) *J. Mol. Biol.* **275,** 347–363
36. Onesti, S., Brick, P., and Blow, D. M. (1991) *J. Mol. Biol.* **217,** 153–176
37. Graves, B. J., Hatada, M. H., Hendrickson, W. A., Miller, J. K., Madison, V. S., and Satow, Y. (1990) *Biochemistry* **29,** 2679–2684
38. Ago, H., Kitagawa, Y., Fujishima, A., Matsuura, Y., and Katsube, Y. (1991) *J. Biochem.* **110,** 360–363
39. Chen, J. M., Dando, P. M., Rawlings, N. D., Brown, M. A., Young, N. E., Stevens, R. A., Hewitt, E., Watts, C., and Barrett, A. J. (1997) *J. Biol. Chem.* **272,** 8090–8098
40. Alvarez-Fernandez, M., Barrett, A. J., Gerhartz, B., Dando, P. M., Ni, J., and Abrahamson, M. (1999) *J. Biol. Chem.* **274,** 19195–19203
41. Bode, W., and Huber, R. (1992) *Eur. J. Biochem.* **204,** 433–451
42. Turk, B., Turk, V., and Turk, D. (1997) *Biol. Chem.* **378,** 141–150
43. Turk, D., Guncar, G., Podobnik, M., and Turk, B. (1998) *Biol. Chem.* **379,** 137–147
44. Fitzgerald, C. E., Patel, S. B., Becker, J. W., Cameron, P. M., Zaller, D., Pikounis, V. B., O'Keefe, S. J., and Scapin, G. (2003) *Nat. Struct. Biol.* **10,** 764–769
45. Ravichandran, S., Sen, U., Chakrabarti, C., and Dattagupta, J. K. (1999) *Acta Crystallogr. D Biol. Crystallogr.* **55,** 1814–1821
46. Micheelsen, P. O., Vévodová, J., De Maria, L., Ostergaard, P. R., Friis, E. P., Wilson, K., and Skjøt, M. (2008) *J. Mol. Biol.* **380,** 681–690
47. Vallée, F., Kadziola, A., Bourne, Y., Juy, M., Rodenburg, K. W., Svensson, B., and Haser, R. (1998) *Structure* **6,** 649–659
48. Jongsma, M. A., Bakker, P. L., Peters, J., Bosch, D., and Stiekema, W. J. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92,** 8041–8045
49. De Leo, F., Bonade-Bottino, M. A., Ceci, L. R., Gallerani, R., and Jouanin, L. (1998) *Plant Physiol.* **118,** 997–1004
50. Cloutier, C., Jean, C., Fournier, M., Yelle, S., and Michaud, D. (2000) *Arch. Insect Biochem. Physiol.* **44,** 69–81

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 285, NO. 13, pp. 10078–10086, March 26, 2010

Stefin B Interacts with Histones and Cathepsin L in the Nucleus*

Received for publication, June 17, 2009, and in revised form, January 8, 2010 Published, JBC Papers in Press, January 14, 2010, DOI 10.1074/jbc.M109.034793

Slavko Čeru[‡], Špela Konjar[‡], Katarina Maher[‡], Urška Repnik[‡], Igor Križaj[§], Mojca Benčina[¶], Miha Renko[‡], Alain Nepveu[‖], Eva Žerovnik[‡], Boris Turk[‡], and Nataša Kopitar-Jerala[‡1]

From the Departments of [‡]Biochemistry and Molecular and Structural Biology and [§]Molecular and Biomedical Sciences, Jožef Stefan Institute, Ljubljana SI-1000, Slovenia, the [¶]Department of Biotechnology, National Institute of Chemistry, Ljubljana SI-1000, Slovenia, and the [‖]Molecular Oncology Group, McGill University, Montreal, Quebec H3A 1A1, Canada

Stefin B (cystatin B) is an endogenous inhibitor of cysteine proteinases localized in the nucleus and the cytosol. Loss-of-function mutations in the stefin B gene (*CSTB*) gene were reported in patients with Unverricht-Lundborg disease (EPM1). We have identified an interaction between stefin B and nucleosomes, specifically with histones H2A.Z, H2B, and H3. In synchronized T98G cells, stefin B co-immunoprecipitated with histone H3, predominantly in the G_1 phase of the cell cycle. Stefin B-deficient mouse embryonic fibroblasts entered S phase earlier than wild type mouse embryonic fibroblasts. In contrast, increased expression of stefin B in the nucleus delayed cell cycle progression in T98G cells. The delay in cell cycle progression was associated with the inhibition of cathepsin L in the nucleus, as judged from the decreased cleavage of the CUX1 transcription factor. *In vitro*, inhibition of cathepsin L by stefin B was potentiated in the presence of histones, whereas histones alone did not affect the cathepsin L activity. Interaction of stefin B with the Met-75 truncated form of cathepsin L in the nucleus was confirmed by fluorescence resonance energy transfer experiments in the living cells. Stefin B could thus play an important role in regulating the proteolytic activity of cathepsin L in the nucleus, protecting substrates such as transcription factors from its proteolytic processing.

Cysteine cathepsins are involved in protein degradation (1) and the development and function of the immune system (2). Cathepsin L is an endopeptidase that is able to perform limited proteolysis in the endosomes and lysosomes of specific cell types. Besides its role in hair formation and skin metabolism, it is involved in T-cell selection and NKT cell development (3). It participates in processing the major histocompatibility complex II invariant chain in thymic cortex epithelial cells (4), encephalin in chromaffin granules of neuroendocrine cells (5), and in the degradation and recycling of growth factors and their receptors in epidermal keratinocytes (6). Cathepsin L is also associated with an endosomal processing step during invasion of cells by Ebola virus (7), severe acute respiratory syndrome (SARS) coronavirus (8), and murine hepatitis coronavirus (9). As the result of gene duplication, the human genome encodes for two cathepsin L-like proteases, namely the human cathepsin L and cathepsin V (cathepsin L2), whereas in mouse only cathepsin L is present (10). At the protein level, mouse cathepsin L displays a higher sequence homology to human cathepsin V than to human cathepsin L (11). Cathepsin V shares 80% protein sequence identity with cathepsin L, but in contrast to the ubiquitously expressed cathepsin L, its expression is restricted to thymus and testis (11, 12).

Recently, the otherwise endosomal proteinase cathepsin L has been reported to be active in the nucleus. It cleaves the CUX1 transcription factor and as a result accelerates progression into the S phase of the cell cycle (13). Cathepsin L deficiency was shown to cause a global rearrangement of chromatin structure and redistribution of specifically modified histones (14). In addition, cathepsin L was found to cleave histone H3.2 in the nucleus during mouse embryonic stem cell differentiation (15).

Cathepsin L is inhibited *in vitro* by a number of proteins as follows: cystatins (16), thyropins (17), and some of the serpins (18, 19). Type 1 cystatins, or stefins, are mainly intracellular, whereas type 2 cystatins are predominantly secreted (20, 21). Stefin B is localized in the cytosol and nucleus of proliferating cells (22). Loss-of-function mutations in the cystatin B (*CTSB*, *stefin B*) gene are found in patients with Unverricht-Lundborg disease (EPM1), but its physiological implication in the pathogenesis of the disease has yet to be defined (23–26). EPM1 is an autosomal recessive inherited disease in which patients suffer from myoclonic jerks, tonic-clonic epileptic seizures, and progressive decline in cognition (26). Histopathological examination of the brain has shown neural degeneration in several areas of the central nervous system, with cerebellar damage and serious alterations of Purkinje cells (27). The most common mutation in EPM1 patients is a dodecamer repeat expansion in the stefin B (*CSTB*) gene promoter region that leads to reduced mRNA and protein levels (23, 25). In addition, four mutations in the coding region were reported in EPM1 (23, 28).

Not only cystatins but also some other proteinase inhibitors are found in the nucleus. A serine proteinase inhibitor (serpin), the myeloid and erythroid nuclear termination stage-specific protein, MENT,[2] was the first described cysteine proteinase inhibitor that interacts with chromatin and influences heterochromatin distribution (29, 30).

* This work was supported by Slovenian Research Agency Grants J3-9324 and J3-0612 (to N. K. J.) and Grant P-0140 (to V. T. and B. T.).

[1] To whom correspondence should be addressed. Tel.: 386-1-477-3510; Fax: 386-1-477-3984; E-mail: natasa.kopitar@ijs.si.

[2] The abbreviations used are: MENT, myeloid and erythroid nuclear termination stage-specific protein; E-64d, L-*trans*-epoxysuccinyl(OEt)-Leu-3-methylbutylamide; FRET, fluorescence resonance energy transfer; NFRET, normalized FRET; GFP, green fluorescent protein; MEF, mouse embryonic fibroblast; NLS, nuclear localization signal; YFP, yellow fluorescent protein; DMEM, Dulbecco's modified Eagle's medium.

The aim of our study was to identify proteins interacting with stefin B in the cell nucleus. The fundamental repeating unit of eukaryotic chromatin is the nucleosome, which is composed of an octamer of the four core histones H3, H4, H2A, H2B, around which 147 bp of DNA are wrapped (31). We have shown that stefin B interacts with the histones H2A.Z, H2B, and H3 and with cathepsin L in living cells. In contrast to MENT, we found that nuclear cystatin/stefin B interacted with cathepsin L and with histones in the nucleus, but it did not bind to DNA. Our results suggest that stefin B regulates the activity of cathepsin L in the nucleus and protects the CUX1 transcription factor and probably other substrates from proteolytic cleavage by cathepsin L.

EXPERIMENTAL PROCEDURES

Antibodies—Antibodies to stefin B were described previously (32). Rabbit polyclonal antibody to histone H3 (ab1791), rabbit polyclonal to trimethyl K4 (H3K4me3) (ab8580), and anti-GFP antibody (ab290) were from Abcam. Antibodies to CUX1 (861 and 1300) have been described (13).

Mice—Stefin B (cystatin B)-deficient mice were created as described previously (33). Stefin B-deficient mice were provided by Dr. R. M. Myers, Stanford University, and bred in our local colony. All mice were genotyped by PCR as described previously (33, 34).

Mouse Embryonic Fibroblasts (MEF), Preparation and Cell Cycle Synchronization—MEFs were prepared from individual embryos at embryonic day 14.5. The head and internal organs were removed, and the torso was minced and dispersed in 0.1% trypsin (45–60 min at 37 °C). Cells were grown for two population doublings (considered as one passage) and then viably frozen. MEFs were maintained in DMEM containing 10% fetal bovine serum (Sigma) and subcultured 1:4 on reaching confluence. For serum starvation experiments, MEFs were plated in DMEM containing 0.4% fetal bovine serum and incubated at 37 °C for 72 h before stimulation with DMEM containing 10% fetal bovine serum.

Plasmids and Constructs—The cDNA clone for *CTSB* was obtained from Image (IMAGE, 3453675). It was PCR-amplified and cloned into pcDNA3 vector (Invitrogen) at HindIII and XhoI restriction sites and into pEF/Myc/Nuc vector (Invitrogen), which contains nuclear localization signal and targets the expressed protein to the nucleus at XhoI and BamHI restriction sites. DNA sequence was determined using an ABI PRISM 310 Genetic Analyzer (PerkinElmer Life Sciences). The multiple cloning site of pcDNA3 vector was changed prior to the insertion of T-Sapphire and Venus. This was done in two steps. The first linker, A, was constructed from two oligonucleotides as follows: A, 5′-AAT TCT GCA GGT ATT CTT CAC ACT GGA GGC CGA CCG GGC C-3′, and B, 5′-CGG TCG GCC TCC AGT GTG AAG AAT ACC TGC AG-3′ complementary to A. This was ligated through EcoRI and ApaI restriction sites into pcDNA3 vector. All restriction sites between EcoRI and ApaI in the multiple cloning site of pcDNA3 vector were removed with linker A, among them the restriction site for XhoI and XbaI. Vector pcDNA3 with inserted linker A is labeled pcDNA3L/A. The second linker, B, was constructed from two oligonucleotides as follows: C, 5′-AGC TTC GTC CGC TCG AGA GCG CTT CTA GAG GTC TGG GAG GTT CAG GTG GAG GTG GAG CTG CTG CCG-3′ (XhoI and XbaI sites underlined), and D, 5′-GAT CCG GCA GCT CCA CCT CCA CCT GAA CCT CCC AGA CCT CTA GAA GCG CTC TCG AGC GGA CGA-3′, complementary to C. It was ligated through HindIII and BamHI restriction sites of pcDNA3L/A vector. pcDNA3L/A vector with inserted linker B is labeled pcDNA3L/AB. The cDNAs from Venus YFP (35) and T-Sapphire GFP (36) were amplified by PCR. The resulting products, after BamHI/EcoRI digestion, were cloned into pcDNA3L/AB expression vector. pcDNA3L/AB vector with inserted Venus is labeled Ven-pcDNA3L/AB and with inserted T-Sapphire as T-Sap-pcDNA3L/AB. The cDNA clone for *CTSB* was PCR-amplified, and the resulting product was cloned into T-Sapphire-pcDNA3L/AB after XhoI/XbaI digestion.T-Sap-pcDNA3L/AB construct with inserted stefin B is labeled as Stefin B-GFP. Met-75 cathepsin L was PCR-amplified from procathepsin L cDNA (37), using forward (5′-GCC CGC CTC GAG ATG GCC ATG AAC GCC TTT GG-3′; XhoI site underlined) and reverse (5′-GTC CGC TCT AGA CAC AGT GGC GTA GCT GGC-3′; XbaI site underlined) oligonucleotides. The resulting product, after XhoI/XbaI digestion, was cloned into Venus-pcDNA3L/AB. The Venus-pcDNA3L/AB construct with inserted Met-75 cathepsin L is named M75 cath L-YFP.

Cell Culture—T98G human glioblastoma cell line, ATCC CRL-1690, was from the American Type Culture Collection (Manassas, VA). Cells were cultured as described previously (38). T98G cells were transfected with pEF/Myc/Nuc/stefin B named NB or empty pEF/Myc/Nuc vector alone named NO, using Lipofectamine 2000 (Invitrogen), according to the manufacturer's instructions. Positive clones overexpressing stefin B in the nucleus were obtained after selection with Geneticin (G418) (Invitrogen) (500 μg/ml) and confirmed with Western blots and stefin B-specific enzyme-linked immunosorbent assay (32). CHO-K1 cells (ATCC CCL-61) were cultured in DMEM supplemented with 10% fetal calf serum, 5 units/0.5 ml penicillin, and 5 μg/0.5 ml streptomycin at 37 °C in 5% CO_2. For FRET analysis, cells were seeded at a density of 1×10^5 on glass coverslips and transiently transfected with 1 μg of Stefin B-GFP and 1 μg of M75 cath L-YFP, using Lipofectamine 2000 (Invitrogen), according to the manufacturer's recommendations. The expression of the GFP fusion proteins was determined by Western blot. MCF-7 cells (ATCC HTB-22) were cultured in DMEM supplemented with 10% fetal calf serum, 5 units/0.5 ml of penicillin, and 5 μg/0.5 ml streptomycin at 37 °C in 5% CO_2. Cells were grown on 10-cm Petri dishes, transiently transfected with pEF/Myc/Nuc/stefin B (NB) or empty pEF/Myc/Nuc vector alone (NO), using Lipofectamine 2000 (Invitrogen), according to the manufacturer's instructions. Cells were lysed 24 h post-transfection and nuclear cell lysates prepared as described previously (13).

Preparation of Cell Lysates—Cell lysates were prepared as described previously (34). Nuclear extracts were prepared by the method of Dignam *et al.* (39), with minor modifications, including the use of a protease inhibitor mixture (catalog no. P8340; Sigma) and the addition of phenylmethylsulfonyl fluoride (Fluka, Basel, Switzerland) (0.5 mM) to the resuspension and lysis buffers. Nuclear cell lysates from MCF-7 cells were

prepared as described by Goulet *et al.* (13). In both cases, the supernatants were transferred to fresh test tubes and, if not used immediately, stored at −80 °C. Total protein concentration was determined using the Bradford assay (Bio-Rad).

Co-immunoprecipitation was performed as described previously (34). 5 μg of anti-stefin B polyclonal or monoclonal antibodies (32) were added to nuclear lysates and allowed to rock at 4 °C for 2 h or overnight. 50 μl of protein A-Sepharose beads were then added to the lysates, which were allowed to rock for another hour at 4 °C. Immunoprecipitates were washed three times in cold phosphate-buffered saline. Samples were resolved on SDS-PAGE and analyzed by Western blots.

Western blots were performed as described previously (13, 34). Equal amounts of protein were loaded and resolved in 15, 12.5, or 6% SDS-polyacrylamide gels and electrotransferred to nitrocellulose membranes. Proteins were visualized with ECL (Amersham Biosciences) according to the manufacturer's instructions.

N-terminal Sequencing—Edman sequence analyses of protein samples were performed on a model 492A Procise Protein Sequencing System (Applied Biosystems, Foster City, CA). Proteins were electrotransferred from SDS-polyacrylamide gels to polyvinylidene difluoride membranes and sequenced using a pulsed-liquid blot sequencing protocol. Phenylthiohydantoin-derivatives were analyzed on line on a microbore high pressure liquid chromatography system 140C (Applied Biosystems) using an RP C18 Spheri-5 column (Brownlee). Any cysteine residues were alkylated before sequencing. All reagents and solvents were of sequencing grade (Applied Biosystems).

Cell Cycle Analysis—Cells were trypsinized and pellets fixed in 70% ethanol overnight at −20 °C. After washing, the cells were incubated for 30 min at 37 °C in phosphate-buffered saline containing 40 μg/ml propidium iodide, 100 μg/ml RNase, and 0.05% (w/v) Triton X-100. Finally, cells were washed and resuspended in 0.5 ml of phosphate-buffered saline. Samples were analyzed with a FACSCalibur flow cytometer (BD Biosciences), and cell cycle profiles were evaluated using CellQuest software (BD Biosciences), version 3.3.

FRET Microscopy—FRET measurements were made using a Leica TCS SP5 microscope equipped with a 405-nm laser. The images were acquired with 405- and 515-nm laser lines. Images were taken through a 63 × 1.4-numerical aperture oil immersion objective. To explore the interaction of the Met-75 truncated form of cathepsin L with stefin B, we used a fluorescent protein-protein pair that has excitation and emission properties favorable for FRET; the emission wavelength of T-Sapphire-GFP partially overlaps with the excitation wavelength of YFP (40). For FRET experiments, transient transfections were performed as described above, and expression levels of stefin B-GFP (donor) and Met-75 cathepsin L-YFP (acceptor) proteins were adjusted to similar levels by Western blot. FRET between stefin B-GFP and M75 cath L-YFP was measured in live cells plated on glass-bottomed culture dishes, 18–24 h after transfection. Fluorescence was recorded at three different settings: GFP_{ex}, 405 nm/GFP_{em} 515 nm; YFP_{ex}, 515 nm/YFP_{em}, 523–535 nm; $FRET_{ex}$, 405 nm/$FRET_{em}$, 528–535 nm. Laser power and detector gain were adjusted in the different channels. The image of FRET was generated with the "PixFRET" plug-in for the ImageJ software. In all FRET experiments, negative FRET controls were analyzed after transfection or co-transfection of GFP/YFP or co-transfection R68*X*-GFP/M75 cath L-YFP (negative controls). Bleed through coefficients were calculated using FRET/donor or FRET/acceptor image stacks captured from cells expressing only the donor or acceptor (41). Coefficients were averaged from at least 80 cells from 40 separate stacks for each experiment.

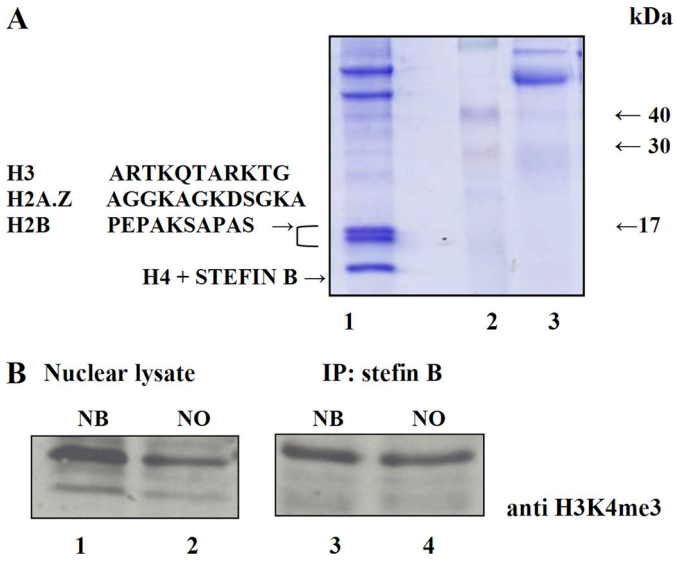


FIGURE 1. **Stefin B in the nucleus co-immunoprecipitates with histones.** *A,* nuclear lysates were prepared as described under "Experimental Procedures" and immunoprecipitated with anti-stefin B antibodies. After SDS-PAGE, gels were stained with Coomassie Blue. The three bands interacting with stefin B were identified as histones by N-terminal amino acid sequencing. N-terminal sequences of the 17- and 14-kDa bands are shown. *Track 1,* nuclear lysates immunoprecipitated with stefin B antibodies; *track 2,* molecular weight standard; *track 3,* control, nuclear lysates immunoprecipitated with antibodies against cathepsin B. *B,* stefin B co-immunoprecipitates with the H3K4me3 histone variant. Nuclear lysates from T98G cells transfected with stefin B in pEF/Myc/Nuc vector (NB) (*track 1*) and control T98G cells transfected with an empty vector pEF/Nuk/Myc alone (NO) (*track 2*) were separated by 15% SDS-PAGE, followed by Western blotting with anti-H3K4me3 antibody. Nuclear lysates from T98G cells transfected with stefin B in pEF/Myc/Nuc vector (NB) (*track 3*) and nuclear lysates from T98G cells transfected with control empty vector (NO) (*track 4*), both immunoprecipitated with stefin B antibodies, were separated by 15% SDS-PAGE, followed by Western blotting with anti-H3K4me3 antibody.

Cathepsin L Inhibition in Vitro—Cathepsin L was preactivated by incubation in 0.1 mM acetate, 1 mM EDTA, 0.1% (w/v) Brij-35, 0.02% (w/v) sodium azide, 10 mM cysteine, pH 5.5, for at least 10 min at room temperature before use. The active enzyme concentration was determined with *trans*-epoxysuccinyl-L-leucylamido-(4-guanidino)-butane (E-64) (Peptide Research Institute, Osaka, Japan) titration. Cathepsin L (final concentration 21.5 nM) was incubated for 5 min with histone isolated from calf thymus (H 4524, Sigma) at final concentrations from 10.8 nM to 2.15 μM. When determining the stefin B and histone interactions, stefin B (final concentration 15 nM) was preincubated with final concentrations of histones from 10.8 nM to 2.15 μM for 5 min, followed by the addition of cathepsin L (final concentration 21.5 nM). After the initial incubation steps, substrate benzyloxycarbonyl-Phe-Arg-*p*-nitroanilide (Bachem, AG, Switzerland) was added to 100 μM

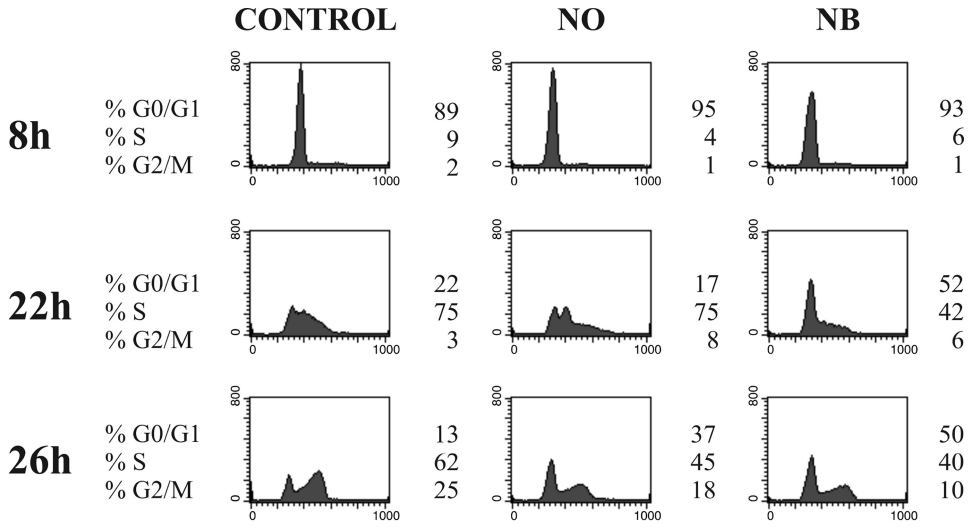


FIGURE 2. **Overexpression of stefin B into the nucleus delays cell cycle progression.** Control T98G cells, T98G cells transfected with an empty vector pEF/Nuk/Myc alone (NO), and T98G cells transfected with stefin B in pEF/Myc/Nuc vector (NB) were harvested at the indicated times, and cell cycle distribution was monitored by flow cytometry analysis after DNA staining with propidium iodide.

final concentration, and initial velocity was determined spectrophotometrically at 410 nM.

RESULTS

Stefin B Interacts with Histones H2A.Z, H2B, and H3 in the Nucleus—With the aim of defining the function of stefin B in the nucleus, we set out to identify proteins that interact with it. Nuclear cell lysates were immunoprecipitated with anti-stefin B antibodies and separated by SDS-PAGE. The 15- and 14-kDa bands were identified as histones H3, H2B, and H2A.Z by N-terminal protein sequencing (Fig. 1*A*). Histone H4 and stefin B were both found around 11 kDa. We found it particularly interesting that stefin B also interacts with the histone H2A.Z variant. The H2A.Z variant is found associated with gene regulatory elements in promoter regions (42). A recent study proposed that H2A.Z is partly co-localized to the same nucleosome as the H3K4me3 histone variant (43). An immunoprecipitation experiment confirmed that the H3K4me3-modified histone also co-immunoprecipitated with stefin B in T98G cells (Fig. 1*B*).

Increased Expression of Stefin B in the Nucleus Delays Cell Cycle Progression—Incubation of cells with the cysteine protease inhibitor E-64d was reported to delay the entry of cells into the S phase of the cell cycle (13, 44). T98G astrocytoma cells were used for cell cycle experiments, as these cells can be synchronized in G_0/G_1 in the absence of serum (38, 45). The effect of E-64d on cell cycle progression was confirmed in synchronized T98G cells (data not shown). Next, we examined whether overexpression of stefin B in the nucleus influences cell cycle progression. T98G cells were transfected with a pEF/Myc/Nuc vector expressing stefin B pEF/Myc/Nuc-stefin B (NB), or an empty vector alone (NO). Expression levels of stefin B in stably transfected T98G cells were quantified by enzyme-linked immunosorbent assay, as described previously (32). In whole cell lysates of cells stably expressing stefin B, the levels of stefin B were 478 ng/mg (ng of stefin B/mg of total cell protein), and in control cells transfected with an empty vector the concentration of stefin B was 178 ng/mg. A marked delay in the progression of cells from the G_1 to the S phase was observed in T98G cells overexpressing stefin B, as compared with control T98G cells and cells transfected with the empty vector alone (NO) (Fig. 2, *22* and *26 h*).

Stefin B Deficiency Leads to Accelerated Entry into S Phase—The influence of stefin B on cell cycle progression was confirmed in MEFs prepared from stefin B-deficient mice and wild type mice. As described previously, the synchronization of MEFs in G_0/G_1 was not as efficient as for T98G cells (46). Nevertheless, cell cycle analysis of synchronized MEFs revealed a small but reproducible acceleration of cell cycle progression from the G_1 to the S phase at 20 and 22 h after serum addition in stefin B-deficient MEFs (Fig. 3). The N-terminally truncated cathepsin L isoform was shown during S phase to localize to the cell nucleus where it cleaves the CUX1 transcription factor (13). The processed isoforms of CUX1 accelerate entry into S phase and stimulate cell proliferation (46). We reasoned that the delay in the cell cycle progression in cells overexpressing stefin B in the nucleus could be attributed to the inhibition of cathepsin L by stefin B.

Increased Expression of Stefin B in the Nucleus Protects CUX1 from Cathepsin L Cleavage—Cathepsin L was shown to cleave CUX1 at multiple sites *in vitro* and *in vivo*, thereby generating a number of processed isoforms collectively called p110 CUX1 (13). Proteolytic processing of CUX1 cannot easily be detected by Western blot analysis in T98G cells.[3] In our experiments, we used MCF-7 cells, which express large amounts of CUX1. MCF-7 cells were transiently transfected with the pEF/Myc/Nuc-stefin B expression vector (NB) or an empty vector (NO), and CUX1 expression was monitored by Western blot analysis. We observed diminished cleavage of full-length CUX1 protein (p200) and diminished p110 fragment formation in cells overexpressing stefin B or treated with the cysteine protease inhibitor E64-d, as compared with cells transfected with an empty vector (NO) (Fig. 4). The reduction in CUX1 cleavage is consistent with the notion that overexpression of stefin B causes the inhibition of cathepsin L in the nucleus.

Visualization of Intracellular Met-75 Cathepsin L-Stefin B Interactions in the Nucleus Using FRET—From structural and kinetics studies that investigated the mechanism of cystatin interaction with papain-like cathepsins, we know that the N-terminal part of cystatins is essential for the interaction with enzymes (47, 48). Therefore, a spectral variant of the GFP-T-Sapphire was fused to the C terminus of stefin B. Venus-YFP was fused to the N- terminal part of the Met-75 human cathepsin L variant. The Met-75-truncated variant of cathepsin L was

[3] A. Nepveu, unpublished data.

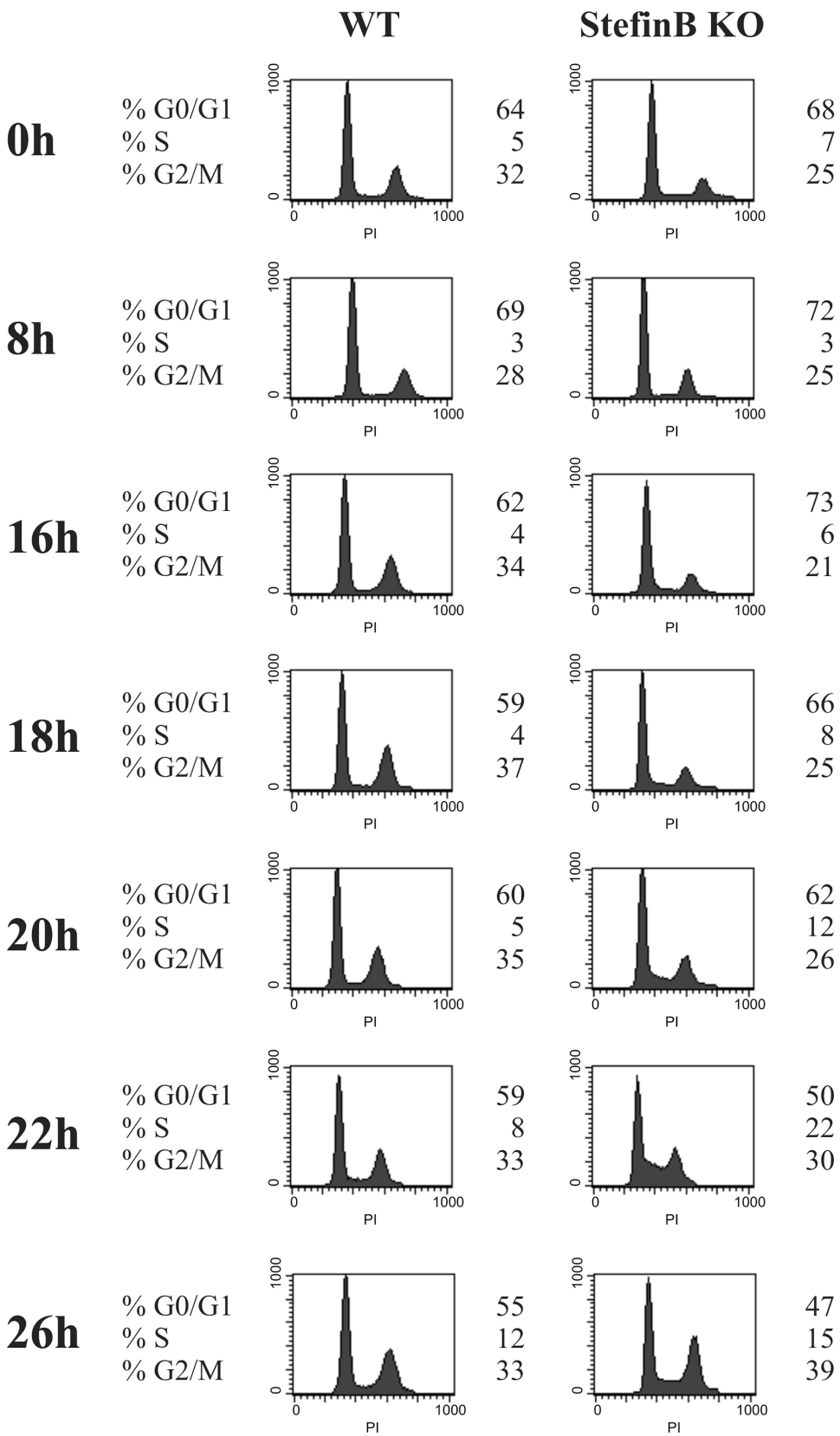


FIGURE 3. **Stefin B deficiency leads to accelerated entry into S phase.** Wild type (*WT*) and stefin B-deficient (*KO*) MEFs were incubated for 72 h in the absence of serum (time point 0). Serum-containing medium was added, and at the indicated time points cell cycle distribution was monitored by flow cytometry analysis after DNA staining with propidium iodide.

found localized in the nucleus (13). In CHO-K1 cells, preferential expression of the Met-75 cathepsin L-YFP fusion protein in the nucleus was observed (Fig. 5*A*). CHO-K1 were chosen for FRET experiments, because they are efficiently transfected (85%) with pcDNA3 expression vector. We used two negative controls as follows: first, a GFP/YFP pair where the two fluorescent proteins are supposed to interact only randomly; second, Met-75 cathepsin L-YFP and a truncated mutant of stefin B, R68*X* (R68Stop), that is found in EPM1 patients and does not inhibit cathepsin L. The biophysical properties of the R68*X* mutant have been studied in detail (49). Spectral bleed through for YFP (acceptor) was 5.2 ± 1.25% and for GFP (donor) was 25.15 ± 0.98%. Images collected in the FRET channel were corrected for donor and acceptor spectral bleed through and normalized for expression levels (NFRET), as described previously (41). NFRET values were measured in the nucleus of 25–30 individual cells for each combination in two separate experiments. Fig. 5*A* provides representative images showing expression of fusion proteins in CHO-K1 cells co-transfected with stefin B-GFP and Met-75 cathepsin L-YFP and control cells co-transfected with R68*X* stefin B-GFP. Normalized FRET (NFRET) values are represented in a gray scale and clearly reveal the existence of FRET in the nuclei of transfected cells but not in control cells transfected with the R68*X* stefin B mutant (Fig. 5*B*). To perform the analysis presented in Fig. 5*B*, the mean values of the intensity, constructed from the NFRET images recorded over the nucleus, were presented as a scatter plot (Fig. 5*C*). When the FRET channel was corrected for the spectral donor (GFP) and acceptor (YFP) bleed through, a NFRET signal in negative control experiments varied between 0 and 20 (Fig. 5*C*). Thus, we have only considered cells with values above a threshold of 20 as positives. In cells transfected with Met-75 cathepsin L-YFP and stefin

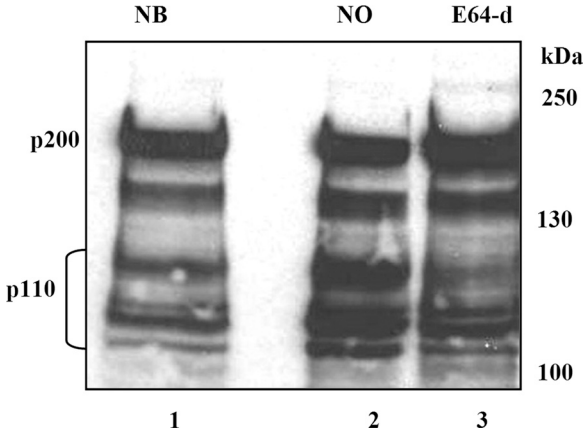


FIGURE 4. **Decreased cleavage of CUX1 transcription factor in MCF-7 cells with increased expression of stefin B in the nucleus.** MCF-7 cells were transfected with pEF/Myc/Nuc-stefin B vector (NB) (*track 1*) in comparison with control cells transfected with pEF/Nuk/Myc alone (NO) (*track 2*) and MCF-7 cells treated with 50 μM E-64d inhibitor (E-64d) (*track 3*).

B-GFP, some of the results were below this threshold (Fig. 5*C*), indicating the possibility that endogenous cathepsin L and stefin B present in the nucleus interacted with GFP fusion proteins and interfered with the results.

Histones Potentiate Inhibition of Cathepsin L by Stefin B—We established that stefin B binds to histones, but the consequences of this interaction are not known. We therefore investigated whether histones influence stefin B inhibitory activity. First, we determined conditions under which inhibition of cathepsin L was in the linear range, *i.e.* up to 70% inhibition. Next, we tested if purified histones inhibit human cathepsin L *in vitro* and found that histones did not inhibit cathepsin L activity even at a 100-fold molar excess (2.15 μM histone concentration) (Fig. 6*A*). To further investigate the role of histones in the inhibition of cathepsin L by stefin B, we preincubated stefin B with increased molar concentrations of histones and measured cathepsin L activity, as described. We found that histone binding to stefin B did not affect the inhibition of cathepsin L by stefin B at lower histone concentrations (10.8–210 nM) (Fig. 6*B*), whereas at higher histone concentrations (1.05–2.15 μM), as found in the nucleus, we observed increased inhibition of cathepsin L by stefin B (Fig. 6*B*).

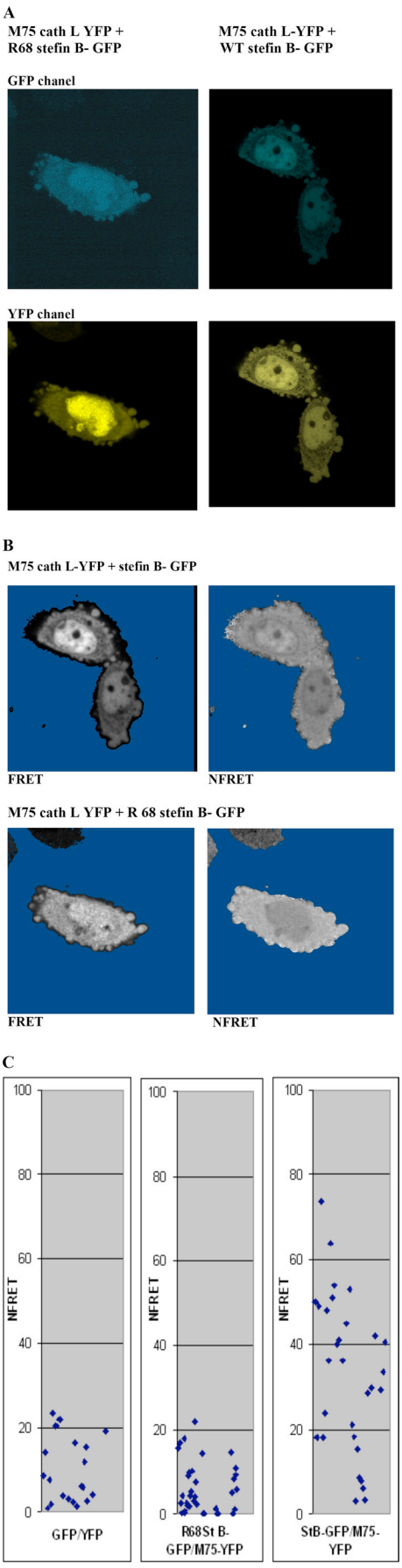
Stefin B Interacts with Histone H3 in the G_1 Phase of Cell Cycle—The above results led us to examine if the stefin B interaction with histones is cell cycle-dependent. Control T98G cells and T98G cells overexpressing stefin B in the nucleus (NB) were synchronized by serum starvation, as described above. Nuclear lysates were immunoprecipitated with anti-stefin B antibodies, and immunoblotting with anti-H3 antibodies showed that most of the binding of stefin B to histone H3 occurred in the G_1 phase of the cell cycle (Fig. 7). In a control experiment, membranes were immunoblotted with anti-stefin B antibody, confirming the presence of immunoprecipitated stefin B.

Stefin B Does Not Bind DNA—MENT, the serpin that inhibits cathepsin L in the nucleus, is known to bind to histones and DNA, thereby influencing heterochromatin distribution (29, 30). It was demonstrated that the inhibitory activity of MENT on cathepsin L, rather than DNA binding, is crucial for mediating its effect (29). DNA was reported to accelerate the rate at which MENT inhibited cathepsin V, a human ortholog of mammalian cathepsin L, up to 50-fold (50). Therefore, the possible interaction of stefin B with DNA was analyzed by gel shift assay, as described previously (50). However, stefin B did not bind to DNA, in contrast to cathepsin V, which was used as a positive control (Fig. 8). These results therefore suggest that the interaction of stefin B with histones is different from the binding of serpins like MENT, which react with both histones and DNA.

DISCUSSION

Proteolytic activity of cathepsin L in the nucleus has to be under strict control to prevent degradation of transcription factors. Among cystatins, stefin B has been observed in the nucleus, which prompted us to investigate the role of stefin B as a regulator of proteolytic activity of cathepsin L in the nucleus. In this study, we describe for the first time an interaction between a nuclear cystatin, stefin B, and the nucleosomal proteins histone H2A.Z, H2B, and H3 (Fig. 1*A*). The histone H2A.Z variant differs from the canonical H2A in its N-terminal tail sequence and also at key internal residues. This variant has been implicated in several biological processes, such as gene activation, chromosome segregation, heterochromatin silencing, and progression through the cell cycle (51, 52). Histones H2A.Z and H3K4me3 are found at promoter regions, and it has been suggested that these two histone variants partly co-localize to the same nucleosomes (42, 53). A co-immunoprecipitation experiment confirmed that the H3K4me3-modified histone also associates with stefin B in T98G cells (Fig. 1*B*).

In T98G cells in which stefin B was targeted to the nucleus, we have observed a delay in cell cycle progression into S phase (Fig. 2). Increased expression of stefin B into the nucleus resulted in a delay in cell cycle progression comparable with treatments with the synthetic cysteine protease inhibitors E-64d or JPM-OEt (13, 44). Our initial observations were confirmed with the reverse experiment using MEFs from stefin B-deficient mice. Following exit from quiescence, stefin B-deficient MEFs reached the S phase faster than wild type MEFs (Fig. 3). Goulet *et al.* (13) has shown that only shorter procathepsin L isoforms (that start at Met-56, Met-75, Met-77, Met-81, or Met-111) could translocate to the nucleus and stimulate processing of the CUX1 transcription factor at the G_1/S transition of the cell cycle. The processed p110 CUX1 isoform exhibits distinct DNA binding and transcriptional properties: cells overexpressing p110 CUX1 reached the next S phase faster than control cells (13, 46). Therefore, we hypothesized that the cell cycle delay in cells overexpressing stefin B could be attributed to cathepsin L inhibition in the nucleus. Indeed, the steady-state level of p110 CUX1 was reduced in cells overexpressing stefin B, suggesting that cleavage of CUX1 in the nucleus was partially prevented (Fig. 4). Interaction of stefin B with the Met-75 truncated form of cathepsin L in the nucleus was shown by FRET experiments in living cells (Fig. 5, *B* and *C*). This interaction was not entirely unexpected, as *in vitro* kinetic studies have shown that stefin B inhibits cathepsin L with a K_i value in the picomolar range (54, 55). Although increased nuclear localization of procathepsin L in *ras*-transformed mouse fibroblasts was reported more than a decade ago (56), the first physiologically relevant substrates of cathepsin L in the nucleus were



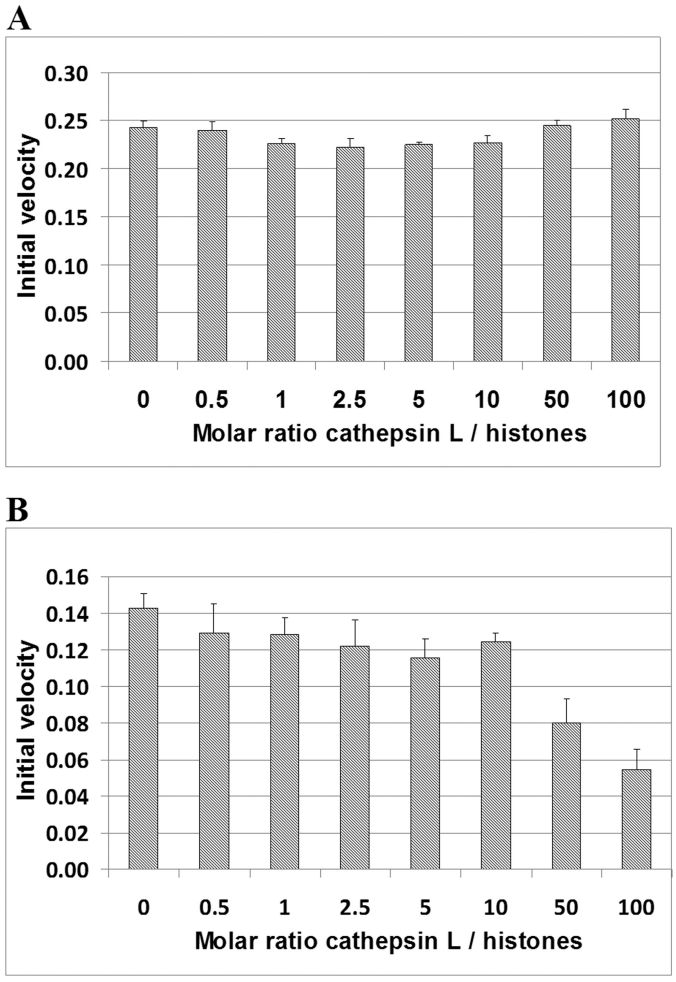


FIGURE 6. **Binding of stefin B to histone increases cathepsin L inhibition by stefin B.** *A,* cathepsin L is not inhibited by histones. Cathepsin L (21.5 nM) was incubated with increasing molar concentrations of histones. The data represent the mean of at least three independent experiments. *B,* addition of histones increased the inhibitory activity of stefin B toward cathepsin L. Stefin B (15 nM) was preincubated with increasing molar concentrations of histones (10.8 nM to 2.15 μM), before the addition of cathepsin L (21.5 nM). The data represent the mean of at least three independent experiments.

described only recently (13, 15, 57). Duncan *et al.* (15) reported that cathepsin L cleaves histone H3 and proposed that this proteolytic activity plays a role in the development and differentiation of mouse embryonic stem cells. We have examined whether the binding of histones to stefin B affects cathepsin L inhibition. *In vitro*, high histone concentrations increased the inhibition of cathepsin L by stefin B. Recently, it was shown that cathepsin L is not the only protease present in the nucleus.

FIGURE 5. **Met-75 cathepsin L and stefin B interact in the nucleus of living cells.** *In situ* analysis of the interaction between stefin B-GFP and Met-75 cathepsin L was measured by FRET. *A,* visualization of Met-75 cathepsin L YFP and Arg-68 stefin B-GFP in CHO K1 cells. The plasmid constructs indicated on *top* of each column were transfected in CHO K1 cells. *B,* visualization of FRET and NFRET in CHO K1 cells pixel-by-pixel analysis of FRET on a cell expressing Met-75 cathepsin L-YFP + stefin B-GFP, performed with the PixFRET plug-in for the ImageJ software. *C,* NFRET in the nuclei of cells transfected with Met-75 cathepsin L YFP + stefin B compared with NFRET in cells transfected with expression vectors for Met-75 cathepsin L/Arg-68 stefin B-GFP. NFRET plot profiles in 25–30 cells were analyzed. Sensitized emission of YFP fusion proteins due to FRET was measured in two separate experiments in the nucleus of 25 individual CHO K1 cells transfected with the indicated combinations of vectors expressing YFP and GFP fusion proteins.

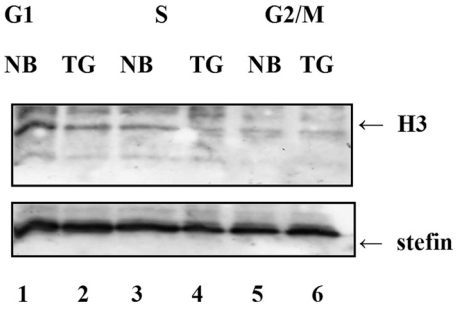


FIGURE 7. **Stefin B interacts with histone H3 predominantly in the G_1 phase of the cell cycle.** T98G cells were synchronized by serum starvation. Nuclear extracts were prepared from T98G cells (*TG*) after 8 h (*G1*), 22 h (*S*), and 26 h (*G2/M*). Protein extracts from T98G cells transfected with stefin B in pEF/Myc/Nuc vector (*NB*) were separated by 15% SDS-PAGE, followed by Western blotting with anti-H3 antibody. Membranes were stripped and immunoblotted with anti-stefin B antibody.

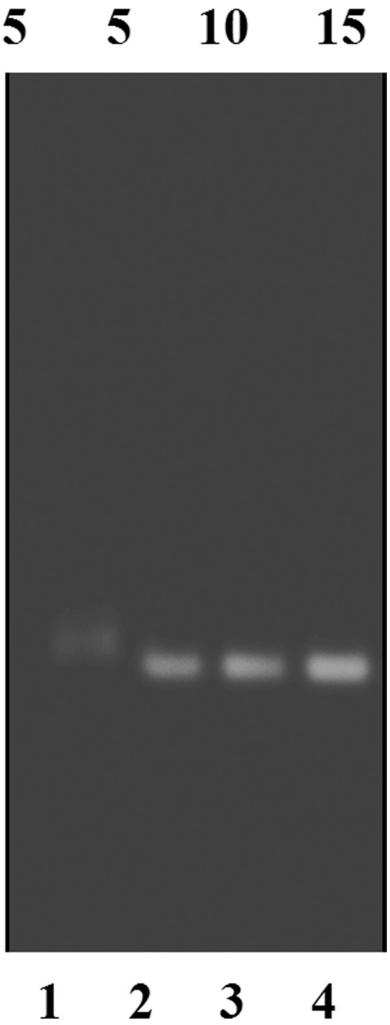


FIGURE 8. **Analysis of stefin B and cathepsin V binding to DNA.** Gel mobility shift analysis of the control protein cathepsin V (*track 1*) and stefin B (*tracks 2–4*) was incubated with ds65-mer DNA. DNA was analyzed by 2% agarose gel electrophoresis. The final concentration (μM) of purified protein in each reaction is indicated at the *top* of each gel panel.

Although it was proposed by Goulet *et al.* (13, 57) that a truncated form of cathepsin F could be active in the nucleus, its nuclear activity was reported only recently. Furthermore, Maubach *et al.* (58) reported that the regulation of nuclear cathepsin F activity by stefin B in hepatic stellate cells was involved in the transcriptional regulation of two activation markers. Together, these findings suggest that the regulatory activity of stefin B in the nucleus may not be limited to its effect on cathepsin L.

MENT, a serpin that also inhibits cathepsins L and V, strongly blocks cell proliferation and promotes condensation of chromatin (29). It was shown that MENT-mediated inactivation of cathepsin L, like cathepsin L deficiency, causes a global rearrangement of chromatin structure and redistribution of specifically modified histones (14, 29). MENT was also reported to interact with DNA (30, 59). Another cysteine protease-inhibiting serpin, SCCA-1 (squamous cell carcinoma antigen-1), has also been shown to localize to the nucleus, but unlike MENT, SCCA-1 is unable to bind to DNA (50). Similarly, we did not observe the binding of stefin B to DNA (Fig. 8). It remains to be tested whether cathepsin inhibitors that do not bind DNA, SCCA-1 and stefin B, have similar effects on heterochromatin redistribution as MENT.

Collectively our results show that stefin B interacts with histones and cathepsin L in the nucleus and regulates cell cycle progression into the S phase. Stefin B was bound to histones preferentially during the G_1 phase of the cell cycle. *In vitro*, high concentrations of histones increased the inhibitory effect of stefin B on cathepsin L activity. Interestingly, entry into S phase is delayed similarly in the presence of the cathepsin L inhibitors E-64d and JPM-OEt or following overexpression of stefin B (Fig. 2) (13, 44).

In addition to the role of stefin B in preventing unwanted cytosolic protease activity resulting from lysosomal leakage, our results demonstrate that stefin B plays an important role in the regulation of cathepsin L proteolytic activity in the nucleus. Stefin B in the nucleus protects cathepsin L substrates from proteolytic processing and consequently participates in transcriptional regulation. It will be interesting to determine whether the loss of cathepsin L inhibition by nonfunctional mutants of stefin B in EPM1 disease is associated with the excessive cleavage of cathepsin L nuclear substrates.

Acknowledgments—The expert technical assistance of Loulou Kroon-Žitko is gratefully acknowledged. We thank Susanne Liebe (Leica Microsystems, Germany) for expert technical help with initial FRET measurements. We thank Roman Jerala (National Institute of Chemistry, Slovenia) for critical reading of the manuscript. We also thank Oliver Griesback (Max-Planck Institute, Germany) for providing plasmid with T-Sapphire GFP and Atsushi Miyawaki (Brain Science Institute, RIKEN, Japan) for plasmid with Venus YFP. We thank. R. M. Myres (Stanford School of Medicine and the Stanford Human Genome Center, Stanford University) for the stefin B-deficient mice.

REFERENCES

1. Turk, V., Turk, B., and Turk, D. (2001) *EMBO J.* **20,** 4629–4633
2. Hsing, L. C., and Rudensky, A. Y. (2005) *Immunol. Rev.* **207,** 229–241
3. Reinheckel, T., Deussing, J., Roth, W., and Peters, C. (2001) *Biol. Chem.* **382,** 735–741
4. Nakagawa, T., Roth, W., Wong, P., Nelson, A., Farr, A., Deussing, J., Villadangos, J. A., Ploegh, H., Peters, C., and Rudensky, A. Y. (1998) *Science* **280,** 450–453
5. Yasothornsrikul, S., Greenbaum, D., Medzihradszky, K. F., Toneff, T.,

Bundey, R., Miller, R., Schilling, B., Petermann, I., Dehnert, J., Logvinova, A., Goldsmith, P., Neveu, J. M., Lane, W. S., Gibson, B., Reinheckel, T., Peters, C., Bogyo, M., and Hook, V. (2003) *Proc. Natl. Acad. Sci. U.S.A.* **100,** 9590–9595
6. Reinheckel, T., Hagemann, S., Dollwet-Mack, S., Martinez, E., Lohmüller, T., Zlatkovic, G., Tobin, D. J., Maas-Szabowski, N., and Peters, C. (2005) *J. Cell Sci.* **118,** 3387–3395
7. Chandran, K., Sullivan, N. J., Felbor, U., Whelan, S. P., and Cunningham, J. M. (2005) *Science* **308,** 1643–1645
8. Simmons, G., Gosalia, D. N., Rennekamp, A. J., Reeves, J. D., Diamond, S. L., and Bates, P. (2005) *Proc. Natl. Acad. Sci. U.S.A.* **102,** 11876–11881
9. Qiu, Z., Hingley, S. T., Simmons, G., Yu, C., Das Sarma, J., Bates, P., and Weiss, S. R. (2006) *J. Virol.* **80,** 5768–5776
10. Puente, X. S., Sánchez, L. M., Overall, C. M., and López-Otín, C. (2003) *Nat. Rev. Genet.* **4,** 544–558
11. Brömme, D., Li, Z., Barnes, M., and Mehler, E. (1999) *Biochemistry* **38,** 2377–2385
12. Santamaría, I., Velasco, G., Cazorla, M., Fueyo, A., Campo, E., and López-Otín, C. (1998) *Cancer Res.* **58,** 1624–1630
13. Goulet, B., Baruch, A., Moon, N. S., Poirier, M., Sansregret, L. L., Erickson, A., Bogyo, M., and Nepveu, A. (2004) *Mol. Cell* **14,** 207–219
14. Bulynko, Y. A., Hsing, L. C., Mason, R. W., Tremethick, D. J., and Grigoryev, S. A. (2006) *Mol. Cell. Biol.* **26,** 4172–4184
15. Duncan, E. M., Muratore-Schroeder, T. L., Cook, R. G., Garcia, B. A., Shabanowitz, J., Hunt, D. F., and Allis, C. D. (2008) *Cell* **135,** 284–294
16. Turk, V., Stoka, V., and Turk, D. (2008) *Front. Biosci.* **13,** 5406–5420
17. Lenarcic, B., and Turk, V. (1999) *J. Biol. Chem.* **274,** 563–566
18. Schick, C., Pemberton, P. A., Shi, G. P., Kamachi, Y., Cataltepe, S., Bartuski, A. J., Gornstein, E. R., Brömme, D., Chapman, H. A., and Silverman, G. A. (1998) *Biochemistry* **37,** 5258–5266
19. Hwang, S. R., Stoka, V., Turk, V., and Hook, V. Y. (2005) *Biochemistry* **44,** 7757–7767
20. Abrahamson, M., Alvarez-Fernandez, M., and Nathanson, C. M. (2003) *Biochem. Soc. Symp.* **70,** 179–199
21. Kopitar-Jerala, N. (2006) *FEBS Lett.* **580,** 6295–6301
22. Riccio, M., Di Giaimo, R., Pianetti, S., Palmieri, P. P., Melli, M., and Santi, S. (2001) *Exp. Cell Res.* **262,** 84–94
23. Lalioti, M. D., Scott, H. S., Buresi, C., Rossier, C., Bottani, A., Morris, M. A., Malafosse, A., and Antonarakis, S. E. (1997) *Nature* **386,** 847–851
24. Lalioti, M. D., Mirotsou, M., Buresi, C., Peitsch, M. C., Rossier, C., Ouazzani, R., Baldy-Moulinier, M., Bottani, A., Malafosse, A., and Antonarakis, S. E. (1997) *Am. J. Hum. Genet.* **60,** 342–351
25. 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., and Myers, R. M. (1996) *Science* **271,** 1731–1734
26. Lehesjoki, A. E. (2003) *EMBO J.* **22,** 3473–3478
27. Eldridge, R., Iivanainen, M., Stern, R., Koerber, T., and Wilder, B. J. (1983) *Lancet* **2,** 838–842
28. Alakurtti, K., Weber, E., Rinne, R., Theil, G., de Haan, G. J., Lindhout, D., Salmikangas, P., Saukko, P., Lahtinen, U., and Lehesjoki, A. E. (2005) *Eur. J. Hum. Genet.* **13,** 208–215
29. Grigoryev, S. A., Bednar, J., and Woodcock, C. L. (1999) *J. Biol. Chem.* **274,** 5626–5636
30. Irving, J. A., Shushanov, S. S., Pike, R. N., Popova, E. Y., Brmme, D., Coetzer, T. H., Bottomley, S. P., Boulynko, I. A., Grigoryev, S. A., and Whisstock, J. C. (2002) *J. Biol. Chem.* **277,** 13192–13201
31. Kouzarides, T. (2007) *Cell* **128,** 693–705
32. Kopitar-Jerala, N., Curin-Serbec, V., Jerala, R., Krizaj, I., Gubensek, F., and Turk, V. (1993) *Biochim. Biophys. Acta* **1164,** 75–80
33. Pennacchio, L. A., Bouley, D. M., Higgins, K. M., Scott, M. P., Noebels, J. L., and Myers, R. M. (1998) *Nat. Genet.* **20,** 251–258
34. Kopitar-Jerala, N., Schweiger, A., Myers, R. M., Turk, V., and Turk, B. (2005) *FEBS Lett.* **579,** 2149–2155
35. Nagai, T., Ibata, K., Park, E. S., Kubota, M., Mikoshiba, K., and Miyawaki, A. (2002) *Nat. Biotechnol.* **20,** 87–90
36. Zapata-Hommer, O., and Griesbeck, O. (2003) *BMC Biotechnol.* ***3***, **5**, 1–6
37. Dolinar, M., Maganja, D. B., and Turk, V. (1995) *Biol. Chem. Hoppe-Seyler* **376,** 385–388
38. Stein, G. H. (1979) *J. Cell. Physiol.* **99,** 43–54
39. Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983) *Nucleic Acids Res.* **11,** 1475–1489
40. Zimmermann, T., Rietdorf, J., Girod, A., Georget, V., and Pepperkok, R. (2002) *FEBS Lett.* **531,** 245–249
41. Feige, J. N., Sage, D., Wahli, W., Desvergne, B., and Gelman, L. (2005) *Microsc. Res. Tech.* **68,** 51–58
42. Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007) *Cell.* **129,** 823–837
43. Schmid, C. D., and Bucher, P. (2007) *Cell* **131,** 831–833
44. Mellgren, R. L. (1997) *Biochem. Biophys. Res. Commun.* **236,** 555–558
45. Hannigan, G., and Williams, B. R. (1986) *EMBO J.* **5,** 1607–1613
46. Sansregret, L., Goulet, B., Harada, R., Wilson, B., Leduy, L., Bertoglio, J., and Nepveu, A. (2006) *Mol. Cell. Biol.* **26,** 2441–2455
47. Machleidt, W., Thiele, U., Laber, B., Assfalg-Machleidt, I., Esterl, A., Wiegand, G., Kos, J., Turk, V., and Bode, W. (1989) *FEBS Lett.* **243,** 234–238
48. Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B., and Turk, V. (1990) *EMBO J.* **9,** 1939–1947
49. Rabzelj, S., Turk, V., and Zerovnik, E. (2005) *Protein Sci.* **14,** 2713–2722
50. Ong, P. C., McGowan, S., Pearce, M. C., Irving, J. A., Kan, W. T., Grigoryev, S. A., Turk, B., Silverman, G. A., Brix, K., Bottomley, S. P., Whisstock, J. C., and Pike, R. N. (2007) *J. Biol. Chem.* **282,** 36980–36986
51. Dhillon, N., Oki, M., Szyjka, S. J., Aparicio, O. M., and Kamakaka, R. T. (2006) *Mol. Cell. Biol.* **26,** 489–501
52. Jin, C., and Felsenfeld, G. (2007) *Genes Dev.* **21,** 1519–1529
53. Schones, D. E., Cui, K., Cuddapah, S., Roh, T. Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008) *Cell* **132,** 887–898
54. Kastelic, L., Turk, B., Kopitar-Jerala, N., Stolfa, A., Rainer, S., Turk, V., and Lah, T. T. (1994) *Cancer Lett.* **82,** 81–88
55. Pol, E., and Björk, I. (2003) *Biochim. Biophys. Acta* **1645,** 105–112
56. Hiwasa, T., and Sakiyama, S. (1996) *Cancer Lett.* **99,** 87–91
57. Goulet, B., Sansregret, L., Leduy, L., Bogyo, M., Weber, E., Chauhan, S. S., and Nepveu, A. (2007) *Mol. Cancer Res.* **5,** 899–907
58. Maubach, G., Lim, M. C., and Zhuo, L. (2008) *Mol. Biol. Cell* **19,** 4238–4248
59. Istomina, N. E., Shushanov, S. S., Springhetti, E. M., Karpov, V. L., Krasheninnikov, I. A., Stevens, K., Zaret, K. S., Singh, P. B., and Grigoryev, S. A. (2003) *Mol. Cell. Biol.* **23,** 6455–6468

The extent of displacement of the occluding loop of cathepsin B corresponds to the size of a ligand

Journal:	*FEBS Journal*
Manuscript ID:	FJ-10-0570
Manuscript Type:	Regular Paper
Subdiscipline:	Structural biology
Date Submitted by the Author:	14-Jun-2010
Complete List of Authors:	Renko, Miha; Jozef Stefan Institute, Department of Biochemistry and Molecular and Structural Biology Pozgan, Urska; Jozef Stefan Institute, Department of Biochemistry and Molecular and Structural Biology Majera, Dusana; Jozef Stefan Institute, Department of Biochemistry and Molecular and Structural Biology Turk, Dusan; Jozef Stefan Inst., Dept. of Biochemistry & Molecular Biology; Jozef Stefan Institute, Department of Biochemistry and Molecular and Structural Biology
Key Words:	Cathepsin B, occluding loop, stefin A, cystatins, crystal structure, complex, conformational flexibility

The extent of displacement of the occluding loop of cathepsin B corresponds to the size of a ligand

Miha Renko, Urška Požgan, Dušana Majera, Dušan Turk

Department of Biochemistry and Molecular and Structural Biology**,** Jozef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia

Corresponding author:
Dušan Turk
Department of Biochemistry and Molecular and Structural Biology**,** Jozef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia, phone: +386 1 477 3215, fax: +386 1 477 3984, e-mail: dusan.turk@ijs.si

RUNNING TITLE
Cathepsin B occluding loop swings out in the complex with stefin A

KEYWORDS:
Cathepsin B, occluding loop, stefin A, cystatins, crystal structure, complex, conformational flexibility

SUBDIVISION: Structural biology

SUMMARY

Cathepsin B is one of the most versatile human cysteine cathepsins. It is important for intracellular protein degradation under normal conditions and is involved in a number of pathological processes. The occluding loop makes cathepsin B unique among cysteine cathepsins. This ~20-residue long insertion imbedded into the papain-like protease scaffold restricts access to the active site cleft and endows cathepsin B with its carboxydipeptidase activity. Nevertheless, the enzyme also exhibits endopeptidase activity and is inhibited by stefins and cystatins. To explain the structural properties of the occluding loop upon binding of stefins, we have determined the crystal structure of the complex between the wild type human stefin A and the wild type human cathepsin B, at 2.6 Å resolution. The papain-like part of cathepsin B structure remains unmodified, whereas the occluding loop residues are displaced. The part enclosed by the disulfide bridge containing histidines 110 and 111, the "lasso" part, is rotated by approximately 45 degrees away from its original position. A comparison of the structure of the unliganded cathepsin B with its complexes with chagasin and stefin A, as well as with the structure of the proenzyme, shows that the magnitude of the shift of the occluding loop is related to the size of the competing ligand, but with no impact on the binding constant. Hence, cathepsin B can dock inhibitors and certain substrates regardless of their size.

INTRODUCTION

Cathepsin B (EC 3.4.22.1), a lysosomal, papain-like cysteine protease is one of the most extensive studied human cathepsins [1]. This enzyme is abundantly expressed in a variety of tissues where it takes part in protein degradation and processing. It is involved in a number of physiological and pathological processes, such as intracellular protein degradation, immune response, prohormone processing, cancer, and arthritis [2-9]. Its proteolytic activity is regulated by stefins and cystatins, endogenous inhibitors of cysteine cathepsins [10]. Cathepsin B differs from other cathepsins by its dual role, exhibiting exo- as well as endopeptidase activity. The crystal structure of this human enzyme [11] has revealed that about a 20 residues long insertion, termed “occluding loop”, occupies the part of the active site cleft on the primed side and blocks access to the active site cleft beyond the S2' substrate binding site [11, 12]. The occluding loop is held together by the disulfide bond between C108 and C119. Its attachment to the body of the enzyme is stabilized by two salt bridges, between H110 and D22, and between R116 and D224. The crystal structure suggested that two histidines, H110 and H111, positioned within the active site cleft, are responsible for docking of the C-terminal carboxylic group of peptidyl substrates. This observation has been later confirmed by the crystal structure of the complex of a substrate-mimicking inhibitor, CA030, interacting through its C-terminal carboxylic group with the two histidine residues [13]. The concept of utilizing additional structural features to block part of the active site cleft in order to restrict the binding of peptidyl substrates and facilitating binding of the substrate termini is not unique to cathepsin B [14]. The amino dipeptidase cathepsin C [15, 16] possesses a large segment of the proregion [17], termed exclusion domain, which remains associated with the mature enzyme and blocks the active site cleft beyond the S2 site. The amino peptidase cathepsin H has covalently attached stretch of eight residues originating from the propeptide, termed the mini chain, which blocks the unprimed binding site [18]. The mini loop in carboxypeptidase cathepsin X blocks the primed side of the active site, restricting access to only one residue [19].

While the structures of the mature native form of cathepsin B clearly exposed the relevance of the occluding loop for the exopeptidase activity [11], they have not explained the mechanisms of endopeptidase activity nor the inhibition of the enzyme by their endogenous protein inhibitors cystatins and stefins [20]. A further step in

understanding of these mechanisms was provided by the crystal structures of human [21] and rat procathepsins B [22]. They have revealed that, in the zymogen form, the propeptide rather than the occluding loop fills the active site cleft. It was shown that the single and double mutations D22A, H110A, R116A, and D224A disrupted the salt bridges between the occluding loop and the body of the enzyme, resulting in enhanced endopeptidase activity [23]. Furthermore, the deletion mutant lacking 12 central residues of the "lasso" region between the disulfide C109-C118 confirmed that their absence yields an enzyme with pure endopeptidase activity, completely lacking exopeptidase activity, and with a 40-fold increase of affinity for cystatins [12]. These results indicated that loop flexibility must be responsible for the endopeptidase activity of cathepsin B, as well as that endopeptidase activity should be associated with the occluding loop displacement from the active site cleft. Recently, crystal structure of the complex between chagasin, a cysteine protease inhibitor from *Trypanosoma cruzi*, and human cathepsin B, a multiple mutant with destabilized affinity of the occluding loop residues towards the active site cleft, has shown that on binding to cathepsin B chagasin displaces the occluding loop from the active cleft [24]. Here we present the crystal structure of the complex between two human proteins: the wild type stefin A and the wild type human cathepsin B. A structural comparison suggests that the extent of the movement of the occluding loop residues necessary for their displacement from the active site cleft is ligand size dependent.

RESULTS AND DISCUSSION

Crystals of the complex of stefin A and cathepsin B contain complete wild type protein sequences. The positioning of the main chains of nearly all residues is clearly revealed by the electron density maps, with the exception of E95, a stretch of four occluding loop residues from V112 to S115 in the first molecule of cathepsin B, G75 and Q76 in the molecule A of stefin A, and M1 and E78 in the molecule B of stefin A. Additionally, eleven side chains lack adequate electron density. The RMS deviation between all pairs of superimposed CA atoms of cathepsin B molecules, excluding residues 105-125 of the occluding loop, is 0.34 Å, whereas the RMS deviation between all pairs of superimposed CA atoms of stefin A molecules exhibits somewhat larger RMS of 0.88 Å. This comparison shows that the differences between the two molecules of cathepsin B are confined to the occluding loop region, whereas the differences between the two stefin A molecules are spread out through the entire

structure, with slightly increased variability in the S72–D79 region that forms the second binding loop.

Cathepsin B has a two-domain, papain-like fold [11]. The N-terminal domain includes the central helix that contains, on its N terminus, the active site C29. The C-terminal domain is based on a 4-stranded β-barrel fold, contributing H199, the other active site residue. The active site cleft is formed at the interface between the two domains, that are also named L- and R- (left and right), according to the standard view used to present the papain-like folds.

The structure of stefin A exhibits the cystatin-like fold composed of a five-stranded β-sheet embracing an α-helix (Fig. 1). This arrangement creates a wedge-shaped structure with the N-terminal trunk and two hairpin loops at its narrow edge [25]. This narrow edge docks into the active site cleft of cathepsin B (Fig. 1). The binding mode is equivalent to those from the related complexes of stefin B-papain [26] and stefin A-cathepsin H[27]. A comparison of the average distances between CA atoms of the active site cysteine and histidine residues in cathepsins B and H and CA atoms of stefins in the structures of both complexes showed that stefin A binds to cathepsin B as deep as does stefin B to cathepsin H, with the equivalent average distances 23.36 Å and 23.43 Å, respectively (Table 1). This shows that the final position of stefin A molecules in the complex is not effected by the additional features of exopeptidases, occluding loop and mini-chain, which occupy parts of the active site cleft (Figure 4). These additional features hinder binding along the whole interdomain interface, yet they both get pushed away upon binding of the ligand.

The N-terminal trunk and the first binding loop occlude the active site C29, blocking the enzymatic activity. The N-terminal trunk binds into the non-primed substrate binding sites, whereas the two loops bind into the primed sites. They occlude the catalytic C29 (surface colored in yellow) in the middle and thereby prevent the approach of substrate molecules. The same approach is utilized by the p41 fragment, a representative of thyropins [28], chagasin [29, 30], and mycocypins [31].

The N-terminal trunk comes down the S1 binding area of cathepsin B, occupies the S2 binding site with proline residue P3 and continues through the S2 binding site upwards (away from the cathepsin B surface). Two hydrogen bonds between the stefin A amide hydrogen (G4) and carbonyl (P3) with cathepsin B carbonyl atom

(G198) and amide hydrogen (G74) attach the first loop to the active site cleft.

The first binding loop of stefin A (V47 to Q51) fills the S1' site with V48. Besides this hydrophobic interaction, the loop is fastened to the cathepsin B surface by the hydrogen bond between the stefin A A49 amide and cathepsin B G24 carbonyl. The binding of this loop is further stabilized by a hydrogen bond between the stefin A N52 side chain amide and the cathepsin B S25 carbonyl group.

The second binding loop (L73 to D79) comes down to the area beyond the S2' site and displaces the occluding loop residues of cathepsin B. It is firmly anchored by the β-sheet hydrogen bonding pattern formed between the three loops in stefin A and an additional hydrogen bond formed between the amide hydrogen of L73 and the side chain carbonyl of E109. A layer of solvent molecules mediates the contacts between the C-terminal part of the second binding loop and cathepsin B.

The occluding loop differs from the native structure (PDB code 1HUC) [11] in the region from S104 and D124 (Figs. 2, 3). The lasso structure between the C108 - C119 disulfide is rotated by approximately 45° and pushed aside. This movement dramatically changes the position of the two occluding loop histidines, H110 and H111. Instead of a parallel positioning within the active site cleft, these two side chains now point into different, almost opposite directions. The side chain of H110 points away from the active site cleft to the back of the molecule, while the side chain of H111 points upwards and away from the surface. In the complex, two stefin A residues, A49 from the tip of the first binding loop and L73 from the second binding loop, fill the places that the two histidines occupy in the native structure. Besides the lasso, the inhibitor also pushes away the chain from C119 to the D124. The position of CA atom of E122 is changed by almost 7 Å from the position it occupies in the native cathepsin B structure. In this respect, stefin interactions with exopeptidases are not unique. The N-terminal trunk of stefin A can displace the mini chain which blocks part of the binding cleft in cathepsin H [27].

Two salt bridges, H110 – D22 and R116 – D224, which additionally stabilize the attachment of the loop to the body of the enzyme, are disrupted in the complex. R116 and D224, however, compensate for the loss of the salt bridge interaction by finding electrostatically favorable partners in K184 of cathepsin B and E78 of stefin A, respectively. The structure presented here shows that weakening of embedding of the

occluding loop into the active site cleft is not mandatory for formation of the crystals of the complex, even though it is associated in a drop of K_i from 0.93 to 0.35 nM, as shown by the chagasin – cathepsin B study. The stefin A - cathepsin B complex contains the wild type sequences and physiologically occurring interactions, as opposed to the crystal structure of chagasin, a parasite inhibitor from *Tripanosoma cruzi*, and cathepsin B complex [24](PDB code 3CBJ). In that complex the first salt bridge interaction has been disrupted by the H110A mutant and the enzyme's reactive site turned off by the C29A mutant. (We assume here that the cathepsin B mutations have not affected the geometry of binding of chagasin.) The wild type sequences have also been preserved in the related structural studies of procathepsin B [21].

These three structures, as well as the structure of native cathepsin B (Figs. 2, 3) demonstrate that the occluding loop can adopt a variety of positions, with the moving part consisting of residues between E109 and D124. The extent of the occluding loop shift from the position in the native enzyme (PDB code 1HUC) is shown in a series of structures starting with the proenzyme form (PDB code 3PBH), complex with stefin A, and chagasin [24] (PDB code 3CBJ) (Figs. 2, 3). The CA atom position of N113 is marked in Fig. 3 to indicate the shifted positions which are 7 Å, 16 Å, and 22.5 Å (14 Å) away from the position that this atom occupies in the native form. Our conclusion is that the size matters. The larger and the wider are the features of the ligands that compete with the occluding loop for binding to the active site, the farther away the occluding loop residues are shifted. Hence, these structures demonstrate that the occluding loop residues can adopt a variety of conformations, whereas the rest of the structure of cathepsin B appears to be rigid. A comparison of the interaction constants of binding of chagasin (K_i=0.93nM [24]) and stefins (1.7 and 2 nM [32, 33], 0.91 nM [34]) to cathepsin B indicate that the extent of the shift does not affect the inhibition constants, even though the interaction surface of chagasin with the occluding loop (160 $Å^2$) is slightly larger than that of stefin A (100 $Å^2$). This observation suggests that the energy cost of ligand binding associated with the occluding loop removal is not related to the magnitude of the occluding loop shift from the active site cleft. Cathepsin B can bind certain ligands along the whole interdomain interface. During docking their size alone likely does not play a role. Cathepsin B will accept inhibitors or substrates – whatever comes across.

MATERIALS AND METHODS

Cathepsin B and stefin A were expressed as previously reported [35, 36], mixed in a molar ratio 1:1.1, and concentrated to 30 mg/mL in 10 mM sodium acetate, pH=5.5. Crystals were grown in 0.2 M sodium sulfate, 24% PEG3000. The initial crystals grown by the sitting drop method were highly mosaic, thereby useless for structure determination. Therefore, the hanging drop method was used in combination with the controlled evaporation approach [37], which greatly improved crystal quality. The crystals, which grew in the form of thin plates, were soaked in mother liquor supplemented with 20-30% glycerol and frozen in liquid nitrogen prior to data collection.

Diffraction data were collected at the XRD1 workstation at Synchrotron Elletra, Trieste, and processed using the HKL2000 package [38]. Determination of the space group was non-trivial. The data were first processed in the $P2_1$ space group due to the higher symmetry, with an acceptable R_{merge} of 0.132 and data completeness of 96.7%. The structure was determined by molecular replacement using Amore [39] with cathepsin B [13] and stefin A [27] as search models. The crystals are extremely dense, having only 28% of solvent, resulting in Matthews coefficient (V_M) of 1.70 [40]. It was surprising that so tightly packed crystals diffracted only to 2.6 Å. The protein database analysis took into account 10,471 crystal forms of proteins, deposited in PDB in 2002 [41]. It showed that more tightly packed crystals (lower V_M) tend to diffract to higher resolutions.

Since we were unable to position the occluding loop residues consistently within the electron density maps, we decided to reprocess the diffraction data in the lower symmetry space group, P1. These data had a lower R_{merge} of 0.084 and slightly lower completeness (92.4%). The lower completeness of the P1 data set is a consequence of highly anisotropic diffraction, which forced us to discard part of the collected data to maintain reasonable merging statistics. The anisotropy was a consequence of the shape of the crystals, which were thin plates diffracting poorly in the direction perpendicular to the beam. The P1 space group data resulted in an improved electron density map for the occluding loop residues and were used for further refinement and model building. The structure was refined using Refmac [42] and MAIN [43].

Data collection and refinement statitistics are summarized in Table 2. The coordinates and structure factors were deposited in the PDB (ID 3K9M).

Distance d (table 1) between stefin A and different enzymes is the average distance between all CA atoms in stefin A and CA atoms of reactive site cysteine and histidine residues.

ACKNOWLEDGEMENTS

This work was supported by the Slovenian Research Agency Grants No. P1-0048, P1-0140, a Marie Curie Fellowship of the European Community programme Drugs for Therapy (MRTN-CT-2004-512385) to D.M. and a Young Researcher fellowship to M.R. and U.P.

REFERENCES

1. Vasiljeva O, Reinheckel T, Peters C, Turk D, Turk V & Turk B (2007) Emerging roles of cysteine cathepsins in disease and their potential as drug targets. *Curr Pharm Des* **13**, 387-403.
2. Yan S & Sloane BF (2003) Molecular regulation of human cathepsin B: implication in pathologies. *Biol Chem* **384**, 845-854, doi: 10.1515/BC.2003.095.
3. Turk V, Turk B, Guncar G, Turk D & Kos J (2002) Lysosomal cathepsins: structure, role in antigen processing and presentation, and cancer. *Adv Enzyme Regul* **42**, 285-303.
4. Mohamed MM & Sloane BF (2006) Cysteine cathepsins: multifunctional enzymes in cancer. *Nat Rev Cancer* **6**, 764-775.
5. Pozgan U, Caglic D, Rozman B, Nagase H, Turk V & Turk B Expression and activity profiling of selected cysteine cathepsins and matrix metalloproteinases in synovial fluids from patients with rheumatoid arthritis and osteoarthritis. *Biol Chem* **391**, 571-579.
6. Gabrijelcic D, Svetic B, Spaic D, Skrk J, Budihna M, Dolenc I, Popovic T, Cotic V & Turk V (1992) Cathepsins B, H and L in human breast carcinoma. *Eur J Clin Chem Clin Biochem* **30**, 69-74.
7. Turk V, Kos J & Turk B (2004) Cysteine cathepsins (proteases)--on the main stage of cancer? *Cancer Cell* **5**, 409-410.
8. Stoka V, Turk B & Turk V (2005) Lysosomal cysteine proteases: structural features and their role in apoptosis. *IUBMB Life* **57**, 347-353.
9. Gocheva V & Joyce JA (2007) Cysteine cathepsins and the cutting edge of cancer invasion. *Cell Cycle* **6**, 60-64.
10. Turk B, Turk D & Salvesen GS (2002) Regulating cysteine protease activity: essential role of protease inhibitors as guardians and regulators. *Curr Pharm Des* **8**, 1623-1637.
11. Musil D, Zucic D, Turk D, Engh RA, Mayr I, Huber R, Popovic T, Turk V, Towatari T, Katunuma N, et al. (1991) The refined 2.15 A X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *EMBO J* **10**, 2321-2330.
12. Illy C, Quraishi O, Wang J, Purisima E, Vernet T & Mort JS (1997) Role of the occluding loop in cathepsin B activity. *J Biol Chem* **272**, 1197-1202.
13. Turk D, Podobnik M, Popovic T, Katunuma N, Bode W, Huber R & Turk V (1995) Crystal structure of cathepsin B inhibited with CA030 at 2.0-A resolution: A basis for the design of specific epoxysuccinyl inhibitors. *Biochemistry* **34**, 4791-4797.
14. Turk D & Guncar G (2003) Lysosomal cysteine proteases (cathepsins): promising drug targets. *Acta Crystallogr D Biol Crystallogr* **59**, 203-213.
15. Turk D, Janjic V, Stern I, Podobnik M, Lamba D, Dahl SW, Lauritzen C, Pedersen J, Turk V & Turk B (2001) 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.
16. Molgaard A, Arnau J, Lauritzen C, Larsen S, Petersen G & Pedersen J (2007) The crystal structure of human dipeptidyl peptidase I (cathepsin C) in complex with the inhibitor Gly-Phe-CHN2. *Biochem J* **401**, 645-650.
17. Paris A, Strukelj B, Pungercar J, Renko M, Dolenc I & Turk V (1995) Molecular cloning and sequence analysis of human preprocathepsin C. *FEBS Lett*

369, 326-330.
18. Guncar G, Podobnik M, Pungercar J, Strukelj B, Turk V & Turk D (1998) Crystal structure of porcine cathepsin H determined at 2.1 A resolution: location of the mini-chain C-terminal carboxyl group defines cathepsin H aminopeptidase function. *Structure* **6**, 51-61.
19. Guncar G, Klemencic I, Turk B, Turk V, Karaoglanovic-Carmona A, Juliano L & Turk D (2000) Crystal structure of cathepsin X: a flip-flop of the ring of His23 allows carboxy-monopeptidase and carboxy-dipeptidase activity of the protease. *Structure* **8**, 305-313.
20. Turk V, Stoka V & Turk D (2008) Cystatins: biochemical and structural properties, and medical relevance. *Front Biosci* **13**, 5406-5420.
21. Podobnik M, Kuhelj R, Turk V & Turk D (1997) Crystal structure of the wild-type human procathepsin B at 2.5 A resolution reveals the native active site of a papain-like cysteine protease zymogen. *J Mol Biol* **271**, 774-788.
22. Cygler M, Sivaraman J, Grochulski P, Coulombe R, Storer AC & Mort JS (1996) Structure of rat procathepsin B: model for inhibition of cysteine protease activity by the proregion. *Structure* **4**, 405-416.
23. Nagler DK, Storer AC, Portaro FC, Carmona E, Juliano L & Menard R (1997) Major increase in endopeptidase activity of human cathepsin B upon removal of occluding loop contacts. *Biochemistry* **36**, 12608-12615.
24. Redzynia I, Ljunggren A, Abrahamson M, Mort JS, Krupa JC, Jaskolski M & Bujacz G (2008) Displacement of the occluding loop by the parasite protein, chagasin, results in efficient inhibition of human cathepsin B. *J Biol Chem* **283**, 22815-22825.
25. Bode W, Engh R, Musil D, Thiele U, Huber R, Karshikov A, Brzin J, Kos J & Turk V (1988) The 2.0 A X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. *EMBO J* **7**, 2593-2599.
26. Stubbs MT, Laber B, Bode W, Huber R, Jerala R, Lenarcic B & Turk V (1990) The refined 2.4 A 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.
27. 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.
28. 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.
29. Redzynia I, Ljunggren A, Bujacz A, Abrahamson M, Jaskolski M & Bujacz G (2009) Crystal structure of the parasite inhibitor chagasin in complex with papain allows identification of structural requirements for broad reactivity and specificity determinants for target proteases. *FEBS J* **276**, 793-806.
30. Ljunggren A, Redzynia I, Alvarez-Fernandez M, Abrahamson M, Mort JS, Krupa JC, Jaskolski M & Bujacz G (2007) Crystal structure of the parasite protease inhibitor chagasin in complex with a host target cysteine protease. *J Mol Biol* **371**, 137-153.
31. Renko M, Sabotic J, Mihelic M, Brzin J, Kos J & Turk D Versatile loops in mycocypins inhibit three protease families. *J Biol Chem* **285**, 308-316.

32. 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.
33. Turk B, Ritonja A, Bjork I, Stoka V, Dolenc I & Turk V (1995) Identification of bovine stefin A, a novel protein inhibitor of cysteine proteinases. *FEBS Lett* **360**, 101-105.
34. Estrada S, Pavlova A & Bjork 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.
35. Kuhelj R, Dolinar M, Pungercar J & Turk V (1995) The preparation of catalytically active human cathepsin B from its precursor expressed in Escherichia coli in the form of inclusion bodies. *Eur J Biochem* **229**, 533-539.
36. Jerala R, Kroon-Zitko L & Turk V (1994) Improved expression and evaluation of polyethyleneimine precipitation in isolation of recombinant cysteine proteinase inhibitor stefin B. *Protein Expr Purif* **5**, 65-69.
37. Govada L & Chayen NE (2009) Crystallization by Controlled Evaporation Leading to High Resolution Crystals of the C1 Domain of Cardiac Myosin Binding Protein-C (cMyBP-C). *Crystal Growth & Design* **9**, 1729-1732.
38. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307-326.
39. Navaza J & Saludjian P. (1997) AMoRe: An automated molecular replacement program package. *Methods Enzymol* **276**: 581-594.
40. Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* **33**, 491-497.
41. Kantardjieff KA & Rupp B (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12**, 1865-1871.
42. Murshudov GN, Vagin AA & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255.
43. Turk D (1992) Weiterentwicklung eines Programms fuer Molekuelgraphik und Elektrondichte-Manipulation and Seine Anwendung auf Verschiedene Protein-Strukturaufklerungen. *PhD thesis, Technische Universitaet Muenchen, Germany*.

TABLES

Table 1: Average distances between CA atoms of the stefins and catalytic residues of cysteine proteases.

	d (Å)
Papain – stefin B	23.93
Cathepsin H – stefin A	23.36±0.23
Cathepsin B – stefin A	23.34±0.15

Table 2: Data collection and refinement statistics for the complex of cathepsin B with stefin A. Numbers in parentheses are for the highest resolution shell. No intensity cutoffs were applied.

Data collection	
PDB ID	3K9M
Space group	P1
Cell dimesions	
a, b, c (Å)	62.0, 31.0, 70.9
α, β, γ (°)	90.0, 104.5, 90.0
Resolution (Å)	68.6 – 2.51
R_{merge} (%)	8.4 (20.6)
I/σI	9.5 (2.6)
Completeness (%)	92.1 (66.7)
Redundancy	2.6 (2.2)
Refinement	
Resolution	40.5 – 2.61
No. of reflections (work/free)	24360 / 713
R_{work}/R_{free}	19.8 / 25.0
B factor (Å 2)	42.0
No. of atoms	
Protein	5454
Water	127
r.m.s. deviation	
Bond lenghts (Å)	0.013
Bond angles (°)	1.71

FIGURE LEGENDS

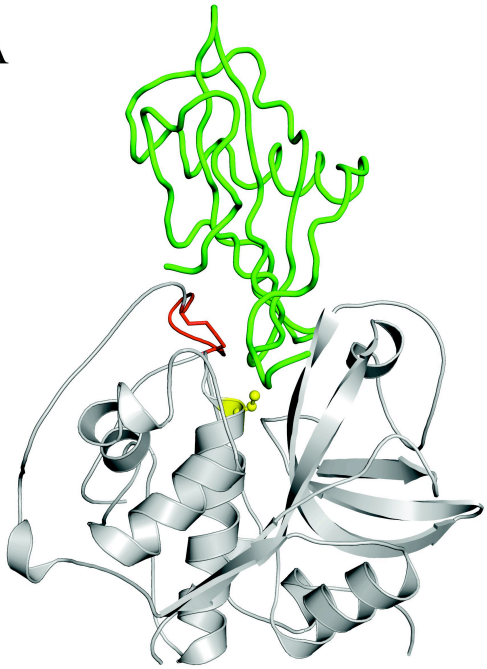
Figure 1: Structure of the cathepsin B – stefin A complex. A) A view along the active site cleft. B) A view perpendicular to the active site cleft. Cathepsin B is shown in gray and stefin A in green. The catalytic cysteine is shown in yellow. The wedge-shaped structure of stefin A fills the active site cleft along the whole length and displaces the occluding loop (the "lasso" is shown in red).

Figure 2: The extent of the occluding loop displacement in the unliganded and liganded structures. The occluding loop (red) is shown in on the surface of the papain-like part of the structure (gray). A) Unliganded cathepsin B (PDB code 1HUC) [11]. B) propeptide in dark blue (PDB code 3PBH) [21]. C) A complex with stefin A, with stefin A in green. D) A complex with chagasin (shown in cyan) (PDB code 3CBJ) [24].

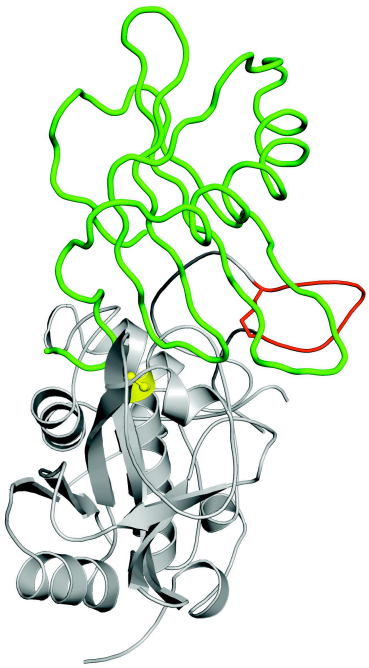
Figure 3: The extent of the occluding loop displacement – superimposed. The papain-like part of cathepsin B is shown as a gray surface with the catalytic cysteine part shown in yellow, while the S1, S1' and S2' binding sites are shown in green and cyan. The occluding loops from various cathepsin B structures (proenzyme, complex with stefin A, complex with chagasin) are shown in dark blue, red, and cyan, respectively. The occluding loop residues, H110 and H111, from the naked cathepsin B, are shown in orange. Spheres represent the position of CA atom of N113, to indicate the extent of movement of the occluding loop.

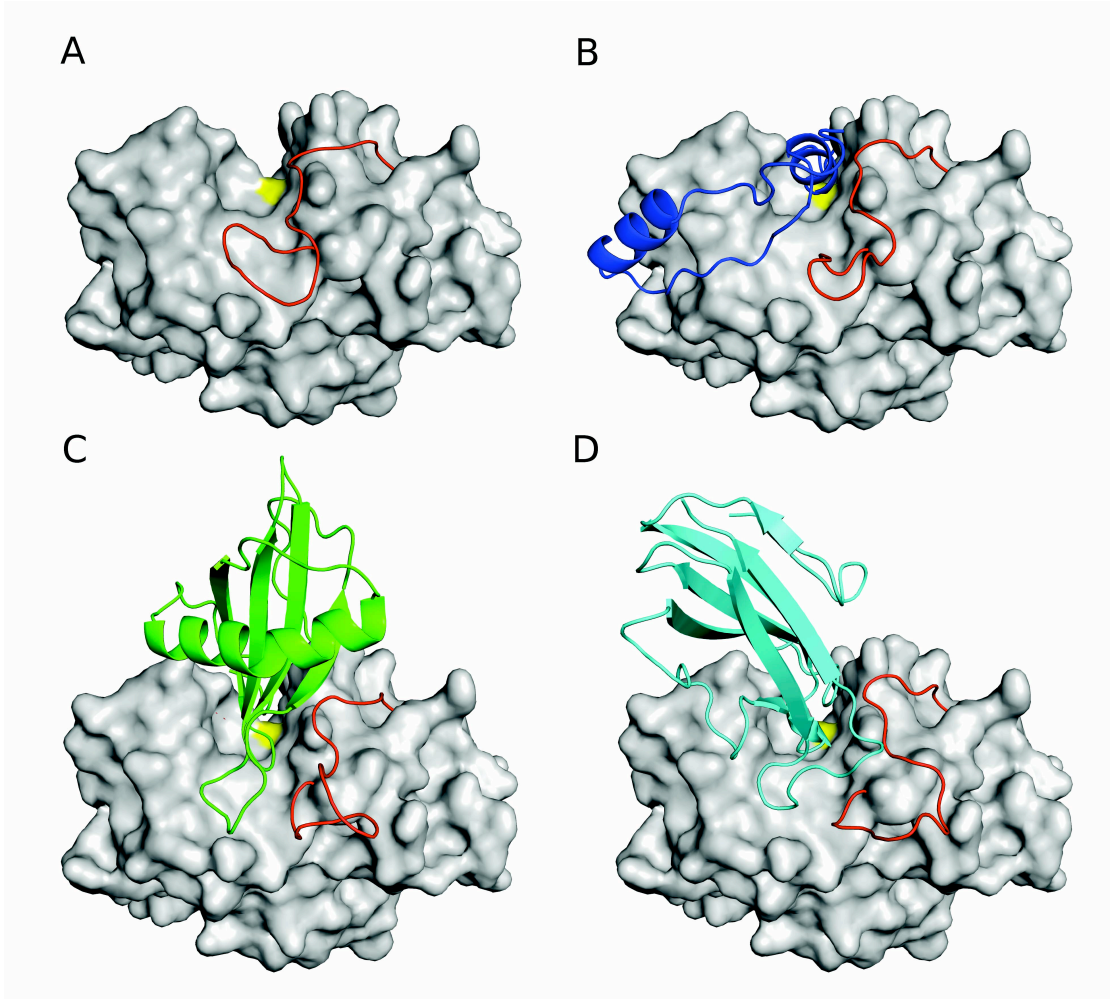
Figure 4: Flexibility of stefin structures. Papain surface (PDB code 1STF) [26] is shown in gray with the part of the reactive cysteine residue shown in yellow. Four structures of stefin A from the complex with cathepsin H are shown in cyan (PDB code 1NB3) [27]. The two structures of stefin A from the complex with cathepsin B are shown in red. The stefin B structure from the complex with papain is shown in green. Six stefin A molecules were moved onto the scaffold of papain using transformation parameters obtained from the superimpositions of their enzymatic partners on the papain structure.

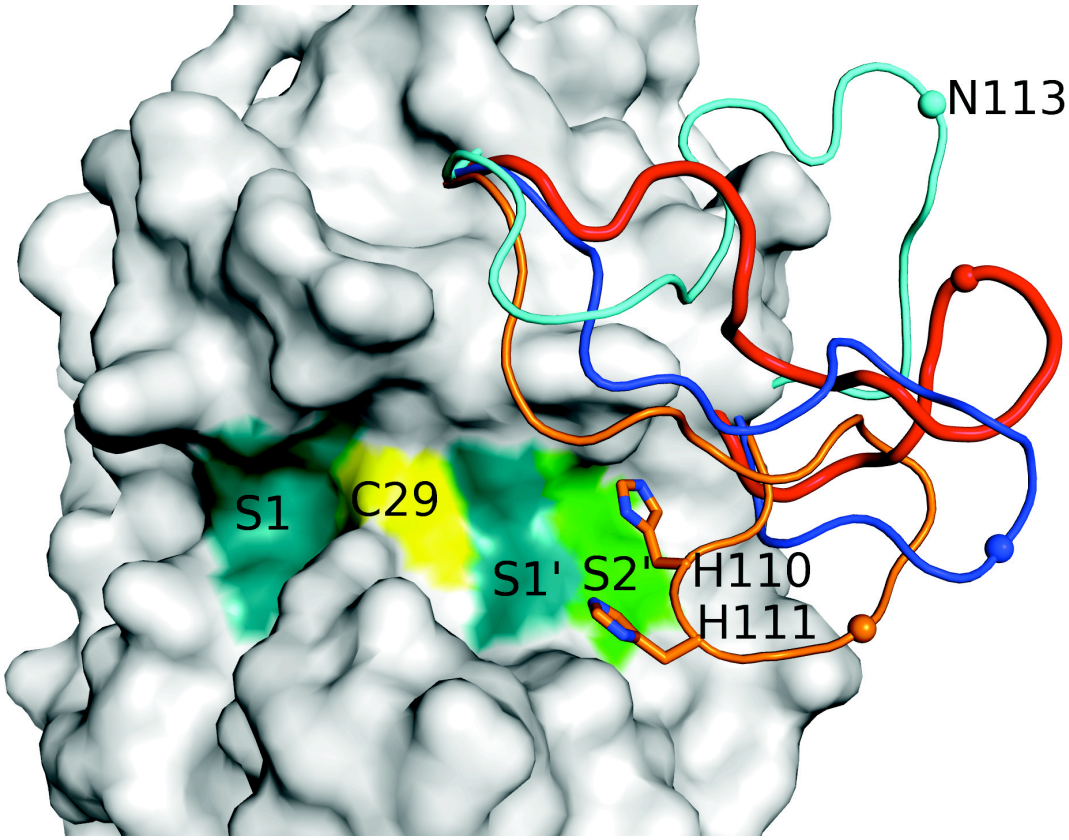
A



B




A
B
C
D


N113
S1
C29
S1'
S2'
H110
H111