STRUCTURAL INSIGHT IN THE SUBSTRATE SPECIFICITY OF CATHEPSINS

Piotr Sosnowski

Doctoral Dissertation
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia

Supervisor: Prof. Dr. Dušan Turk, Ljubljana, Slovenia

Evaluation Board:
Prof. DDr. Boris Turk, Chair, Jožef Stefan Institute, Ljubljana, Slovenia
Prof. Dr. Nataša Poklar Ulrih, Member, Biotechinical Faculty, University of Ljubljana, Ljubljana, Slovenia
Prof. Dr. Walter Keller, Member, Institut für Molekulare Biowissenschaften, KatL Franzens Universität, Graz, Austria

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL



Piotr Sosnowski

STRUCTURAL INSIGHT IN THE SUBSTRATE SPECIFICITY OF CATHEPSINS

Doctoral Dissertation

STRUKTURNI VPOGLED V SUBSTRATNO SPECIFIČNOST KATEPSINOV

Doktorska disertacija

Supervisor: Prof. Dr. Dušan Turk

Ljubljana, Slovenia, March 2016

I dedicate this work to my loving family . . .

Acknowledgments

CIPKEBIP and ARRS for the funding.
Evaluation board: Prof. DDr. Boris Turk, Prof. Dr. Nataša Poklar Ulrih and Prof. Dr. Walter Keller for critical and thorough reading of the thesis and their corrections.
My supervisor Dušan Turk for support during the work and thesis writing.
Miha Renko for support and introduction to protein structure determination.
Saša Usenik for broad help in work preparation.
Ajda Taler-Verčič for thorough thesis reading and valuable corrections.
Livija Tušar for thorough thesis reading and valuable corrections.
Gregor Pretnar for help in laboratory.
All other co-workers at the Jožef Stefan Insitute.
My family for support and patience.

Abstract

Lysosomal cysteine cathepsins are proteolytic enzymes that take part in a number of biological processes such as endosomal protein degradation, antigen processing, bone remodeling. Besides their roles under normal physiological processes they are also involved in a number of pathological processes such as cancer, arthritis and osteoporosis. In these processes the cysteine cathepsins perform limited and unlimited proteolysis. The current challenge in understanding of physiological roles of cysteine cathepsins is to gain insight in the interactions with their natural protein substrates and their selection. Since not much is known about endosomal cysteine cathepsins, we first made an attempt to provide insight in the substrate binding of cathepsins V and L. By structural approaches we tried to shed light on digestion data of natural samples analyzed by proteomics. This work describes how we established an experimental platform to study interactions between cysteine cathepsins and their natural substrates. We used presumably catalytically inactive cathepsins and selected peptides. So far we were able to express, crystallize and determine structures of catalytically inactive cathepsins L and V (C25S mutation) as models for binding peptides within the active site.

The crystal structures of complexes of cathepsin V C25S with peptides and cathepsin L C25S show peptide residues along the whole active site cleft from the S3 to S4' subsite, confirming now also structurally that there are indeed seven residues interacting with the active site of a papain-like enzyme as shown biochemically by the pioneering work of Schechter and Berger back in 1968. This information was obtained by combining several structures of complexes with peptides bound into the active site. The peptides were designed on the basis of analysis of proteomic data of cathepsin V degradation pattern of cell lysate. The structures showed peptides bound either in the non-primed or in the primed side of the active site, but not in both. The structures revealed the binding of peptides in the S3' and S4' subsites, whereas the subsequent residues are not in the contact with the enzyme and are exposed to the solvent. The complexes of cathepsins V C25S mutants which bind residues in the S3, S2 and S1 subsites indicated the existence of low enzymatic activity of cathepsins. The final evidence for the remaining catalytic activity came from the structure of cathepsin L C25S, which was cleaved during the two-week crystallization process. The crystal structure caught the cathepsin L protease – cathepsin L substrate complex “in the act” after cleavage and before the final separation of the two molecules. This analysis showed that the cysteine mutants only reduced the catalytic activity of the enzymes to a level that is not detectable by fluorogenic substrates. We suggest that given enough time all C25S and C25A active site mutants can hydrolyze peptide bonds and that for the same reason also the complex of human cathepsin L C25A mutant published by Adams-Cioaba et al. (2011) contained residues only in the non-primed region.

In addition, we solved structures of cathepsin K in complex with different nitrile inhibitors which show specific interactions at the S subsites of cathepsin K.

Povzetek

Lizosomski cisteinski katepsini so proteolitski encimi, ki sodelujejo v vrsti bioloških procesov, kot so endosomska proteinska razgradnja, antigensko procesiranje in remodeliranje kostnega tkiva. Poleg njihovih vlog pri normalnih fizioloških procesih so katepsini vključeni še v vrsto patoloških procesov, kot so rak, artritis in osteoporoza. V teh procesih cisteinski katepsini izvajajo omejeno in neomejeno razgradnjo proteinov. Trenutni izziv pri razumevanju fizioloških vlog katepsinov je pridobiti vpogled v interakcije med njihovimi naravnimi proteinskimi substrati in njihovo izbiro. Ker je izbira substratov manj znana za endopeptidaze, smo najprej poskusili pridobiti vpogled v vezavo substrata katepsinov V in L. S pristopi strukturne analize smo poskusili osvetliti podatke razgradnje naravnih vzorcev s proteomskimi pristopi. To delo opisuje, na kakšen način smo vzpostavili eksperimentalno platformo za študij interakcij med cisteinskimi katepsini in njihovimi naravnimi substrati. Pri tem smo uporabili predvidoma neaktivne katepsine in nabor peptidov. Do sedaj nam je uspelo izraziti, kristalizirati in določiti strukture predvidoma katalitsko neaktivnih katepsinov L in V (C25S mutacijo) kot modelov za vezavo peptidov v aktivno mesto.

Strukture kompleksov katepsina V C25S s peptidi in katepsina L C25S v kristalu kažejo peptidne preostanke vezane vzdolž aktivnega mesta od mest S3 do S4'. Ti kompleksi potrjujejo zdaj tudi s strukturno analizo, da se v aktivno mesto papainu podobnih encimov zares veže sedem aminokislinskih preostankov - kot sta v svojem pionirskem delu pokazala Schechter in Berger leta 1968. To informacijo smo pridobili iz struktur več različnih kompleksov peptidov, vezanih v aktivno mesto katepsina V. Peptidi so bili pripravljeni na osnovi analize proteomskih podatkov razgradnje celičnega lizata. Strukture so pokazale, da se peptidi vežejo v vezavna mesta s črtico ali pa v mesta brez črtice, ne pa v obe. Strukture so povečale obseg vpogleda v vezavo peptidov, in sicer v vezavna mesta S3' in S4'. Nadaljnji preostanki substratnega modela niso več v stiku z molekulo encima. Kompleksi katepsina V C25S mutant, ki so vezali peptide v vezavna mesta S3, S2 in S1, so nakazali obstoj nizke encimatske aktivnosti katepsinov. Dokončni dokaz za preostalo katalitsko aktivnost pa smo pridobili s strukturo katepsina L C25S, v kateri so se molekule katepsina L med dvotedenskim kristalizacijskim postopkom razcepile. Struktura v kristalu je ujela kompleks med encimatsko aktivno molekulo katepsina L in drugo molekulo, v kateri katepsin L igra vlogo substrata med dejanjem samim, takoj po cepitvi in pred separacijo obeh molekul. Ta analiza je pokazala, da je katalitska aktivnost cisteinskih mutant zgolj zmanjšana do nivoja, ko je s fluorogenimi substrati ne moremo več zaznati. Kadar pa je na voljo dovolj časa, sklepamo, da bodo mutante C25S in C25A lahko hidrolizirale peptidne vezi in da so zato tudi v kompleksu katepsina C25A, ki so ga objavili Adams-Cioaba in sod. (2011), prisotni samo preostanki v substratnih mestih brez črtice.

Poleg tega smo rešili več struktur katepsina K v kompleksu z nitrilnimi inhibitorji in tako razkrili specifične interakcije na S vezavnih mestih katepsina K.

Contents

List of Figures **xvii**

List of Tables **xix**

Abbreviations **xxi**

1 Introduction **1**
1.1 Definition of Substrate Binding Sites ... 2
1.2 Cysteine Proteases ... 3
1.2.1 Papain-like proteases ... 3
1.2.1.1 3D structure of papain ... 3
1.2.1.2 Cysteine Cathepsins ... 5
1.2.1.2.1 Physiological relevance of cathepsins L, V and K ... 6
1.2.1.3 Proteomic studies ... 7

2 The Aim of the Work **9**

3 Methods **11**
3.1 Protein Expression ... 11
3.1.1 Recombinant proteins ... 11
3.1.2 Theoretical background ... 13
3.1.3 Protocols for cathepsins expressions ... 14
3.1.3.1 Cathepsins L, L-C25S, V, V-C25S and K expression in erlenmeyer flasks ... 14
3.1.3.2 Cathepsin L expression in bioreactor ... 14
3.1.3.3 Cathepsins expression summary ... 15
3.2 Protein Purification ... 15
3.2.1 Theoretical background ... 15
3.2.1.1 Ni affinity chromatography ... 15
3.2.1.2 Size exclusion chromatography ... 15
3.2.1.3 Ion exchange chromatography ... 15
3.2.2 Protocols for protein purification ... 16
3.2.2.1 First step purification ... 16
3.2.2.2 Second step purification ... 16
3.2.2.2.1 Human cathepsin L-C25S ... 16
3.2.2.2.2 Human cathepsin V-C25S ... 16
3.2.2.2.3 Human cathepsin K ... 17
3.3 Cathepsins Propeptide Removal ... 17
3.3.1 Theoretical part ... 17
3.3.2 Protocols for propeptide removal ... 17
3.3.2.1 Cathepsin L-C25S and cathepsin V-C25S propeptide removal and purification ... 17

3.4 Protein Properties, Differential Scanning Fluorimetry (DSF) 19
3.4.1 Theoretical background 19
3.4.2 Protocols for buffer optimization 20
3.4.2.1 Buffer test with DSF 20
3.5 Crystallization 20
3.5.1 Theoretical background 20
3.5.2 Crystallization protocols 22
3.5.2.1 Cathepsin L-C25S 22
3.5.2.2 Cathepsin V-C25S 22
3.5.2.3 Cathepsin V-C25S peptide soaking 23
3.5.2.4 Cathepsin V-C25S peptide cocrystallization 23
3.5.2.5 Cathepsin K cocrystallization with nitrile inhibitors 23
3.5.2.6 Crystallization and peptide soaking summary 23
3.6 Structure Determination 25
3.6.1 Theoretical background 25
3.6.1.1 X-ray diffraction 25
3.6.1.2 Data collection 27
3.6.1.3 Data processing 28
3.6.1.4 Phase determination 28
3.6.1.4.1 Isomorphous replacement 28
3.6.1.4.2 Anomalous scattering 28
3.6.1.4.3 Direct methods 29
3.6.1.4.4 Molecular replacement 29
3.6.1.5 Density modification 29
3.6.1.6 Model building 29
3.6.1.7 Protein model refinement 30
3.6.1.8 Structure validation 31
3.6.2 Structure determination protocols 31
3.6.2.1 Human cathepsin L-C25S 31
3.6.2.2 Human cathepsin V-C25S 31
3.6.2.3 Cathepsin V-C25S with peptides 31
3.6.2.4 Cathepsin K complex structure determination 31
3.6.3 Data collection and refinement statistics 32

4 Results **37**
4.1 Cathepsin V-C25S Structures 37
4.1.1 Residue binding to cathepsin V-C25S active sites and beyond 44
4.1.1.1 P3 binding to S3 binding site 44
4.1.1.2 P2 binding to S2 binding site 44
4.1.1.3 P1 binding to S1 binding site 44
4.1.1.4 P1' binding to S1' binding site 44
4.1.1.5 P2' binding to S2' binding site 45
4.1.1.6 P3' binding to S3' position 45
4.1.1.7 P4' substrate position in cathepsin V-C25S 46
4.1.1.8 P5' and P6' substrate position in cathepsin V-C25S 47
4.2 Cathepsin L-C25S Complex 47
4.2.1 Residues binding to cathepsin L-C25S active sites and beyond 49
4.3 Cathepsin K with Nitrile Inhibitors 51

5 Discussion **53**
5.1 Evaluation of Experimental Model for Substrate Subsite Studies 53

5.2 Unique Complex of Cathepsin L-C25S....56
5.3 Cathepsin L C25S Mutant Partial Activity....58
5.4 Cathepsin K with Nitrile Inhibitors....59

6 Conclusions **61**

References **63**

Bibliography **75**

7 Materials **77**
A.1 Culture Media....77
A.1.1 Purification buffers for affinity chromatography and activation buffer....78
A.1.2 Other chemicals....79
A.1.3 Laboratory equipment....79
A.1.4 Peptides 80

Appendix B **81**

List of Figures

Figure 1.1: Standard view of cartoon (A) and surface (B) representation of papain. 4
Figure 1.2: Binding sites of cysteine proteases presented on human cathepsin L. 4
Figure 1.3: Subsrate proteomic ice logos of cathepsins L (A) and V (B). 7
Figure 3.1: Scheme presenting main steps for determination of a 3D protein structure. ..11
Figure 3.2: Recombinant protein schemes. 12
Figure 3.3: Sequence alignment of expressed cathepsins. 13
Figure 3.4: CatL-C25S and catV-C25S propeptide removal. 18
Figure 3.5: CatV-C25S propeptide removal time scale screen. 19
Figure 3.6: Presentation of crystallization phase diagram. 21
Figure 3.7: Vapour diffusion crystallization overview. 22
Figure 3.8: Human cathepsin crystal images: 24
Figure 3.9: Human cathepsin K complexes crystal images: 25
Figure 3.10: Representation of several Miller indices h, k, l 25
Figure 3.11: Representation of X-ray diffraction and Bragg's law 26
Figure 3.12: A diffraction pattern of our cathepsin V-C25S viewed in iMOSFLM 27
Figure 4.1: Two asymmetric units of cV-C25S crystals. 38
Figure 4.2: Structural insight in the cathepsin V-C25S complex_1_soak. 40
Figure 4.3: Structural insight to cathepsin V-C25S complex_2_soak. 41
Figure 4.4: Structural insight to cathepsin V-C25S complex_3_soak. 42
Figure 4.5: Structural insight to cathepsin V-C25S complex_4_cocrys. 43
Figure 4.6: Cathepsin V-C25S structural insight to binding sites with peptides. 46
Figure 4.7: Cathepsin V-C25S structural insight to the P4' residue binding 47
Figure 4.8: The crystal packing of cathepsin L-C25S exhibits zig-zag alignment: 49
Figure 4.9: Cleaved peptide binding into the active site cleft 50
Figure 4.10: Auto-processing of human cathepsin L-C25S during crystallization 51
Figure 4.11: Human cathepsin K-inhibitor complex. 52
Figure 5.1: Superimposition of peptides binding into the active site of cathepsin V-C25S54
Figure 5.2: Binding of peptides 55
Figure 5.3: The cathepsin L-C25S cleaved region 58
Figure 5.4: Superimposition of crystal structures of nitrile inhibitors on OLB inhibitor ..60

List of Tables

Table 3.1: Summary of the crystallization conditions.....24
Table 3.2: Data collection and statistics of cathepsins L-C25S and V-C25S complexes. 33
Table 3.3: Data collection and statistics of cathepsin V-C25S complexes. 34
Table 3.4: Data collection and statistics of cathepsins K complexes 35
Table 4.1: The RMSD values of cathepsin V complexes.....39
Table 4.2: Summary of peptide binding to the cathepsin V-C25S subsites. 39
Table 5.1: Data Calculated statistics of determined cathepsin L-C25S 57

Abbreviations

ALAASS	···	peptide N-Ala-Leu-Ala-Ala-Ser-Ser-C
AVAEKQ	···	peptide N-Ala-Val-Ala-Glu-Lys-Gln-C
AVRARL	···	peptide N-Ala-Val-Arg-Ala-Arg-Leu-C
Bromo-benzoyl	···	N-(2-Amino-4-bromobenzoyl)homocycloleucyl-glycinonitrile
C25A	···	mutation of 25th Cys (mature cathepsin numbering) into alanine
C25S	···	mutation of 25th Cys (mature cathepsin numbering) into serine
catL	···	Human Cathepsin L
catV	···	Human Cathepsin V
cDNA	···	Circular DNA
Chloro-benzoyl	···	N-(2-Amino-4-chlorobenzoyl)homocycloleucyl-glycinonitrile
Cocrys.	···	short of “cocrystallization “
Cryo	···	Cryogenic conditions
EDTA	···	2,2',2'',2'''-(Ethane-1,2-diyldinitrilo)tetraacetic acid
Fluoro-benzoyl	···	N-(3-Fluoro-4-aminobenzoyl)homocycloleucyl-glycinonitrile
Fobs	···	Observed reflections amplitudes
Fcalc	···	Reflection amplitudes calculated from the model
HEPES	···	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
LLKVAL	···	peptide N-Leu-Leu-Lys-Val-Ala-Leu-C
MAD	···	Multiple wavelength anomalous dispersion
MIR	···	Multiple isomorphous replacement
ML	···	Maximum likelihood
MPD	···	2-methyl-2,4-pentadiol
MR	···	Molecular replacement
N/A	···	Short of “Not available”
PDB	···	Protein Data Bank
PICS	···	Proteomic identification of protease cleavage sites
RMSD	···	Root-mean-square deviation of atomic positions
SAD	···	Single wavelength anomalous dispersion
SDS-PAGE	···	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TRIS	···	2-Amino-2-hydroxymethyl-propane-1,3-diol

Chapter 1

Introduction

Cells are the elementary structural, functional and biological units of all organisms. They consist of protoplasm enclosed with a membrane where organelles and biomolecules are found [1]. The protoplasm makes an environment for most biological processes, which are driven by biomolecules. Biomolecules differentiate into macromolecules, such as nucleic acids, lipids, other small molecules and the most numbered proteins [2]. Proteins have various functions, such as stimulation of the production of other molecules, transporting, signaling, catalysis and many others [1]. Shortly, transporting proteins move ions, molecules and macromolecules through cell membrane [3], while the signaling proteins are responsible of triggering processes and communication between cells [4]. Proteins capable of catalyzing the chemical reactions are called enzymes. They lower the activation energy of chemical reactions and therefore speed up reactions [5]. Depending on the reaction enzymes are divided into six classes: transferases (transfer a functional group from one molecule to another), ligases (join two molecules with each other), oxidoreductases (transfer electrons, between molecules), isomerases (convert a molecule from one isomer to another), hydrolases (break bonds using water) and lyzases (break bonds without water and oxidation) [6].

The one of the most interesting enzyme groups are proteases (EC 3.4.) [7]. Proteases or proteinases or proteolytic enzymes hydrolize the peptide bonds of other proteins and peptides [8]. They are involved in cleaving of long peptides into short fragments by removing the peptide bond linking amino acid residues. Depending on the cleavage position, we differentiate between exopeptidases, which cleave at the termini of a peptide, and endopeptidases, which cleave the internal peptide bonds. According to the function we recognize two types of proteolysis: limited and unlimited. During limited proteolysis only specific substrate peptide bonds are cleaved, whereas during the unlimited proteolysis substrates are degraded to short peptides or single amino acids continuously.

Based on the residues involved in catalysis we differentiate aspartic, cysteine, glutamic, metallo-, asparagine, serine, threonine and other not classified proteases, where the active site residues are involved in a nucleophilic attack on the carbonyl carbon of the peptide bond [7].

Because proteases can cleave essentially any protein they can reach there are several mechanisms in place which prevent their harmful activity. Proteases are usually synthesized in a zymogen form, which blocks their premature activity, like trypsinogen, chymotrypsinogen or pepsinogen. The zymogen usually consists of an inhibitory segment and an active protein [9]. The zymogen is activated by removal of the inhibitory segment by auto cleavage or by another enzyme. The protein activation occurs when the precursor reaches its destination –for example intracellular component (e.g. lysosome) or extracellular environment (e.g. stomach) [10].

Proteases are found in all organisms and are essential for cell physiology. In human body, there are at least 747 identified proteases [7], involved in many biological functions, from simple food digestion to regulation of complicated cascades. Trypsin, chymotrypsin and pepsin are examples of proteases responsible for protein digestion to short peptides or single amino acids in stomach and intestine. The blood clotting example shows a classical model of proteolytic cascade which requires a number of events that involve activation of different serine proteases zymogens to produce fibril cloths [11]. Yet another example is the process of programmed cell death (apoptosis), which is regulated by granzymes and cysteine proteases caspases. The initiation of the apoptosis removes a damaged cell, for example to prevent further spread of viral infection [11].

In order to understand and be able to control the proteolytic pathways a number of studies are ongoing. For example, the functional studies uncover interactions between proteases and their substrates. Such studies include determination of protease cascade pathways, substrate inhibition or activation, substrate specificity and others. At the other end, the structural studies give insight to the substrate protease binding mechanisms, explain the factors responsible for substrate specificity and provide information on the specific inhibitor design.

Since proteases can perform both limited and unlimited proteolysis, the understanding of the binding and selection mechanism of substrates is important [10]. Proteases performing specific cleavages have their structure modified in a way that only a desired substrate is able to reach their active site, while the less specific proteases have their active site accessible to many different substrates [12]. The active site of proteases consists of active site residues and substrate binding sites [13]. It is the interplay of complementarity of features of a substrate and a protease which makes them partners in proteolysis [14]. The specificity of interactions is thus embedded in both and defines the outcome of their interaction.

1.1 Definition of Substrate Binding Sites

In 1967, Schechter and Berger [13] reported their fundamental work on substrate binding sites of papain. They studied substrate kinetics dependency on the length of a polyalanine chain and found out that the kinetics is influenced by the polypeptide chain length until the number of amino acids residues did not exceed seven and concluded that there are seven substrate binding sites on the papain molecule. Therefore they proposed the nomenclature for the enzyme subsites (S) and the substrate amino acid residues positions (P). Positions and subsites are numbered from the cleavage site in both directions. To further differentiate between the N-terminal and C-terminal sides of the peptide the substrates sites and amino acid positions on the C-terminal side of the cleaved peptide were denoted with "'" and called the primed sites and positions on the N-terminal side of the cleavage were called non-primed. This definition of substrate-enzyme interaction became a standard for assignment of sites of interaction between a polypeptide substrate and a proteolytic enzyme.

Insights into the specificity of later described cysteine proteases were obtained sporadically. The studies done on recombinant cysteine proteinase revealed specificity of the S1 subsite [15]. The studies of cathepsin B revealed specificity of the S2 subsites for the arginine residue [16], while the other studies determined the role of H110 and H111 for S2' subsite [17]. Studies on cathepsin L revealed specificity of the S2 subsite [18]. Other studies described the S1' subsite specificity of all cysteine cathepsins [19]. Described examples refer to cysteine proteases which are the main focus of this thesis. Lately the protease substrate interactions are studied systematically by systematic

studies of interactions between combinatorial libraries [20] of substrates and inhibitors and enzymes as well as proteomic approaches [21].

The proteomic studies provide large databases of natural substrates with different approaches: case by case, libraries [22] and PICS [23]. However, they cannot provide insight into the mechanism of substrate binding and their selection. We hope to complement this by combining these results with structural studies on cysteine cathepsins, members of the papain family of enzymes.

1.2 Cysteine Proteases

Cysteine proteases are the third largest group of hydrolases with 113612 sequences. The group is divided into 14 clans with 84 families according to the MEROPS peptidase data bank [7]. Proteases studied during this work come from the C1 (papain-like proteases) clan. The representative protease of this group is a papain structure, which is a well-studied protease from the fruit of the *Carica papaya* tree [24].

The catalytic activity of these proteases resides on two catalytic residues cysteine and histidine, which are supported with two other residues glutamine and asparagine. The glutamine is a part of the oxyanion hole, which is an electrophilic center that stabilizes the intermediate within the active site. The asparagine is interacting with the catalytic histidine and properly orientates its position in the active site [25].

1.2.1 Papain-like proteases

The family of papain-like proteases is most abundant among the cysteine proteases and their representatives are found in bacteria, archaea, protozoa, fungi, plants animals and viruses. It is one of few families which are found in all organisms [7].

The family is represented by papain and related proteins like chymopapain, caricain, bromelainm ficain and cysteine cathepsins. The proteins are basically small with molecular mass range from 20-35 kDa [26], with the exception of cathepsin C which is a tetrameric enzyme with molecular mass of 200 kDa [27].

The family contains broad range specific endopeptidases such as papain, endopeptidases with narrow specificity such as glycyl endopeptidases, aminopeptidases, peptidyl-dipeptidase, carboxypeptidases and peptidases with both exo- and endopeptidase activity such as cathepsin B and H.

1.2.1.1 3D structure of papain

Structurally cysteine proteases are monomeric with exception of cathepsin C. The standard structure for papain-like cysteine proteases is papain [28]. In agreement with the standard view, the structure of papain consists of two domains, referred to as the left (-L) and the right (-R) (Figure 1.2). The left domain is built from three α-helices, where the right one is a kind of β-barrel with the front strand(s) forming a coiled structure. In between the two domains a V shaped active site is located [28].

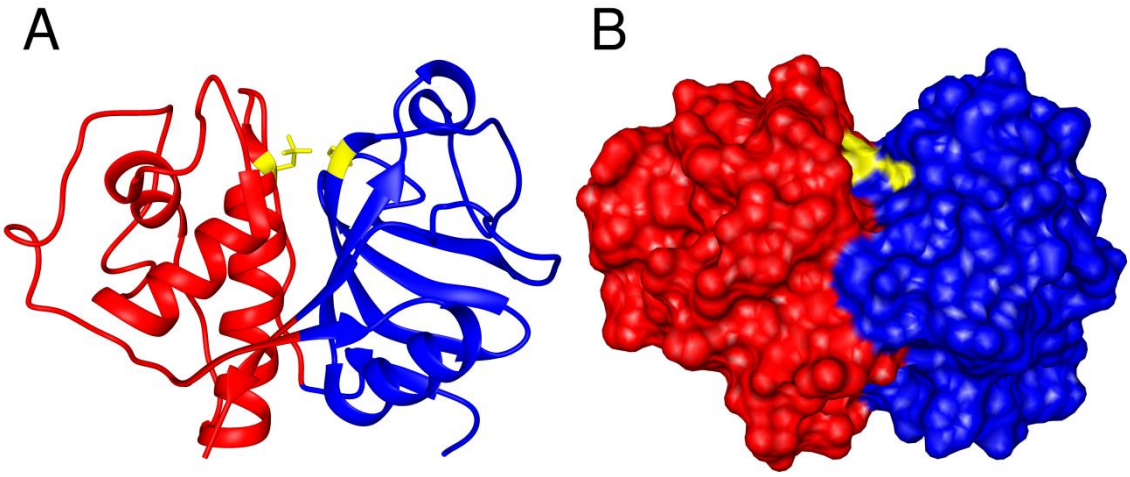


Figure 1.1: Standard view of cartoon (A) and surface (B) representation of papain. -L and -R domains are colored red and blue, respectively. The active site residues are colored yellow and presented as sticks. The picture was prepared using the Open-source PyMOL™ v.1.7.2.1. (The PyMOL Molecular Graphics System, Version 1.7.2 Schrödinger, LLC).

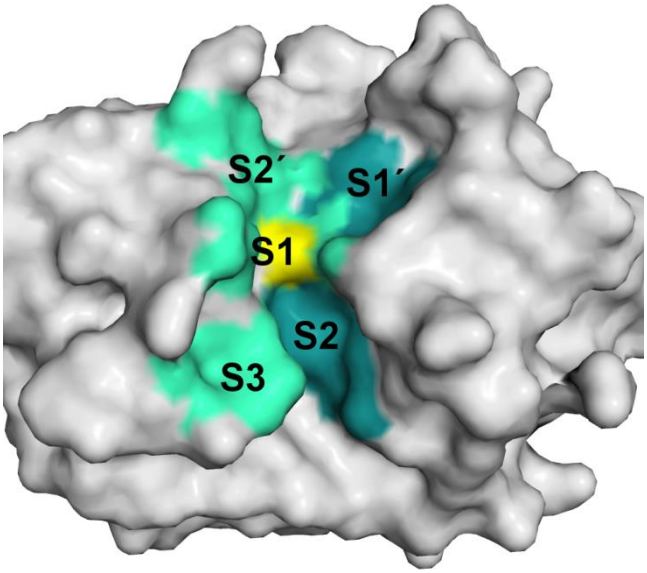


Figure 1.2: Binding sites of cysteine proteases presented on human cathepsin L. The protease surface is colored grey. The S3, S1 and S2' binidng sites are light green. The S2 and S1' binding sites are colored dark teal. The active site cysteine is colored yellow. The picture was prepared using the Open-source PyMOL™ v.1.7.2.1. (The PyMOL Molecular Graphics System, Version 1.7.2 Schrödinger, LLC).

Cysteine cathepsin active site consists of five subsites (S3, S2, S1, S1', S2') presented in Figure 1.2. with the active site residues Cys25 and His163 coming from both domains [29]. In addition, Gln19 and Asn187 supplement the active site and together with the active site residues form a reactive oxyanion hole [25]. The substrate's N-terminal is positioned on the S3, S2, S1 binding sites where the C-terminal positions in the S1' and S2' subsites. The binding sites were deducted from the structures of papain-like enzymes with substrate analogue inhibitors. The S1, S2 and S3 subsites were determined using chloromethyl ketone inhibitor [30]. The S1' and S2' subsites were determined from cathepsin B structure with the epoxysuccinyl inhibitor CA030 [31].

In the 90s, when a sufficient number of protease-inhibitor structures became available, the definition of Schechter and Berger substrate binding sites on the papain-like enzymes was revisited and revised [32]. The superimposed structures of complexes of substrate analogue inhibitors and cathepsins revealed that substrate residues bind along the active site cleft in an extended conformation, with the side chains oriented alternately toward the L- and R-domain. Residues P2, P1 and P1' bind into well-defined binding sites, where the S2 binding site is a deep pocket and the S1 and S1' sites provide a binding surface. Positioning of these residues is governed by interactions, which involve both main-chain as well as side-chain atoms. The P2 main chain interacts with the enzyme through hydrogen bonds with the G66 (cathepsins L numbering), while the side chain fits between two domains through hydrophobic interactions [33]. The P1 residue binds to the S1 site interacting with both active site residues, while the side chain points out to the solvent region [32]. The P1 specificity is probably turned towards long positive side chains, since the cathepsin H was processing substrates with small side chains poorly, as shown by Kirschke, et. al [34]. The P1' residue forms hydrogen bonds with the enzyme active site H163 and Q19 (cathepsins L numbering). The P1' residue shows specificity towards aromatic residues in cathepsins B [19, 35]. In addition the P2' residue seems to bind to a well-defined binding site [36]. The residue showed preferences for hydrophobic residues for cathepsins L, B and papain; for basic residues for papain and for acidic residues for cathepsins B only [37].

For the other subsites there is no strong evidence that they are actually binding sites. Therefore it was suggested that the substrate residue binding regions beyond S2 and S2' should not be called sites but areas [32]. The papain-like proteases thus represent a special class of proteolytic enzymes with the smallest number of substrate binding sites as opposed to chymotrypsin-like serine proteases with six [38] and aspartic proteases with eight binding sites [39]. The specificity studies did not reveal strong preferences for specific type of residues [37].

The crystal structures of exopeptidases cathepsins B [40], C [41], H [42] and X [43] revealed structural determinants responsible for docking of termini of peptides. However, the substrate binding mechanism for endopeptidase is still eluding.

There is only one structure available with a real peptidyl substrate bound to the active site cleft of cysteine proteases [44]. It shows binding of a histone H3 Q19-G33 peptide to the inactive recombinant cathepsin L C25A mutant, where three amino acids (Q19, L20, A21) of the histone H3 19-33 peptide are bound to the active site. In the structure of the complex, Q19, L20, A21 occupy the S3, S2 and S1 binding sites, respectively. The remaining residues were not observed in the structure leaving the primed region of the active site empty.

1.2.1.2 Cysteine Cathepsins

There are eleven human cysteine cathepsins, belonging to the papain family. Among them there are endo (cathepsins F, K, L, O, S, V, W) and exopeptidases (cathepsins B,

C, H, X) [7]. The cysteine cathepsins are endosomal proteases located in a membrane-bound compartment (endosome) and involved in intracellular protein degradation and extracellular biological processes regulation. Studies show their role in activation of MHC class II molecules by degradation of invariant chain and generation of antigenic peptides during the MHC class-II restricted antigen presentation, bone remodeling, keratinocyte differentiation, prohormones activation and development of diseases, like cancer, neural disorders, cardiovascular diseases, obesity, and inflammatory diseases. [32, 45-48]

Cysteine cathepsins are expressed as precursors (in zymogen form), similarly to trypsin and pepsin. The gene of cathepsin consists of coding region for the signaling peptide, propeptide and mature enzyme. The signaling peptide is responsible for direction of the protein to the final destination. The propeptide sterically blocks the activity of cathepsins and is removed when the enzyme reaches its destination. The mature protein is the active form of the enzyme ready for interaction with the substrates. In order to prevent unwanted activation of cathepsins, special conditions need to be achieved. Cysteine cathepsins can be activated by propeptide removal by another enzyme or by autoprocessing in acidic and reducing environment such as present in endosome. In the mature form cysteine cathepsins are stable in slightly acidic pH of lysosomes and are unstable in neutral pH. They are optimally active in slightly acidic, reducing environment of lysosomes with the exception of cathepsins S which was found to be active and stable in both conditions [12].

1.2.1.2.1 Physiological relevance of cathepsins L, V and K

To provide structural insight in interactions of cysteine cathepsins with natural substrates we decide to work with endopeptidases cathepsins L, V and K. Cathepsin L was chosen because it was shown [44] } that it can bind a part of a natural substrate from histone H3 in the non-primed active site. In addition to cathepsin L, cathepsin V was chosen. Cathepsin V shares high sequence homology with 77% amino acids identity with cathepsin L. In addition, cathepsin L and V have 72% and 75% identity towards mouse cathepsin L, respectively [49]. Cathepsin L is expressed ubiquitously while the cathepsin V expression is restricted to thymus and testis [50, 51]. Both human cathepsins were found to play a crucial role during the MHC class II restricted antigen presentation. The localization of cathepsin V suggests its role in controlling the αβ-CLIP complexes similar to mouse cathepsin L, while cathepsin L is considered to take part in various metabolic and cardiovascular disorders [52]. Catalytically active cathepsin L was found in nucleus, where it regulates the cell-cycle progression and histone H3 N-terminal proteolytic processing [53]. Moreover, cathepsins were found extracellular indicating that the environmental pH is not a limiting factor for enzyme activity. Cathepsins L, V and others were found to cleave elastin [45, 48]with different activity rates, where cathepsin V showed highest activity. On the other hand, it was shown that cathepsin L has some collagenase activity [46]. The high similarity, different cell distribution and functions of both cathepsins make them interesting targets for substrate specificity studies.

Besides, we studied cathepsin K inhibitor interactions in collaboration with the Chemical Institute in Slovenia. They targeted cathepsin K because it was found to have a major role in osteoporosis and other bone-related pathologies, diabetes, obesity, cancer, and atherosclerosis, [54-60]. The most advanced drug candidates, such as ONO-5334 and Odanacatib, are currently in clinical trials [61-69]. However, one of the earlier candidates Balicatib was discontinued due to unexpected off-target effects resulting in accumulation of the drug in the lysosomal environment [70]. This prompted Borisek et al [71] to develop N-(“low basic” functionalized benzoyl)-homocycloleucyl-glycinonitriles as cathepsin K inhibitors. Selected candidates were evaluated for their selectivity against

cysteine cathepsins K, S, and L. The most selective inhibitors were crystallized with cathepsin K and determined structures provided the structural insight into the inhibitor binding within the active site.

1.2.1.3 Proteomic studies

Recent information of interactions between cathepsins and their substrates is carried out from functional proteomic studies. The studies address biological functions of unknown proteins by determination of their interactions with possible targets. The study uses mass spectrometers as a tool for the identification of proteins separated by gel electrophoresis or by enzymatic digestion [72, 73]. This approach of studying enzyme targets was further developed and implemented to the functional studies between enzymes and their substrates. Nowadays, proteomics provide large numbers of cleavage sites, which are gathered and combined into proteomic libraries. The libraries allow access to large data bases and analysis of them over different aspects. Each library has its own features which can be used for prediction of the cleavage sites [74]. Another proteomic study outcome is the Protease specificity prediction server (PROSPER), which allows prediction of new substrates and cleavages out of 24 different protease families [75]. On the other hand, the determination of substrate specificity can be carried out by using peptide libraries and Proteomic Identification of protease Cleavage Sites (PICS). Peptide libraries are large databases of peptides from digestion of cell lysates resulting in a random combination of amino acids. Peptide libraries are used in determination of protein-protein interactions, drug design and others [76], while PICS uses the proteome derived from peptide libraries for identification of the amino acid preferences of proteases within the active site [77].

Cysteine cathepsins are able to cleave numerous types of substrates. However, thanks to the proteomics studies we are able to gain insight in their similarity as well as differences. We received proteomic data of different cathepsins cleavage products which were done on full cell lysates, from the Kris Gevaert group. The analysis showed 673 specific cleavages for cathepsin L, 614 specific cleavages for cathepsin V, 769 cleavages for cathepsin L and V only and 1130 cleavages for cathepsin L and V and other cathepsins. Figure 1.3, which shows amino acid preference for substrate positions P15-P14' of cathepsins L and V.

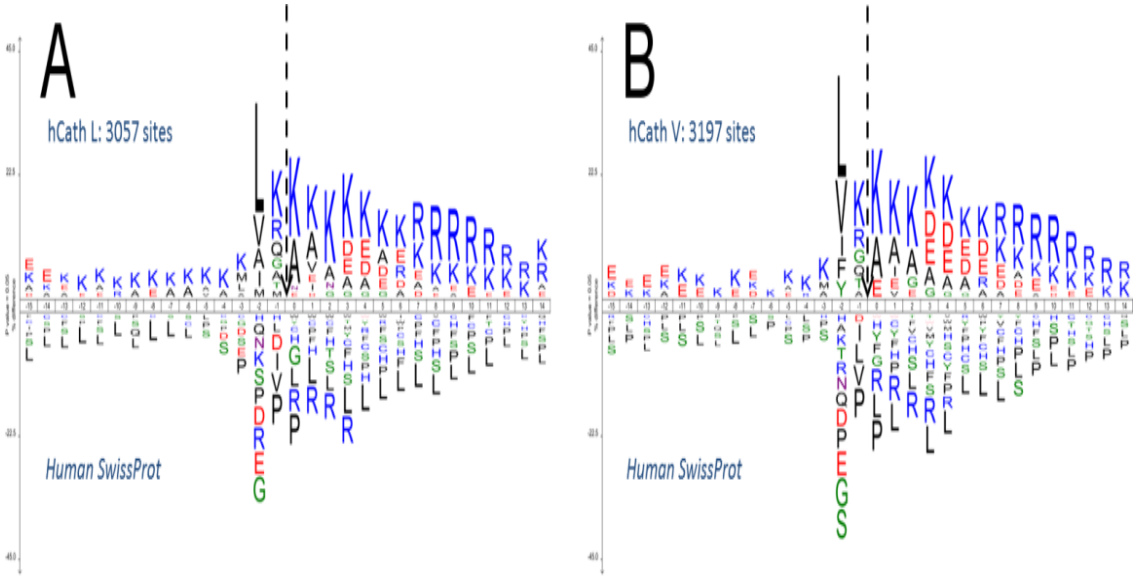


Figure 1.3: Subsrate proteomic ice logos of cathepsins L (A) and V (B). Residues are

colored according to their properties, red for negative charges, blue for positive charges, green for hydrophobic, and black for neutral. Figures were prepared by the Kris Gevaert group. The character size corresponds to the abundance of a certain residue at the position. The positive values refer to the preferable residues, whereas the negative numbers show unpreferable ones.

For the initial interaction insight we selected four synthetic peptides: ALAASS, AVAEKQ, LLKVAL and AVRARL. Peptides were designed to bind through the whole active site (S3-S3') with the scissile bond located between the third and fourth amino acid, with the consideration of the amino acid preferences at certain positions. The P3 residue was selected to be a small hydrophobic or a hydrophilic residue, like leucine and alanine which are additionally well defined at this position. The P2 residue was selected to be a hydrophobic residue like leucine or valine, since both of them are well defined at this position. The P1 residue was selected to have either large or short side chain to additionally address the possible interactions between residues P1 and P2' bound into the S1 and S2' subsites, which are structurally close to each other. Therefore, a set of small-small, small-large, large-small and large-large side chain peptides at P1 and P2' positions were prepared, where we combined lysine and arginine which have large side chains with alanine and serine which have a small side chain. The P1' was selected to have small side chains with well-defined binding to the subsite. The P2' residues were selected so they will correspond to the restrictions we introduced with the P1 residue. Finally, the P3' residue was selected to investigate binding of different residues at the S3' subsite.

Chapter 2

The Aim of the Work

The main aim of this work is to provide a structural insight into the binding interactions of protein substrates of cysteine cathepsins and thereby assist in interpretation of protein degradation data collected with proteomic approaches.

The additional aim is to provide structural insight into the binding interactions of drug candidate nitrile inhibitors in the S subsites of cathepsin K.

Chapter 3

Methods

To peruse the goals described in the Introduction and The Aim of the Work the following experimental steps (Figure 3.1), methods and procedures described below were used.

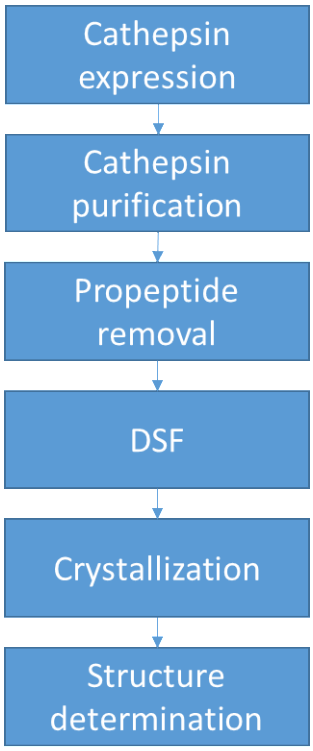


Figure 3.1: Scheme presenting main steps for determination of a 3D protein structure.

3.1 Protein Expression

3.1.1 Recombinant proteins

Procathepsin L and V cDNA were obtained from Deutches Ressourcenzentrum fűr Genomforschung. Two point mutations were introduced for both proteins at positions 123 and 204 (procathepsin numbering). The cysteine at position 123 was mutated into serine to remove enzymatic activity. The threonine at position 204 was mutated into alanine to prevent glycosylation of the protein in *Pichia pastoris* (*P. pastoris*) system.

Recombinant proteins were inserted into pPIC9 plasmids between restriction sites recognized by Not I and Sac I restriction enzymes. The pPIC9 plasmid consists of an α-factor secretion signal responsible for secretion of a protein out from the cell to media. In addition, proteins cDNA were inserted so the His-tag region is located at the N-terminal end of procathepsin. We prepared four recombinant protein constructs (Figure 3.2, Figure 3.3).

The cDNA of human procathepsin K was obtained from Deutches Ressourcenzentrum fűr Genomforschung. cDNA for human procathepsin K was amplified by PCR. PCR products were cloned into vector pPIC9 (Invitrogen). The proteins cDNA were inserted without the His-tag region.

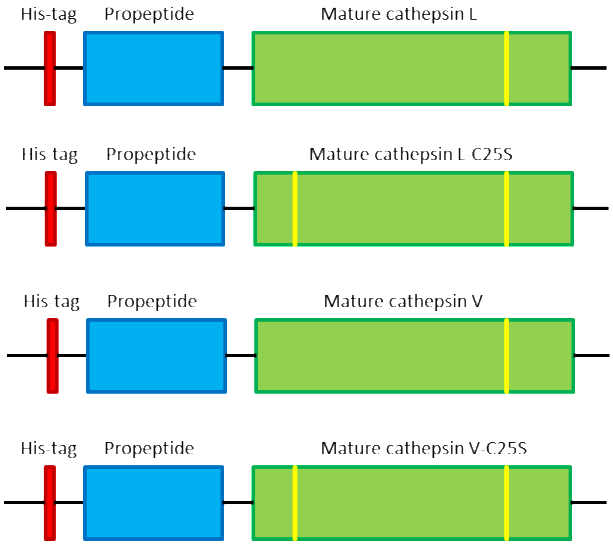


Figure 3.2: Recombinant protein schemes. Parts of proteins presented in a block scheme. The His-tag is colored red. The propeptide is colored blue. The mature protein form is colored green. Introduced point mutations are presented as yellow lines.

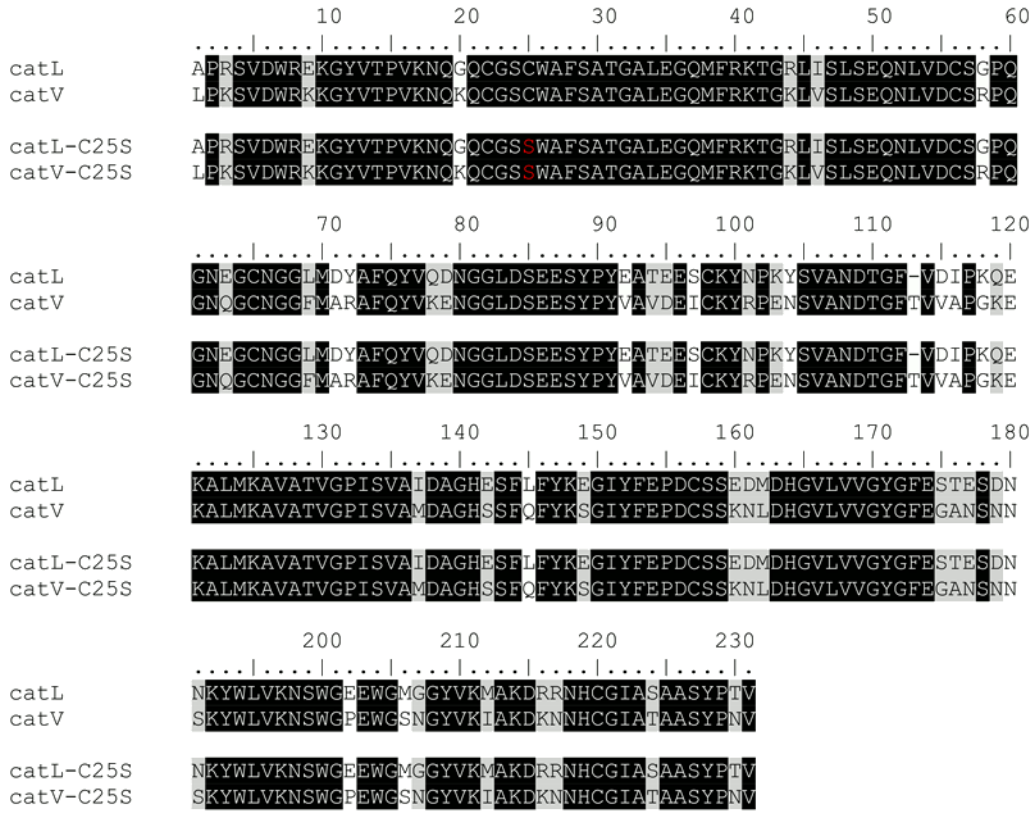


Figure 3.3: Sequence alignment of expressed cathepsins. L, V, L-C25S, V-C25S. The active site mutation C25S is marked red.

3.1.2 Theoretical background

Pichia pastoris (*P. pastoris*) is used frequently as an eukaryotic expression system for the protein production. It has eukaryotic advantages but still remains easy to handle as *Escherichia coli* (*E. coli*) or *Saccharomyces cerevisiae* (*S. cerevisiae*). Protein expression yield is much higher than in mammalian or baculovirus cell cultures, which makes it a useful tool for protein crystallization. *P. pastoris* has high growth rate and grows on a simple inexpensive medium, performs posttranslational modifications, protein folding and others and is suitable for small and large scale expressions.

P. pastoris is methylotrophic yeast, which allows it to metabolize methanol as its source of sole carbon [78]. Its metabolism is based on oxidation of methanol to formaldehyde using molecular oxygen by the enzyme alcohol oxidase. The side effect of this reaction is production of peroxide. That is why to avoid its toxicity, this reaction undergoes in peroxisomes. Since alcohol oxidase has low oxygen affinity, *P. pastoris* compensates with high enzyme production. The enzyme is coded under two genes, AOX1 and AOX2. The AOX1 is the main gene for enzyme production and loss of AOX1 gene results in significant reduction of the enzyme amount in the cells and methanol utilization ratio. Such impaired strains are phenotypically called MutS and grow poorly on methanol media. On the other hand, Mut^+ refers to wild type strain, which grows

very good on methanol media [79]. These two phenotypes are used to evaluate transformants for integration of an implemented gene [80].

3.1.3 Protocols for cathepsins expressions

The expression clones were prepared for the following recombinant proteins: human cathepsin L, L-C25S, V, V-C25S and K. Prepared protein DNA sequences were transferred to the competent *E. coli* cells for amplification. The DNA sequences were extracted using Wizard Plasmid purification KIT and electroporated to the competent *P. pastoris* GS115 cells. The electroporation was done under the following conditions: capacitance 25 µF, resistance 200 Ω, power 1.5 kV. Electroporated cells were spread over MD plates and incubated for 3 days at 30°C. Grown colonies were transferred on new MD and MM plates and used for small scale expression. Protein expression was evaluated with SDS PAGE and the best cells were transferred on new MD plates and stored at 4°C. Each month cells were transferred to a new MD plate to keep them alive. Best colonies were used for large scale expression in erlenmeyer flasks or bioreactor [80].

3.1.3.1 Cathepsins L, L-C25S, V, V-C25S and K expression in erlenmeyer flasks

The colonies from the plate were used for inoculation of 200 ml BMGY media in sterile 2 L erlenmeyer flasks. Cells were left to grow overnight at 30°C and 250 rpm. Afterwards, cell solution was redistributed to ten new erlenmeyer flasks filled with 400 ml BMGY media and left to grow again overnight at 30°C and 250 rpm. Next day, cell solutions were centrifuged for 15 min at 1000 rpm at 4°C to separate supernatant from cells. Gathered pelleted cells were resuspended in same the amount of BMMY media and incubated at 30°C and 250 rpm. For the next four days, 100% methanol was added to 1% final concentration in every flask each day. After four days, solution was centrifuged at 7000 rpm, for 10 min at 4°C. Pellet was discarded and the supernatant containing protein was collected and concentrated to ~200 ml at 4°C and set up for purification [80].

3.1.3.2 Cathepsin L expression in bioreactor

Protein expression in bioreactor was used to acquire a larger amount of protein from the same batch. The expression was followed according to the Invitrogen manual. Firstly, the machine needed to be prepared for expression. Both pH and pO2 sensors were equilibrated and all bioreactor components were autoclaved. Sterile bioreactor was filled with the autoclaved expression media and temperature, agitation, aeration and pH were set to 30°C, 1000 rpm, 0.5 v/vm and pH 6.0, respectively, before cell inoculation. That prepared system was inoculated with previously prepared 100 ml of culture in BMGY media. At the beginning of each phase and at the end of last phase 10 ml of sample was taken for further analysis of cell growth and protein production. The equilibration phase lasted for 18-24 hours until the DO (dissolved oxygen) parameter does not change anymore. The following glycerol feeding phase was started with feed rate of 18.15 ml/h/L of initial fermentation volume. When, all glycerol was used and the DO parameter does not change the methanol feed phase was started. The methanol Fed-Batch phase started with feed rate of 3.6 ml/h/L of initial fermentation volume and was increased after 2 hours to 10.9 ml/h/L of initial fermentation volume. The entire methanol feed phase took around 70 hours. Afterwards, collected media was centrifuged at 7000 rpm, for 10 min at 4°C, cell pellet was discarded and the supernatant was collected and concentrated to ~200 ml at 4°C and set up for purification [80].

3.1.3.3 Cathepsins expression summary

Pro-cathepsins L, L-C25S and V, V-C25S and K were successfully expressed in P. pastoris. The expression yield was ~20 g/L. After purification the pro-cathepsin L and V were frozen at -80°C, while the inactive mutants of pro-cathepsins L-C25S and V-C25S and pro-cathepsin K were moved forward for protein purification and propeptide removal.

3.2 Protein Purification

3.2.1 Theoretical background

Protein purification is a series of steps done to get a homogeneous solution of single protein. It is recommended to use as little different steps, necessary to clean protein from other molecules. Each separation step should exploit one of the differences in protein size, physicochemical properties, binding affinity and biological activity. The most common are affinity and size exclusion chromatography, which were used during this work [10].

3.2.1.1 Ni affinity chromatography

The affinity chromatography is a separation technique which uses reversible interaction between a protein and a specific ligand coupled to chromatography matrix. It is based on biological interactions between the ligand and a target molecule, such as electrostatic, van der Waals' forces, hydrophobic interactions or hydrogen bonding. Such variety of possible interactions together with high selectivity and high capacity for the protein made the technique commonly used for protein purification. There are several types of ligands that can bind to the matrix making the technique useful in many cases.

The method was firstly developed by Roche [81] and uses resins which have bounded Ni^{2+} ions as ligands. The binding compound is called His-tag, which is composed of polyhistidines where the histidine number varies from 6 to 8. Depending on the purification purpose the tag can be attached to either C-terminus or N-terminus of the molecule. The compound utilizes a mechanism as "lock and key" when binding to the ion coupled resins during a run through a column. For release of bounded compound an elution procedure uses high concentration of imidazole which competes with the compound for the position in the resins.

3.2.1.2 Size exclusion chromatography

It uses porous gels for separation of proteins in solution. It was first developed by Lathe and Ruthven in 1955 [82]. The principle uses molecule size for their separation. Depending on the pore size of the matrix smaller molecules have to travel longer paths in order to reach the end of the column, while larger molecules go faster through and molecules of similar mass range leave the column at the same point [10].

3.2.1.3 Ion exchange chromatography

This technique is based on a different net charge of each molecule in certain buffer. Each protein has different isoelectric point (pI), where the molecule carries zero electrical charge, which means equal amount of positive and negative charges. Depending on solutions pH, a compound can have one of the two possible net charges, positive or negative. If the pH is higher than the compounds pI, the compound will have a negative net charge and travel towards a positive charge. Similarly, if the solution pH is smaller

than the compounds pI, it will have a positive net charge and travel towards a negative charge. These properties are used in ion exchange chromatography [83]. The method uses resins that are either positively (Q and DEAE sepharose) or negatively (SP and CM) charged and further either strong (Q and SP sepharose) or weak (DEAE and CM) ion exchangers [84].

Before a separation is started, a column is adjusted with desired pH of a buffer to equilibrate the opposing ions in the matrix. Upon the injection of the sample, charged molecules will compete with buffer ions for the binding place next to the opposed charged matrix resins. The retention in the column depends upon the strength of the molecule charge. For elution rising salt concentration is used where salt ions compete with bounded molecules to the matrix. Molecules with weakest net charge will elute first, followed by those with higher net charge [84].

3.2.2 Protocols for protein purification

The protocols for protein purification in two steps were applied after the expression. First step of purification was used to separate pure fractions of procathepsins (precursors of cathepsins L, V, L-C25S, V-C25S) from other molecules. The second step of protein purification was used for the inactive proteins (cathepsin L-C25S, V-C25S) after the propeptide was removed. Only the second step was used in case of cathepsin KK since the cathepsin K construct did not have the His-tag end.

3.2.2.1 First step purification

The concentrated crude procathepsin solution after the expression in P. pastoris was applied on AKTA Xpress system. The protein was automatically applied on the IMAC HisTrap HP 5 ml column. The protein was washed with AKTA washing buffer and eluted with AKTA elution buffer. Directly after that the protein was applied on HiPrep 26/60 Sephacryl S-200 HR column and washed with AKTA Xpress size exclusion chromatography buffer. Pure procathepsin fraction was automatically collected, concentrated to approximately 1 mg/ml and stored at -80°C or prepared directly for propeptide removal.

3.2.2.2 Second step purification

3.2.2.2.1 Human cathepsin L-C25S

Mature protein was dialyzed against 50 mM sodium acetate, pH 5.5 and applied to previously equilibrated SP sepharose column. The protein was eluted with gradient from 0 M to 2 M NaCl. Pure fractions of human apo-cathepsin L-C25S were collected at the beginning of gradient. Joined fractions of pure protein were dialyzed against 20 mM sodium acetate, pH 4.5, 100 mM NaCl, 1 mM DTT, 5% glycerol, concentrated to 44 mg/ml and stored at -20°C.

3.2.2.2.2 Human cathepsin V-C25S

Mature protein was dialyzed against 50 mM sodium acetate, pH 5.0 and applied to previously equilibrated SP sepharose column. The protein was eluted with gradient from 0 M to 0.5 M NaCl. Pure fractions of human apo-cathepsin V-C25S were collected near 0.3 M NaCl. Joined fractions of pure protein were dialyzed against 20 mM sodium acetate, pH 4.5, 100 mM NaCl, 1 mM DTT, 5% glycerol, concentrated to 40 mg/ml and stored at -20°C.

3.2.2.2.3 Human cathepsin K

Activated cathepsin K was dialyzed to buffer_2 (50 mM sodium acetate, pH 5.5) and blocked reversibly with 10-fold molar excess of S-methylmethanethiosulfonate (MMTS). The prepared protein solution was applied to SP sepharose. Cat K was eluted using gradient from 0 to 2 M NaCl in buffer_2. Fractions containing pure Cat K were collected and dialyzed again against buffer_2. Afterward, pure cat K was concentrated to a final concentration of 36 mg/mL.

3.3 Cathepsins Propeptide Removal

3.3.1 Theoretical part

Cathepsins are expressed in a zymogen form which consists of signal peptide, propeptide and mature protein. The zymogen form is an inactive form of cathepsins, where propeptide is blocking the active site of cathepsin. In cells cathepsins are activated in two ways, either with autoactivation or with other proteases, like pepsin or cathepsin D. The autoactivation of a cathepsin undergoes in acidic pH usually in lysosomes. The way how cathepsins are activated depends on a biological process in which they are involved [85].

3.3.2 Protocols for propeptide removal

Since the studied cysteine cathepsin mutants were designed to be inactive, the autocatalytic process of protein activation was not even considered. Therefore, the propeptide removal with addition of other enzyme was the only option. We avoided accumulation of other enzymes with the C25S mutated cathepsins (no catalytical activity) by adding low amount of an active enzyme needed for propeptide removal within one day of incubation. Activation of procathepsins with catalytical activity was done automatically by addition of a suitable buffer.

Pro-cathepsins were dialyzed against the activation buffer (100 mM sodium acetate, 1 mM EDTA, 5 mM DTT, pH 5.0 [86]). Active cathepsins L and V were auto activated for 4 hours at 37°C and used immediately. The mutant cathepsins L-C25S and V-C25S were mixed with previously activated cathepsin L and V in different ratios. The experiments were performed at three different temperatures: 4°C, 23°C and 37°C. Cathepsin L-C25S was additionally incubated with pepsin. Mixed cathepsins were incubated for 48 hours and time point fractions were taken for analysis.

The medium from cathepsin K expression was dialyzed to 50 mM sodium acetate, pH 5.5, 2.5 mM EDTA, 2.5 mM DTT buffer concentrated to 350 mL. Procathepsin K was activated by adding pepsin to final concentration of 0.03 mg/mL and reducing buffer pH to 4.0. The activation took 4 h at 37°C. Activated cathepsin K was purified as described in 3.2.2.2.3.

3.3.2.1 Cathepsin L-C25S and cathepsin V-C25S propeptide removal and purification

Inactive mutants of human procathepsin L-C25S and V-C25S were tested for propeptide removal. We decided to test different temperatures and different concentration of an active enzyme added to a procathepsin sample. Propeptide removal results are presented in Figures 3.4 and 3.5.

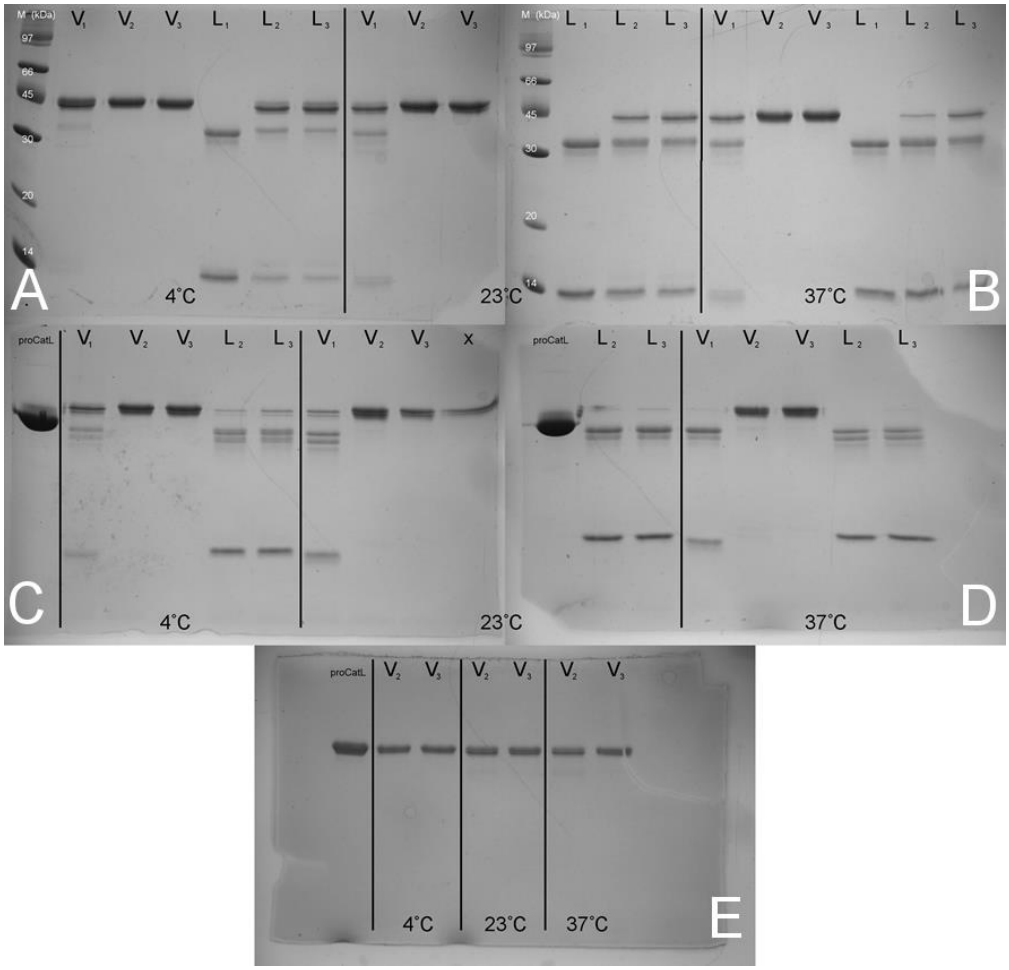


Figure 3.4: CatL-C25S and catV-C25S propeptide removal. The procathepsin L-C25S mixed with the active cathepsin L (L_n) or V (V_n) with the following active/inactive ratios for propeptide removal: V_1 = 0.64%, V_2 = 0.016%, V_3 = 0.012%, L_1 = 0.7%, L_2 = 0.02%, L_3 = 0.014%. The reaction time was as follows: A), B) 1 hour; C), D) 8 hours; E) 20 hours.

The propeptide is cleaved with addition of active cathepsin L in all cases, while active cathepsin V cleaves the propeptide only at the highest active/inactive protein ratio. Three main bands are visible: 43 kDa (which corresponds to procathepsin form), 32 kDa (which corresponds to mature cathepsin form) and 12 kDa (which corresponds to the propeptide). The propeptide is successfully cleaved when only two bands are visible: 32 kDa and 12 kDa. The real masses of procathepsin (36 kDa) and mature enzyme (24 kDa) are different from the ones the gel, because of the travel time of the protein during the electrophoresis. From these results we have decided to remove propeptide from our samples at 22°C using active human cathepsin L.

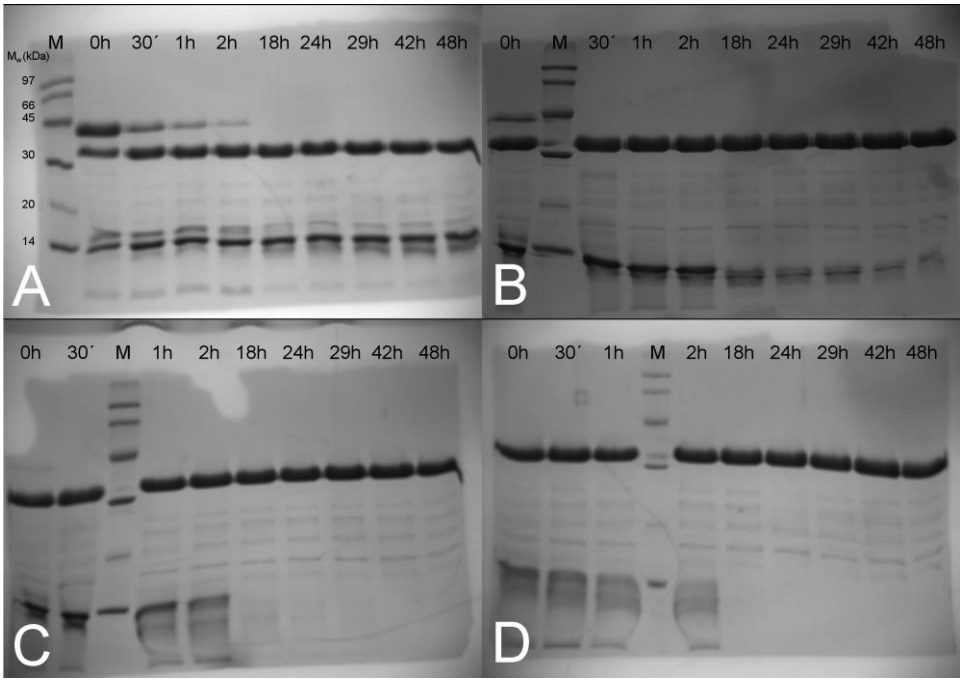


Figure 3.5: CatV-C25S propeptide removal time scale screen. The active/inactive cathepsin ratio for propeptide removal was A) 0.55% B) 2.75%C) 5.5% D) 11% at 23°C.

Afterwards, we optimized the amount of active human cathepsin L which will cleave the whole peptide and degrade it within 24 hours. We tested four different active/inactive protein ratios and the result was achieved with the 5.5% active/inactive ratio.

Cathepsin K was excluded from the propeptide removal condition screen, since it was not the purpose of the work.

3.4 Protein Properties, Differential Scanning Fluorimetry (DSF)

To increase the probability of successful protein crystallization, knowledge of protein properties should be known from different techniques, such as DSF [87], circular dichroism (CD). CD can be used to estimate the protein secondary structure, which might be useful when dealing with protein mutants, since each mutation can distort the protein structure [88]. DSF is used to determine the best conditions, where the protein is most stable. That information helps to find the best conditions according to protein stability and therefore increases the chance that the protein will form crystals. There are several techniques available. We were using differential scanning fluorimetry (DSF) to determine the best stabilization buffer for targeted proteins.

3.4.1 Theoretical background

One of important aspects for protein crystallization is protein stability in solution. This property defines how protein behaves in different conditions. To determine best stability conditions a technique called differential scanning fluorimetry (DSF) was developed and

applied [89, 90]. It is a quick temperature based assay for determination of melting temperatures (temperature when half of the protein undergoes denaturation process). The method is using the SYPRO Orange dye property of increasing fluorescence when binding to exposed hydrophobic patches of the protein (SYPRO Orange, which binds to hydrophobic parts and fluoresces at 590 nm after excitation at 490 nm). While the temperature is rising, the protein starts to unfold, therefore opening new binding places for the dye. When the protein unfolds, the increase in fluorescence is not observed anymore. The temperature and fluorescence are monitored using qPCR machine (real time PCR). The melting temperatures of a protein in different conditions are compared and those which increase the melting point are considered to present the best stabilizing buffer for the protein.

3.4.2 Protocols for buffer optimization

After several unsuccessful crystallization trials with cathepsin L and V C25S mutants using the 20 mM NaAcetate, pH 5.0 buffer, we decided to improve the protein stability by modifying the buffer. The human cathepsin V-C25S was dialyzed against 20 mM NaAcetete, pH 5.0 and concentrated to 1 mg/ml and a buffer screen of 96 different conditions was prepared. For the experiment 10 µL of protein was mixed with 10 µL of 100x diluted SyproOrange and 20 µL of each screen buffer. Prepared plate with solutions was interested into Real Time PCR machine and heated up steadily from 25°C to 95°C by 1°C steps. The results were analyzed using "Transform Agilent Mx3005p data for DSF Analysis v3.0" and "DSF Analysis v3.0.2" spreadsheets, both freely available on internet [87].

We used the same screen of stabilizing conditions for human cathepsin L-C25S and V-C25S based on their sequence similarity.

We did not perform the DSF for cathepsin K, because the protein was successfully crystallized with inhibitors from the 50 mM sodium acetate, pH 5.5 buffer.

3.4.2.1 Buffer test with DSF

The DSF analysis showed small temperature changes for both cathepsins L-C25S and V-C25S. The largest changes were noticeable in buffer pH, sodium chloride, DTT and glycerol concentrations and were not higher than +2°C each. We decided to combine all positive temperature changes and finished up with the following buffer: 20 mM NaAcetate, pH 4.5, 100 mM NaCl, 1 mM DTT and 10% glycerol. All buffer compounds increased protein stability for both cathepsins so we expected successful protein crystallization using those initial protein conditions.

3.5 Crystallization

3.5.1 Theoretical background

Protein crystals, needed for structure determination, are formed during protein crystallization process, which consists of two main events called nucleation and crystal growth [91].

Nucleation is driven by the suppression of the activation energy barrier of nuclei forming in a buffer supersaturated with a protein. When the activation energy will be reached, the protein clusters will eventually grow into crystals. In order to start crystal growth, the nuclei need to reach a critical size for and be stable so they do not redissolve. The critical nuclei size depends on the operating conditions like temperature, buffer

parameters, supersaturation, etc. When the stable state is reached, protein molecules arrange together defining the crystal structure [92].

When nuclei are formed, crystal growth takes place. During the supersaturation state, crystal growth and formations of the nuclei take place simultaneously. Depending on the conditions one of the events might be predominating resulting in different crystal sizes and shapes. When the supersaturation is exhausted, the protein-liquid system reaches equilibrium and both events are stopped, meaning that the crystallization is finished. The described process is presented in Figure 3.6 on a phase diagram, where protein concentration is marked on one of the axis. In addition, the adjusted parameters (temperature, pH, ionic, strength, etc.) are marked and have an influence on the crystallization process [92].

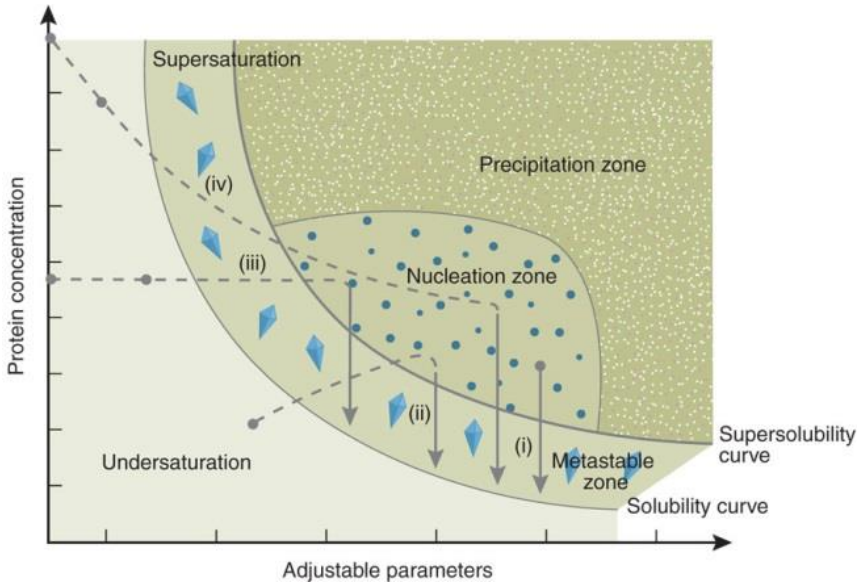


Figure 3.6: Presentation of crystallization phase diagram. Nuclei are forming in the nucleation zone, crystals are growing in the metastable zone until the solution is undersaturated. Grey arrows represent succesful crystallization [93].

Two main crystallization methods are used. Their mechanism is based on the vapour diffusion. They are known under the term of sitting or hanging drop [94, 95]. Both methods are based on the same principle. They are using two volumes: a small drop and more than 100 bigger reservoir solution. The drops are a mix of purified protein and reservoir solution and the reservoir solution is a mixture of a buffer, salt and a precipitant at higher concentrations. Both volumes are sealed together to allow condition equilibration. The solution concentrations in the drop are lower than the concentrations in the reservoir, therefore in order to achieve the equilibrium between both volumes, the water needs to evaporate from the drop. The water evaporation is the driving factor of the vapour diffusion, since the reservoir volume is much higher than the drop; the evaporated water from the drop barely changes the concentrations in the reservoir. When the stable conditions are reached, the concentrations in the drop and the reservoir are the same [96]. When water is leaving the drop, the reagents increase in relative supersaturation moving towards the nucleation zone. When the nucleation zone is reached, the nuclei start to form and if they are stable, the crystals start to grow.

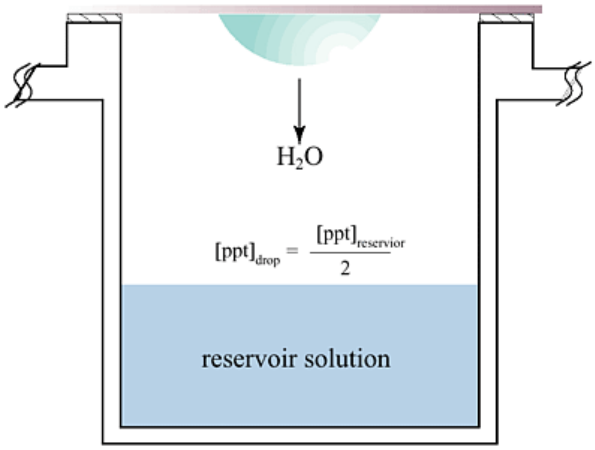


Figure 3.7: Vapour diffusion crystallization overview. System contains a protein drop and a reservoir solution sealed under vapour equilibration [97].

Cocrystallization [98] and ligand soaking [99] are two of the most common methods of obtaining protein-ligand complexes crystals. Both of them are at based on vapour diffusion. The cocrystallization method is usually used for compounds that are insoluble or when the protein aggregates easily. During cocrystallization a ligand is mixed together with a protein, incubated together for some time and applied for crystallization. While the soaking method uses already grown crystals of a protein. The crystals are transferred to another drop containing the same crystallization conditions and a ligand. Crystals are soaked so long that the ligand binds to the protein. Soaking time should not be too long, since a binding ligand might destroy a crystal pattern [100].

3.5.2 Crystallization protocols

The sitting drop technique was used for all crystallization trials. Initial trials were done on 96-well plates using commercial QIAGEN screens. The drop consisted of 0.2 µL of a protein and 0.2 µL of a reservoir solution against 70 µL of a reservoir. The screens were incubated at 8°C and imaged. For the optimization of achieved crystals 24-well plates with 4 sitting drops were used. The optimization drops consisted of 0.2 µL of a protein and 0.2 µL of a reservoir solution against 300 µL of a reservoir.

3.5.2.1 Cathepsin L-C25S

The human apo-cathepsin L-C25S was concentrated to 44 mg/ml and dialyzed against 20 mM NaAcetate, pH 4.5, 10 mM NaCl, 1 mM DTT, 5% glycerol. First crystals grew within one month in 0.1 M HEPES/NaOH, pH 7.5, 2.0 M ammonium sulfate. Optimized crystals grew within one month in the 0.107 M HEPES/NaOH, pH 7.5, 2.14 M ammonium sulfate. Harvested crystals were transferred to cryo solution containing 20% glycerol in reservoir buffer and flash-cooled in liquid nitrogen prior to the diffraction data collection.

3.5.2.2 Cathepsin V-C25S

The human apo-cathepsin V-C25S was concentrated to 40 mg/ml and dialyzed against 20 mM NaAcetate, pH 4.5, 10 mM NaCl, 1 mM DTT, 5% glycerol. First crystals grew from

the precipitate within 3 weeks in 0.1 M TRIS/HCl, pH 8.0, 65% MPD. Optimized crystals grew within one month in the 0.06 M TRIS/HCl, pH 8.0, 77% MPD. Since the buffer itself contained cryo conditions, harvested crystals were directly flash-cooled in liquid nitrogen prior to the diffraction data collection.

3.5.2.3 Cathepsin V-C25S peptide soaking

All four designed peptides: ALAASS, AVAEKQ, LLKVAL, AVRARL were dissolved in water and mixed with the optimized 0.06 M TRIS/HCl, pH 8.0, 77% MPD. The final concentration of a peptide was 6.34 mM. The final molar ratio between protein and peptide during soaking was 1:4.5. Human cathepsin V-C25S crystals were harvested and transferred to a new drop containing a peptide. Soaking took from 1 to 24 hours. After this time, protein crystals were harvested and flash cooled prior to the diffraction data collection.

3.5.2.4 Cathepsin V-C25S peptide cocrystallization

All four designed peptides: ALAASS, AVAEKQ, LLKVAL, AVRARL were dissolved in water to final concentration 37.5 mM and added to the 0.06 M TRIS/HCl, pH 8.0, 77% MPD where the total peptide concentration was 6.34 mM. The molar ratio between protein and peptide in the drop was 1:4.5. Crystals of human cathepsin V-C25S and peptide LLKVAL grew in one month. Protein solutions with peptides AVAEKQ, ALAASS, AVRARL did not crystallize under those conditions. Crystals were harvested and flash cooled in liquid nitrogen prior to the diffraction data collection.

3.5.2.5 Cathepsin K cocrystallization with nitrile inhibitors

For crystallization trials, active cathepsin K blocked with MMTS was reactivated by addition of DTT to the 1 mM final concentration and incubation for 10 min at 37°C. Nitrile inhibitors N-(2-Amino-4-chlorobenzoyl)homocycloleucyl-glycinonitrile, N-(2-Amino-4-bromobenzoyl)homocycloleucyl-glycinonitrile and N-(3-Fluoro-4-aminobenzoyl)homocycloleucyl-glycinonitrile were added to reach 2 mM final concentration and incubated for 30 min at room temperature. Final complexes of enzyme and inhibitor were concentrated to 10-15 mg/mL and centrifuged at 20000 rpm for 30 min. Sitting drops of 0.1 µL protein solution and 0.1 µL reservoir solution against 70 µL of reservoir solution were set at 22°C using commercial screens from QIAGEN and Hampton Research. Crystals grew in 1 week. Crystals of Cat K with N-(2-Amino-4-chlorobenzoyl)homocycloleucyl-glycinonitrile grew in 0.05 M HEPES, pH 7.5, 0.2 M KCl, 35% pentaerythritol propoxylate (5/4 PO/OH) solution. Crystals of Cat K with N-(2-Amino-4-bromobenzoyl)homocycloleucyl-glycinonitrile grew in 0.1 M NaOAc, pH 4.5, 0.2 M Li2SO4, 30% PEG 8000 solution. Crystals of Cat K with N-(3-Fluoro-4-aminobenzoyl)homocycloleucyl-glycinonitrile grew in 0.1 M NaOAc, pH 5.0, 2.4 M (NH4)2SO4. All crystals were transferred for 5 s into cryo buffer, reservoir solution supplemented with 20% glycerol, and flash-cooled in liquid nitrogen.

3.5.2.6 Crystallization and peptide soaking summary

Human cathepsin V-C25S and human cathepsin L-C25S were successfully crystallized. Crystals grew within one month for both cathepsins (Figure 3.8). Cathepsin V-C25S crystals were formed out from a precipitate, while the cathepsin L-C25S crystals from a clear drop. Cathepsins crystals had a shape of a tetrahedron developed from the 'striped' faces of an octahedron. Table 4.1 summarizes the conditions of the crystal growth.

Harvested crystals were soaked in a peptide, containing the crystallization solution for 5 min, 15 min, 30 min, 1 hour, 2 hours, 4 hours, 8 hours and 24 hours and flash cooled.

Harvested crystals of cathepsin K complex with N-(2-Amino-4-chlorobenzoyl)homocycloleucyl-glycinonitrile, N-(2-Amino-4-bromobenzoyl)homocycloleucyl-glycinonitrile and N-(3-Fluoro-4-aminobenzoyl)homocycloleucyl-glycinonitrile inhibitors (Figure 3.9) were flash cooled and tested for diffraction. Table 3.1 summarizes crystallization conditions.

Table 3.1: Summary of the crystallization conditions.

Protein	Crystallization conditions	Crystal growth time	Max. size of the crystal
catV-C25S	0.06 M TRIS/HCl pH 8.0 77% MPD	3-4 weeks	~ 300 mm^3
catL-C25S	0.107 M HEPES/NaOH pH 7.5 2.14 M Ammonium sulfate	3-4 weeks	~ 100 mm^3
catK with N-(2-Amino-4-chlorobenzoyl)homocycloleucyl-glycinonitrile	35% (5/4 PO/OH) 0.2 M KCl 0.05 M HEPES/NaOH, pH7.5	1 week	~ 100 mm^3
catK with N-(2-Amino-4-bromobenzoyl)homocycloleucyl-glycinonitrile	30% PEG 8000 0.2 M Lithium sulfate 0.1 M NaAcetate, pH 4.5	1 week	~ 150 mm^3
catK with N-(3-Fluoro-4-aminobenzoyl)homocycloleucyl-glycinonitrile	2.4 M Ammonium sulfate 0.1 M Tri-sodium citrate/HCl, pH 5.0	1 week	~ 150 mm^3

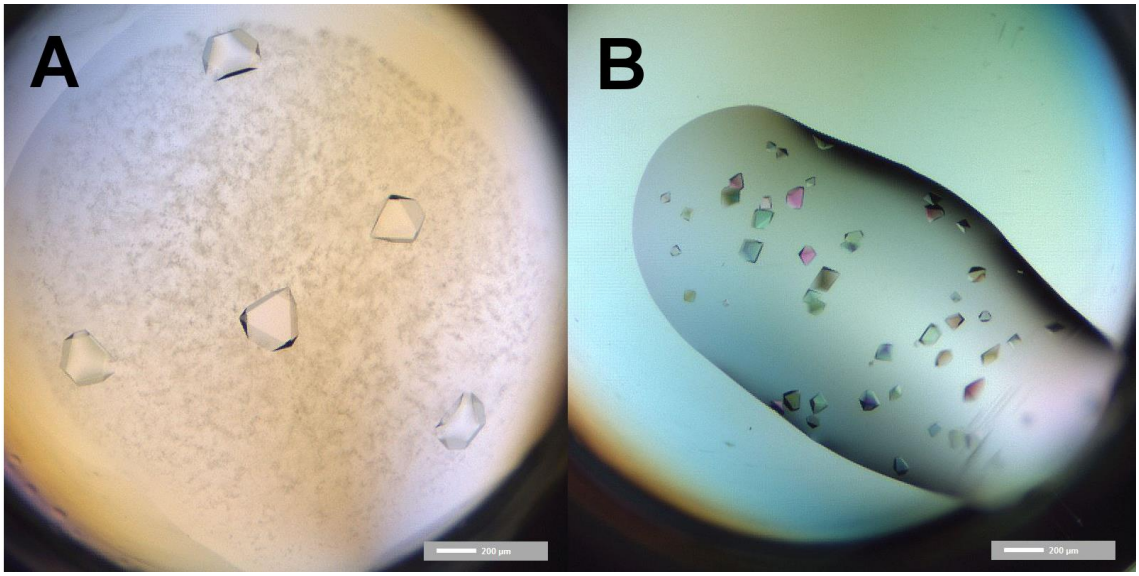


Figure 3.8: Human cathepsin crystal images: A) cathepsin V-C25S crystals, the visible

drop shade comes from the precipitated protein, B) cathepsin L-C25S.

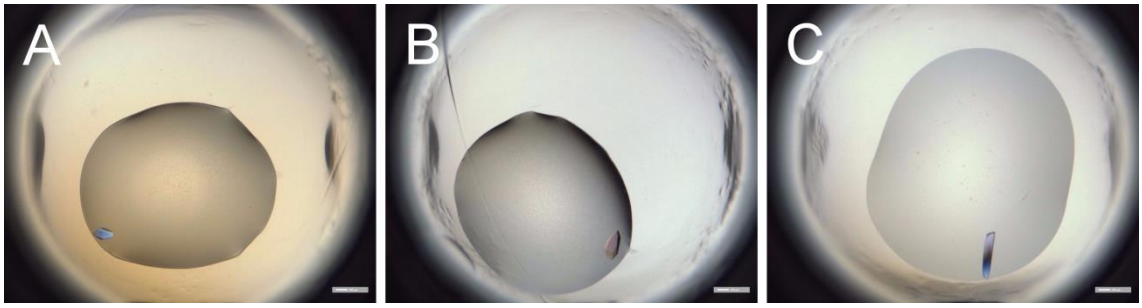


Figure 3.9: Human cathepsin K complexes crystal images: A) with N-(2-Amino-4-chlorobenzoyl)homocycloleucyl-glycinonitrile, B) with N-(2-Amino-4-bromobenzoyl)homocycloleucyl-glycinonitrile and C) with N-(3-Fluoro-4-aminobenzoyl)homocycloleucyl-glycinonitrile. Crystals are visible in the lower part of a drop.

3.6 Structure Determination

3.6.1 Theoretical background

3.6.1.1 X-ray diffraction

X-ray diffraction of crystals is a technique which reveals the atomic structure within the crystal. Since a diffraction pattern from a single molecule is too weak to be measured, crystals as 3-dimensional periodic lattice are used. Crystals are composed out of many molecules which are ordered three dimensional, therefore increasing the output signal, giving rise to both constructive and destructive interference effects in the diffraction pattern.

For the description of crystal X-ray diffraction the Bragg model is used. In this model X-rays when hitting an atom are diffracted with the same frequency, throughout evenly spaced (d) planes identified as Miller indices (h, k, l) [101].

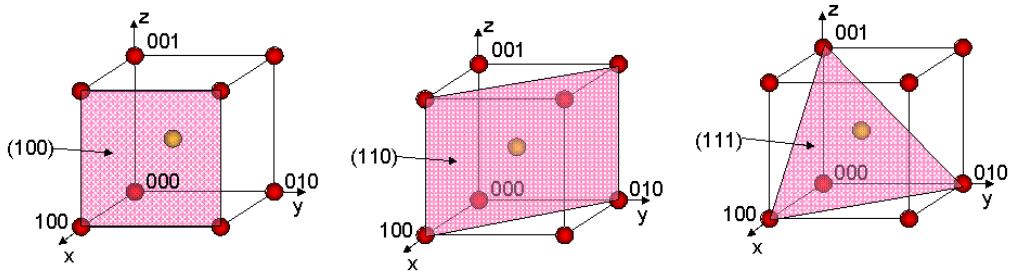
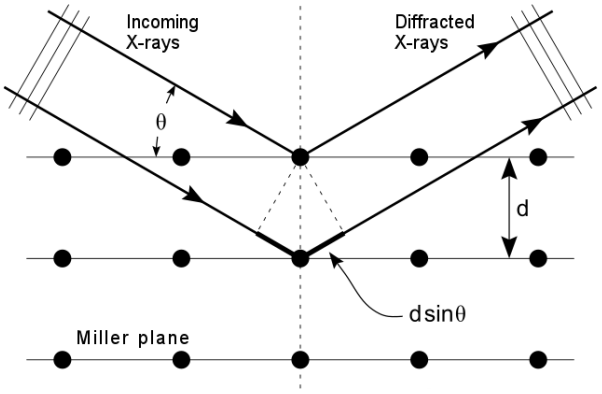


Figure 3.10: Representation of several Miller indices h, k, l [102].

The Bragg law is based on a model in which X-rays are scattered the same way from each plane. In this case, X-rays scattered from adjacent planes will combine constructively and 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. Therefore, X-

rays fulfilling the described rules will result in constructive interference (increasing the signal), when the other x-rays will combine destructively (nullifying the signal) [103].



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

Figure 3.11: Representation of X-ray diffraction and Bragg's law [104].

Combined x-ray diffractions are a result of electrons colliding with the crystal. The out coming x-ray waves can add up in phase (constructively) or out of the phase (destructively). The resulting diffraction is collected by detectors as shown in a diffraction pattern of a crystal (Figure 3.12). The pattern is composed of series of discrete spots, known as reflections. Each reflection contains information about all atoms positioned in the crystal structure and the visible peaks represent integrated intensities of each x-ray diffraction reflections (amplitude). Such picture has encoded information about the crystal symmetry, unit cell (dimension and angles) and the crystal orientation and is also related to the scattering object by amathematical operation known as a Fourier transform (Eq. 3.1) [105, 106].

$$F(s) = \int \rho(\boldsymbol{r}) \exp(2\pi i \boldsymbol{s} \times \boldsymbol{r})\, d\boldsymbol{r} \tag{3.1}$$

Fourier transforms can be inverted. Therefore the Fourier transform of the diffraction pattern can be used to reveal the structure of a scattering object (protein in a crystal). In order to transform the diffraction pattern into the electron density of a model the amplitude of diffracted x-rays and their relative phases need to be known. During the x-ray measurement the amplitudes are measured, while the phases need to be calculated other way. This situation is referred to phase problem, where amplitudes are known and the phases are not [105, 106].

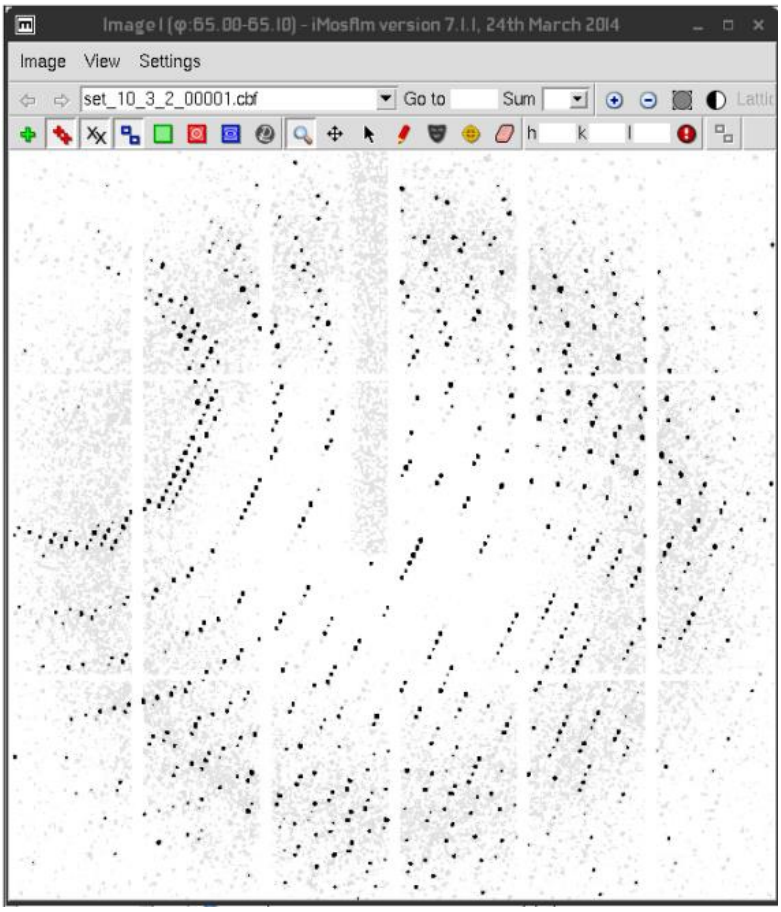


Figure 3.12: A diffraction pattern of our cathepsin V-C25S viewed in iMOSFLM [107].

3.6.1.2 Data collection

Data collection is the last experimental step before the computational structure determination. Its result set up for the outcome of the following computational steps. Data collection result in a series of diffraction images of a protein crystal at a defined wavelength (λ). The purpose of data collection is to measure the structure factor amplitudes and ascertain the crystal symmetry, unit cell parameters, the resolution limit and the crystal orientation. The structure factor amplitudes describe positions of each atom in the crystal structure, which fulfil the Bragg's law. The unit cell parameters describe symmetry molecules in the crystal. The resolution limit defines how well it is possible to invert the x-ray diffraction to model electron density using Fourier transformation. The orientation of the crystal sets up the angles at which the x-rays will be shooting on the crystal resulting in the set of diffraction images. When decided on the purpose of the data collection (native data set collection, isomorphous replacement, anomalous diffraction, molecular replacement) the collection strategy is calculated which will maximize the resolution and completeness of the data set. During the data collection the crystal is rotated through a small angle, usually 1° and the x-ray diffraction patterns are recorded [103, 106].

3.6.1.3 Data processing

Collected diffraction patterns are processed to extract the relative intensities of the x-rays. Several programs are used for this process: Mosflm [107], CCP4 [108], XDS [109] and HKL2000 [110]. The programs use Fourier transformation, where vectors projected along the same space axis direction will result in transformation as a large peak, while the vectors projected on different (random) directions will have different peak size and will not be a peak in the transform. The collected images contain a list of diffraction intensities, which contain information about atom position in the crystal structure. In order to determine the unit cell from atom structure factors, structure factors are mapped back to reciprocal space and the unit cell geometries are determined. Then the unique Miller index is assigned, the intensity value is measured and the cell constants values and symmetry are assigned [103, 106].

3.6.1.4 Phase determination

X-ray diffraction is a result from the interaction of x-rays (waves) with electrons. Reflections on the diffraction pattern (structure factor of electron density in the reciprocal space) correspond to a wave consisting of amplitude and a phase. The waves can be described by a function with an amplitude and phase. The amplitude is calculated by taking the square root of the intensity, while the phase is lost. Structure factors can be represented as vectors. The vectors add together when they travel the same distance differing by exactly an integral number of wavelengths. The scattering vectors fulfil the Bragg's law (Figure 3.11). Therefore, vectors manifest the reciprocal relationship between real distances in the crystal and the dimensions of the diffraction pattern [103, 111].

X-ray crystallography uses several methods to solve the phase problem, by perturbing the x-ray scattering in a predictable way (isomorphous replacement, anomalous scattering) or by guessing the phases (molecular replacement, direct methods).

3.6.1.4.1 Isomorphous replacement

The method uses crystal soaking with heavy atom solution and compares the diffraction pattern of this derivative crystal to the diffraction pattern of the native crystal. The heavy atoms in the structure contribute enormously to the diffraction which makes their contribution easy to detect and the difference of the two diffraction patterns can be taken. Knowing the positions of heavy atoms, which can be determined by direct methods, it is possible to calculate the heavy atom contribution to each structure factor [111, 112].

Two approaches can be taken SIR (single isomorphous replacement) and MIR (multiple isomorphous replacement). In both methods the amplitudes of reflection data sets are measured for the native and the derivative crystal. The difference is used for the estimation of the heavy-atom structure factor amplitude, which is used to determine the heavy-atom positions using Patterson methods. When the parameters of a heavy atom are located, the calculation of the corresponding phases becomes possible [111, 112].

3.6.1.4.2 Anomalous scattering

The complexity of the phase problem is reduced by determining the positions of few heavy atoms with direct methods, similar to isomorphous replacement. To get the information about the positions within one single crystal anomalous scattering (or dispersion) of heavy atoms can be used. If the wavelength of the X-ray corresponds to a transition between different electron shells of the atom, there will be a modification of the phase. This anomalous scattering results in effects in the diffraction pattern that can

be distinguished from the rest (e.g. unequal amplitudes of the Fridel pairs). The atoms do not need to be as large as for isomorphous replacement and typically proteins are used where methionine is replaced by selenomethionine. Generally data is collected at the peak (maximum of heavy atom fluorescence scan, f") for single anomalous dispersion (SAD) and additionally at inflection (half-way down the peak absorption of heavy atom fluorescence scan) and remote (minimum of heavy atom fluorescence scan, f') for multiple anomalous dispersion (MAD). The total scattering factor of an atom is a combination of different contributions. The anomalous scattering factors (f', f") change with the wavelength. At high and low wavelength these are negligible but with a tunable wavelength they can be optimized to have a maximal anomalous signal [113-115].

3.6.1.4.3 Direct methods

Direct methods use Patterson maps, where all phases are zero. A Patterson map can be constructed by taking all vectors between atoms from molecule and plotting them from the origins of the unit cell. This means that the amplitudes of each reflection can be interpreted as vectors between atoms. Such approach allows thinking of a molecule that explains the map. The limitations of the method allow solving the phase problem only for small molecules up to 100 atoms [111].

3.6.1.4.4 Molecular replacement

The method uses a homology model for solving the phase problem. For the method to be successful, a minimal sequence identity of 40% is required to calculate initial phases of the structure factors. The method uses the Patterson map which is the result of Fourier transform of measured data only (amplitudes). In Patterson maps all phases correspond to where peaks correspond to a distance between individual atoms. The MR uses the Patterson map of a model protein and by rotation and translation determines the initial phases of studied structure. During the phase determination the Patterson methods are usually followed by translation of the properly oriented model in the new unit cell [101, 111, 116].

3.6.1.5 Density modification

The density modification procedures are used to improve phases by modification of electron density maps obtained during the solution of the phase problem. Traditional techniques include solvent flattening, histogram matching and non-crystallographic symmetry (NCS) density map averaging. Solvent flattening is based on the understanding that the regions occupied by macromolecule have higher and more variable electron density maps than the solvent region, which is during the procedure flattened. Subsequently the defied density is transformed iteration through reciprocal space. Therefore, a set of phases leading to a flat map in the solvent regions will be correct. The new resulting map contains fewer variables within the solvent regions [117]. The histogram increases the electron density maps by modifying the macromolecule regions. The method adjusts the macromolecular electron density to an ideal distribution [118]. The non-crystallographic symmetry is used when the crystals contain non crystallographic symmetry. Averaging of related regions improves the signal-noise level [119] and can lead to iterative maps improvements.

3.6.1.6 Model building

Model building is the step in the macromolecular crystal structure determination when the model is built into the electron density map. It can be done automatically or

manually. Automatic building uses a model protein to fit its atoms with known chemical identity into the new electron density map. The fitted atoms are later combined into partial models. Either approach uses already determined atomic positions as guides for further model building. Automated building usually has additional features which help to speed up the process. The typically used programs are PHENIX [120] and ARP/WARP [121] or BUCANEER [122]. The features consist of main chain tracing (determination of main chain atoms positions), side chain building (best rotamer conformation fitting), loop building (building missing residues from the protein sequence), secondary structure recognition (determination of chain direction), ligand building (addition of possible ligands) and solvent building [121]. This feature helps the automatic fit to result in a protein model fitted into the electron density. However, automated model building usually does not result in a complete model. Besides, each automated fitting needs manual evaluation. Manual investigation of the model is not very different from the automated one. The advantage of this method is immediate validation of each inspected part of the model and its extension. Manual model building is usually done with programs such as MAIN [123] or Coot [124] crystallographic programs.

3.6.1.7 Protein model refinement

The role of model refinement is to fit model atoms to experimental data. As a consequence of a better fit, the electron density maps resulting from improved structures lead to better and more accurate electron density maps. Usually two types of maps are calculated: 2Fobs-Fcalc (map used to fit the model) and Fobs-Fcalc (map where positive and negative densities indicate absence and wrong positioning of atoms in a model). Fobs are observed reflection amplitudes, Fcalc are structure factors calculated from the model [125]. Refinement combines several energy terms restraining models to chemical restraints such as restraints of bond lengths, bond angles, planarity, improper dihedrals, dihedral angles, electrostatic interactions and Van der Waals or repulsion and crystallographic terms [103].

Common refinement method uses maximum likelihood (ML) crystallographic target. This method uses Bayes' probability theorem, which defines that the probability of x and y is the probability of x, given y is multiplied by the probability of y. Incorporation of this probability determines the ML basics that the best model is consistent with the observations. Each of the results from MR and refinement generates a set of predicted structure factors. How good they are is determined by the likelihood. Each reflection is treated separately as an independent observation, which leads to a set of predicted structure factors. All of those sets are then refined to give out the best explanation of the observed data [103, 126]. The ML relies on cross-validation, which uses statistical quantity for model validation. The cross-validation checks the agreement between observed and computed structure factors [127].

The newest approach uses free kick instead of cross-validation in ML refinement [128]. Typical cross-validation ML uses 5% of the diffraction data for justifying the model accuracy, while the free kick uses 95% or more of the measured data for calculating phase error estimates. The free kick refinement is based on estimation of the coordinate error from comparison of structure factors from model against the working set of observed structure factors. The model atoms are kicked from their current positions, which frees them from their restraints imposed by the energy function and are rebuild into the electron density with new refined structure factors using the ML method. The resulting model is a little worse but is better evaluated [128]

3.6.1.8 Structure validation

The validation of the built model is carried out during all model building steps. The final validation determines the reliability of the data and model quality. The Rfree factor, which is calculated from the 5% of the diffraction data during cross-validation, should not exceed 7% of the R value, which is the indicator of the quality of the structure (the lower the R value, the model is more probable) [125]. However, the new free kick refinement removes the relevance of the Rfree factor, since the model is refined against all data [128]. Model validation includes deviations from ideal bonding terms, planarity, chirality, packing of atoms, and Ramachandran plots. They provide insight in the energetically allowed regions of the main chain backbone dihedral angles (φ,ψ) [125]. Additional validation parameter is the temperature (B-) factor, which describes the displacement of atoms from their positions [129].

3.6.2 Structure determination protocols

All diffraction data sets were collected at the BESSY BEAMLINE 14.1 in Berlin on PILATUS 6M detector using 1.0 Å wavelength [130]. The data were processed using XDSAPP [109] package.

3.6.2.1 Human cathepsin L-C25S

The structure was determined by molecular replacement using PHASER [131] with cathepsin L PDB ID: 3K24 [44] as a search model. The structure was refined using REFMAC5 [132] and MAIN [123].

3.6.2.2 Human cathepsin V-C25S

The structure was determined by molecular replacement using PHASER [131] with cathepsin V PDB ID: 3KFQ [133] as a search model. The structure was refined using REFMAC5 [132] and MAIN [123].

3.6.2.3 Cathepsin V-C25S with peptides

The structures were determined by molecular replacement using PHASER [131] with cathepsin V PDB ID: 3H6S [133] as a search model. The structures were refined using REFMAC5 [132] and MAIN [123].

3.6.2.4 Cathepsin K complex structure determination

Crystals of Cat K with N-(2-Amino-4-chlorobenzoyl)homocycloleucyl-glycinonitrile, N-(2-Amino-4-bromobenzoyl)homocycloleucyl-glycinonitrile inhibitors were diffracting to 1.6 Å at home source using Bruker Microstar rotation anode, whereas the crystals of Cat K with N-(3-Fluoro-4-aminobenzoyl)homocycloleucyl-glycinonitrile inhibitor diffracted to 1.0 Å at BESSY BEAMLINE 14.1 in Berlin on PILATUS 6 M detector using 1.0 wavelength. 59 data sets of the first two crystals were processed with PROTEUM (Proteum Software Pty Ltd.), whereas the third one was processed with XDSAPP.60. The diffraction pattern of all crystals revealed the presence of a single crystal lattice. Data sets were scaled in P21 space group for Cat K with N-(2-Amino-4-chlorobenzoyl)homocycloleucyl-glycinonitrile, in P 212121 space group for Cat K with N-(2-Amino-4-bromobenzoyl)homocycloleucyl-glycinonitrile and Cat K with N-(3-Fluoro-4-aminobenzoyl)homocycloleucyl-glycinonitrile. Data statistics are presented in Table S4.

Matthews coefficient analysis [134] ambiguously defined one molecule in the asymmetric unit for all three Cat K structures. The structures were solved using PHASER [131] with the PDB model 4DMX [135]. Refinement was performed using REFMAC5 [132] and MAIN [123]. Model building was done using Coot [124] and MAIN [123]. Complete chain of Cat K from 1st to 215th in all three crystals is well resolved by electron density maps. Several residues were built with alternative conformations. One potassium and one chloride ion were built in the N-(2-Amino-4-chlorobenzoyl)homocycloleucyl-glycinonitrile complex. Five SO_4^{2-} anions were built in the crystal complex of Cat K with N-(2-Amino-4-bromobenzoyl)homocycloleucyl-glycinonitrile. Six SO_4^{2-}, two NH^{4+}, and an AcO^- anion were built in the N-(3-Fluoro-4-aminobenzoyl)homocycloleucyl-glycinonitrile complex. Determined coordinates of Cat K with N-(2-Amino-4-chlorobenzoyl)homocycloleucyl-glycinonitrile, N-(2-Amino-4-bromobenzoyl)homocycloleucyl-glycinonitrile and N-(3-Fluoro-4-aminobenzoyl)homocycloleucyl-glycinonitrile were deposited in PDB as 4X6J, 4X6I, and 4X6H, respectively.

3.6.3 Data collection and refinement statistics

The data collection and refinement statistics of human cathepsin L-C25S and cathepsin V-C25S complexes are presented in Tables 3.2 and 3.3. Structure of human cathepsin V-C25S from the crystallization before any further manipulation is referred to as native complex. Successful complexes of human cathepsin V-C25S got from peptide soaking with ALAASS, AVAEKQ and LLKVAL peptides are called complex_1_soak, complex_2_soak and complex_3_soak, respectively. The only successful complex of cathepsin VC25S with LLKVAL peptide obtained by cocrystallization is called complex_4_cocrys.

The data collection and refinement statistics of successful human cathepsin K complexes with N-(2-Amino-4-chlorobenzoyl)homocycloleucyl-glycinonitrile, N-(2-Amino-4-bromobenzoyl)homocycloleucyl-glycinonitrile and N-(3-Fluoro-4-aminobenzoyl)homocycloleucyl-glycinonitrile are presented in Table 3.4.

Table 3.2: Data collection and statistics of cathepsins L-C25S and V-C25S complexes. Values in parentheses are for the highest resolution shell. One crystal was used for the structure. No reflection cutoffs were applied. Structures are not deposited in PDB database. N/A – means that the structures were not yet deposited.

Data collection	catL-C25S	Native complex	Complex_1_soak
PDB ID	5I4H	N/A	N/A
Space group	P 21 21 21	P 43 21 2	P 43 21 2
Cell dimensions			
a, b, c (Å)	40.0 67.4 77.4	94.0 94.0 124.0	94.0 94.0 125.7
$\alpha = \beta = \gamma$	90.0	90.0	90.0
Resolution (Å)	38.7 – 1.4	43.93 – 1.5	45.7 – 2.1
R_{merge} (%)	0.053 (0.437)	0.113 (0.942)	0.110 (0.839)
I/σI	20.46 (2.56)	14.67 (2.06)	17.79 (2.8)
Completeness (%)	95 (69)	99 (98)	100 (100)
Redundancy	6.1 (3.7)	12.9 (8.5)	13.3 (12.9)
Refinement			
Resolution (Å)	38.7 - 1.4	43.93 – 1.5	45.7 – 2.1
No. of reflections (work/free)	38421/2741	89222/2000	35514/1120
R_{work}/R_{free}	0.147/0.184	0.162/0.186	0.178/0.224
B factors			
Protein	15.70	18.28	33.81
Water	29.70	31.00	48.52
No. of atoms			
Protein	1749	3500	3455
Water	237	320	267
r.m.s. deviation			
Bond lengths (Å)	0.025	0.005	0.027
Bond angles (°)	2.15	0.77	1.81

Table 3.3: Data collection and statistics of cathepsin V-C25S complexes. Values in parentheses are for the highest resolution shell. One crystal was used for the structure. No reflection cutoffs were applied. Structures are not deposited in PDB database. N/A – means that the structures were not yet deposited.

Data collection	Complex_2_soak	Complex_3_soak	Complex_4_cocrys
PDB ID	N/A	N/A	N/A
Space group	P 43 21 2	P43 21 2	P43 21 2
Cell dimensions			
a, b, c (Å)	93.5 93.5 125.9	94.2 94.2 126.2	96.1 96.1 125.6
$\alpha = \beta = \gamma$	90.0	90.0	90.0
Resolution (Å)	45.6 – 1.6	47.1 – 1.7	48.1 – 1.7
R_{merge} (%)	0.114 (0.989)	0.109 (1.079)	0.117 (1.011)
$I/\sigma I$	17.93 (2.64)	15.23 (2.31)	14.89 (2.24)
Completeness (%)	100 (100)	100 (100)	100 (95)
Redundancy	13.5 (13.7)	13.5 (13.4)	13.3 (12.2)
Refinement			
Resolution (Å)	45.6 – 1.6	47.1 – 1.7	48.1 – 1.7
No. of reflections (work/free)	75383/2098	61888/2096	60130/2100
R_{work}/R_{free}	0.159/0.185	0.164/0.198	0.157/0.185
B factors			
Protein	19.90	25.55	25.04
Water	32.63	39.45	34.95
No. of atoms			
Protein	3516	3528	3481
Water	393	387	361
r.m.s. deviation			
Bond lengths (Å)	0.025	0.023	0.02
Bond angles (°)	2.18	2.01	1.86

Table 3.4: Data collection and statistics of cathepsins K complexes with N-(2-Amino-4-chlorobenzoyl)homocycloleucyl-glycinonitrile, N-(2-Amino-4-bromobenzoyl)homocycloleucyl-glycinonitrile and N-(3-Fluoro-4-aminobenzoyl)homocycloleucyl-glycinonitrile. Values in parentheses are for the highest resolution shell. One crystal was used for the structure.

Data collection	N-(2-Amino-4-chlorobenzoyl)homocycloleucyl-glycinonitrile	N-(2-Amino-4-bromobenzoyl)homocycloleucyl-glycinonitrile	N-(3-Fluoro-4-aminobenzoyl)homocycloleucyl-glycinonitrile
PDB ID	4X6J	4X6I	4X6H
Space group	P 1 21 1	P 21 21 21	P 21 21 21
Cell dimensions			
a, b, c (Å)	47.22, 44.40, 51.35	36.46, 54.45, 87.87	36.46, 54.45, 87.87
α, β, γ	90.0, 116.42, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	45.99 – 1.59 (1.642 – 1.585)	46.28 – 1.87 (1.937 – 1.87)	46.37 – 1.0 (1.036 – 1.0)
R_{merge} (%)	0.085 (1.124)	0.259 (0.981)	0.052 (0.598)
I/σI	9.12 (1.17)	9.96 (3.78)	15.89 (1.79)
Completeness (%)	97.53 (90.01)	98.31 (91.60)	96.78 (71.72)
Redundancy	1.9	9.5	5.9
Refinement			
Resolution (Å)	45.99 – 1.59 (1.642 – 1.585)	46.28 – 1.87 (1.937 – 1.87)	46.37 – 1.0 (1.036 – 1.0)
No. of reflections (work/free)	25360	14804	92375
R_{work}/R_{free}	0.189/0.241	0.181/0.237	0.111/0.135
B factors			
Protein	17.3	7.9	10.3
Water	30.6	16.4	33.5
No. of atoms	2004	1909	2118
Protein	215	215	215
r.m.s. deviation			
Bond lengths (Å)	0.019	0.019	0.022
Bond angles (°)	1.91	1.76	2.1

Chapter 4

Results

The structure of the peptides H3 19-33 derived from the histone H3 in the complex with cathepsin L [44] prompted us to approach structural studies of substrate binding to cysteine cathepsins in a systematic way. First we made an attempt to obtain diffracting crystals of inactive mutants of cathepsins V and L and then tried to select representative peptides by informatics analysis, which included proteomic digestion data, and soak them in the crystals.

We made an attempt to explore halide derivatives of Balicatib (a drug candidate) interactions with the S subsite of cathepsin K.

4.1 Cathepsin V-C25S Structures

We were able to establish a model for cathepsin substrate binding interaction studies. The model is based on cathepsin V-C25S protein crystals, which structure provides space for peptide binding studies. Simultaneous attempts of soaking and cocrystallization of peptides ALAASS, AVAEKQ, LLKVAL and AVRARL with cathepsins V-C25S and L-C25S were made. Three peptides ALAASS, AVAEKQ and LLKVAL were successfully soaked with cathepsin V-C25S, the AVRARL attempt failed. The cocrystallization was successful only for one of the peptides, the LLKVAL peptide. The following structures were determined: a native complex of cathepsin V-C25S with a propeptide part bound to the active site, called the "native complex" and the complexes of cathepsin V-C25S soaked with ALAASS, AVAEKQ and LLKVAL peptides called "complex_1_soak", "complex_2_soak" and "complex_3_soak", respectively. In addition, we were able to determine the crystal structure of complex of LLKVAL peptide by cocrystallization, but failed to determine the crystal structures of cathepsin V-C25S in complex with peptides ALAASS, AVAEKQ and AVRARL. The description of amino acid residues positions in complexes, peptides will be numbered from left to right with digitals 1 to 6, respectively.

The cathepsin V-C25S model or native complex crystallized in tetragonal P43212 space group with the unit cell dimensions: a = 93.97, b = 93.97, c = 123.96, $\alpha=\beta=\gamma$ = 90.00°. Figure 4.8 shows two asymmetric molecules A (white) and B (black) with sequences from D-1 to V221 and L1 to V220 using mature enzyme nomenclature, respectively. Protein residues were unambiguously resolved by the electron density map. The positioning of the D-1 residue is partially supported by the electron density map. We found that the cathepsin V-C25S active site is occupied with a tripeptide. The density interpretation suggests sequence E48, Y49, S50 originating from the cathepsin V propeptide. The peptide is bound to S3-S1 binding sites. In addition, 385 solvent molecules, four glycerol and two MPD molecules, one sodium ion and one chloride ion were incorporated in the structure.

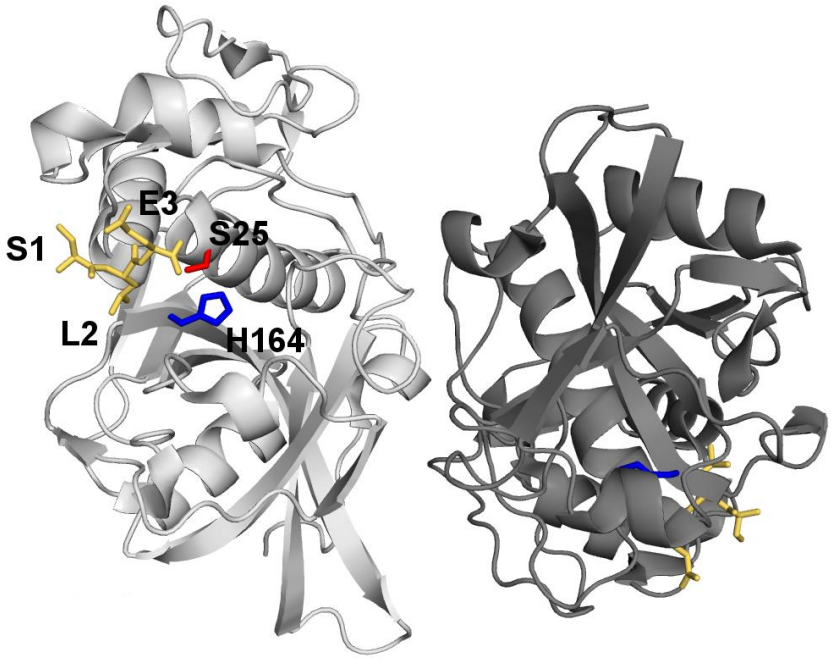


Figure 4.1: Two asymmetric units of cV-C25S crystals. Two asymmetric units are colored white and black. The active site serine and histidine are colored red and blue respectively. The SLA peptide is colored light orange. Active site and peptide residues are labelled. Left and right domain are labelled with (L-) and (R-), respectively. The picture was prepared using the Open-source PyMOL™ v.1.7.2.1. (The PyMOL Molecular Graphics System, Version 1.7.2 Schrödinger, LLC).

Both molecules of cathepsin V superimpose well with each other and with three other structures of cathepsin V, the crystal structure of human cathepsin V complexed with an irreversible vinyl sulfone inhibitor [136] (rhcV_1), the cathepsin V complex with clitocypin [133] (rhcV_2) and the unpublished complex of cathepsin V complex with stefin A [137] (rhcV_3) (Table 4.1).

Table 4.1: The RMSD values of cathepsin V complexes native complex with complex_1_soak, complex_2_soak and complex_3_soak. Both molecules A and B from native complex were aligned with available structures. Nalign stands for the number of aligned residues.

	Proteins	Native complex molecule A	Nalign (A)	Native complex molecule B	Nalign (B)
Cα-atoms RMSD values (Å)	Native complex molecule B	0.44	221	-	-
	Complex_1_soak	0.44	221	0.32	221
	Complex_2_soak	0.35	220	0.37	220
	Complex_3_soak	0.35	221	0.37	221

Table 4.2: Summary of peptide binding to the cathepsin V-C25S subsites. The S3-S2' specifies the substrate binding sites according to Schechter and Berger nomenclature while the S3'-S6' were used to keep the consistency of the nomenclature of possible binding areas for longer peptides. Presented peptide residues in the table are bound to the molecule A from their structures, unless specified differently. The (B) letter next to the residues refers to the second molecule B in the structure. Subsites with n/e describe that the electron density in those subsites was not visible for residue determination. "–" means no binding.

Structure	Noticed peptides	Active site binding subsites								
		S3	S2	S1	S1'	S2'	S3'	S4'	S5'	S6'
Native complex	EYS	E1	Y2	S3	-	-	-	-	-	-
Complex_1_soak	ALAA	-	-	-	A1	L2	A3	A4	n/e	n/e
Complex_2_soak	AVAEKQ	-	-	-	A1	V2	A3	E4	n/e K5 (B)	n/e Q6 (B)
Complex_3_soak	EY, LLKV	E1	Y2	-	L1	L2	K3	V4	n/e	n/e
Complex_4_cocrys	LLK	L1	L2	K3	-	-	-	-	-	-

We expected that the EYS peptide is exchangeable with another peptide when crystals of cathepsin V C25S mutant were soaked with a ~5x molar excess of the soaked peptide. However, in several attempts the EYS peptide remained bound within the primed region of the active site of molecule B. We noticed that all soaked peptides bound to the primed region of the active site. Besides, in complex_1_soak, complex_2_soak and complex_3_soak we determined a chlorine ion and MPD molecule in the S1 and S2 subsites, respectively. We were not able to determine the biological relevance of the chlorine ion, however we suspect that it is mimicking the carboxyl group of a peptide and stabilizes overall structure. Hydrophobic molecule MPD was found in a hydrophobic subsite. The complex_4_cocrys structure has LLK peptide bound to the non-primed region, which is opposite to observed interactions in the complexes from peptide soaking. Besides, of the available space and length of the peptide, which would suggest binding to

the primed region we were not able to determine the last three peptide residues (VAL). The peptide binding to the native complex was summarized in Table 4.2 and presented in Figures 4.2-5.

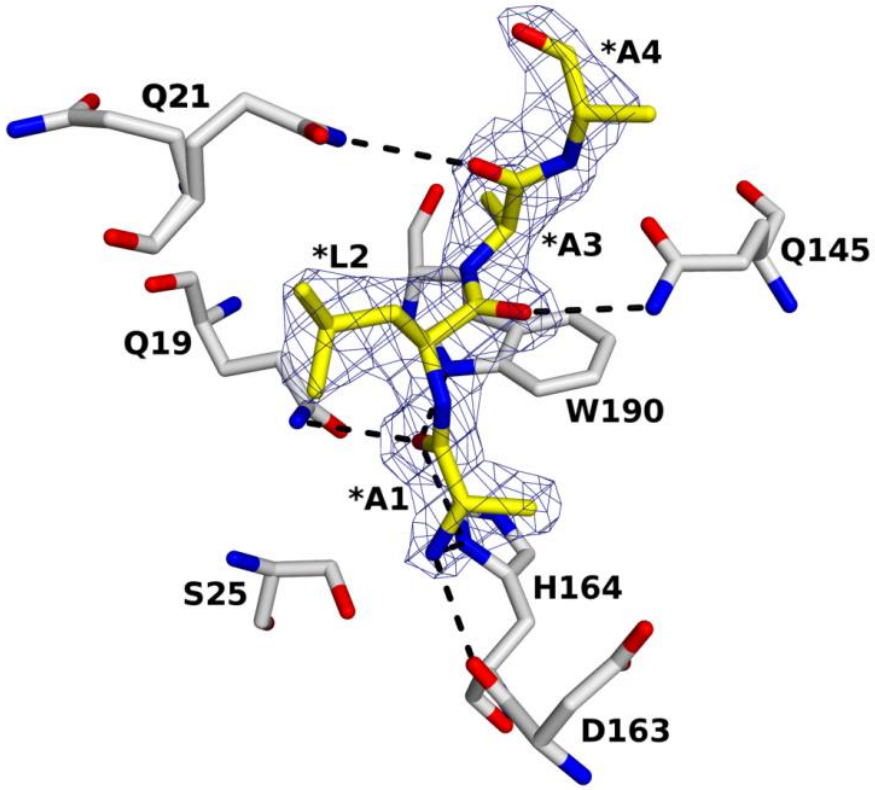


Figure 4.2: Structural insight in the cathepsin V-C25S complex_1_soak. The ALAASS peptide and protein residues are presented with sticks. The chlorine atom is presented as a green sphere. The A1-S6 peptide and protein residue carbon atoms are colored yellow and white, respectively. Oxygen and nitrogen atoms are colored red and blue, respectively. Peptide and protein residues are labelled. The peptide residues are additionally marked with *. Electron density around the peptide is shown as a blue mesh contoured at 1.0 σ. The picture was prepared using the Open-source PyMOL™ v.1.7.2.1. (The PyMOL Molecular Graphics System, Version 1.7.2 Schrödinger, LLC).

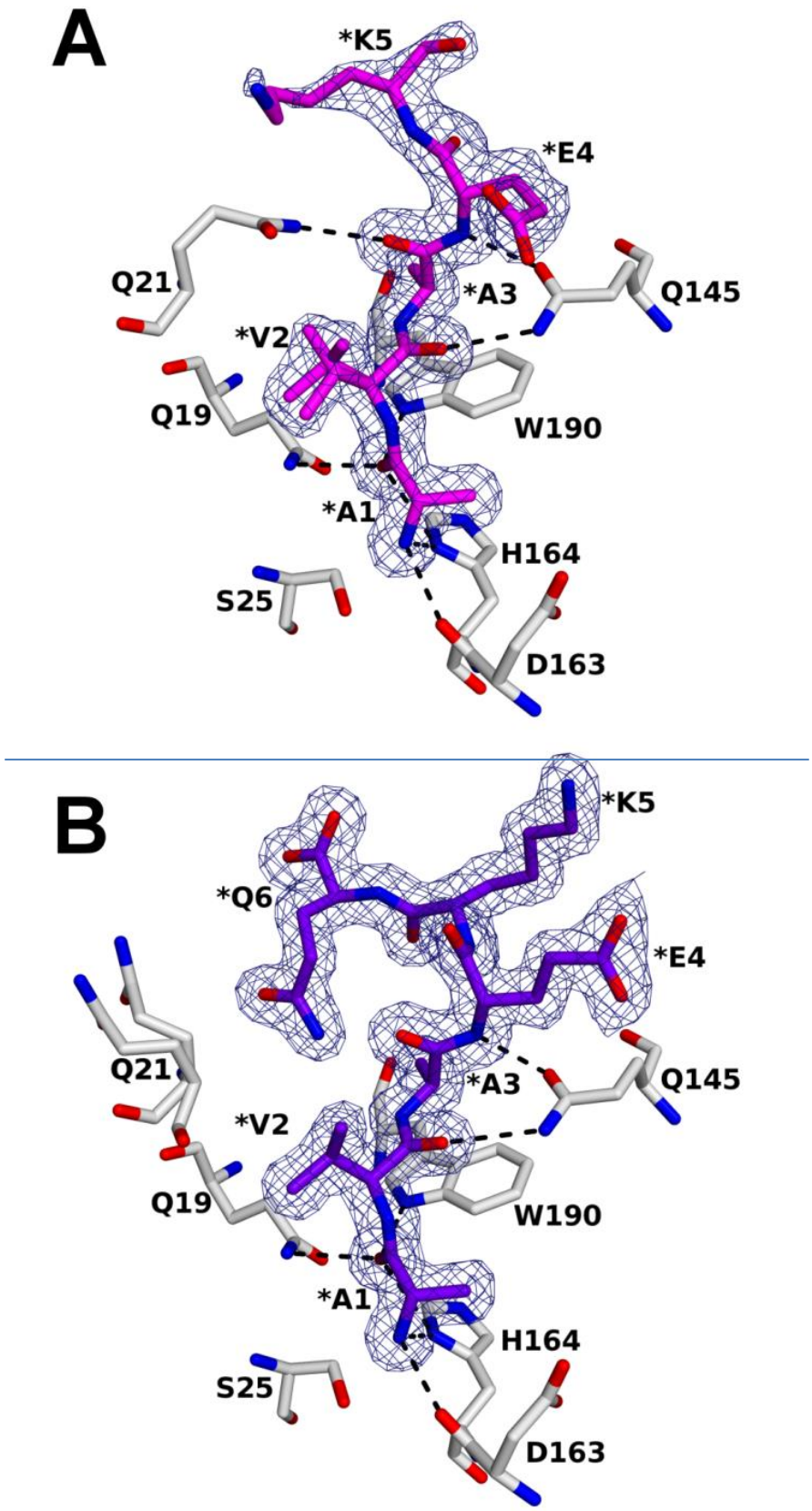


Figure 4.3: Structural insight to cathepsin V-C25S complex_2_soak. The AVAEKQ

peptide and protein residues are presented with sticks. The A1-Q6 peptide in molecule A, B and protein residue carbon atoms are colored magenta, violet and white, respectively. Oxygen and nitrogen atoms are colored red and blue, respectively. The peptide and protein residues are labelled. The A1-Q6 peptide residues are additionally marked with *. Electron density around peptide is shown as a blue mesh contoured at 1.0 σ. The picture was prepared using the Open-source PyMOL™ v.1.7.2.1. (The PyMOL Molecular Graphics System, Version 1.7.2 Schrödinger, LLC).

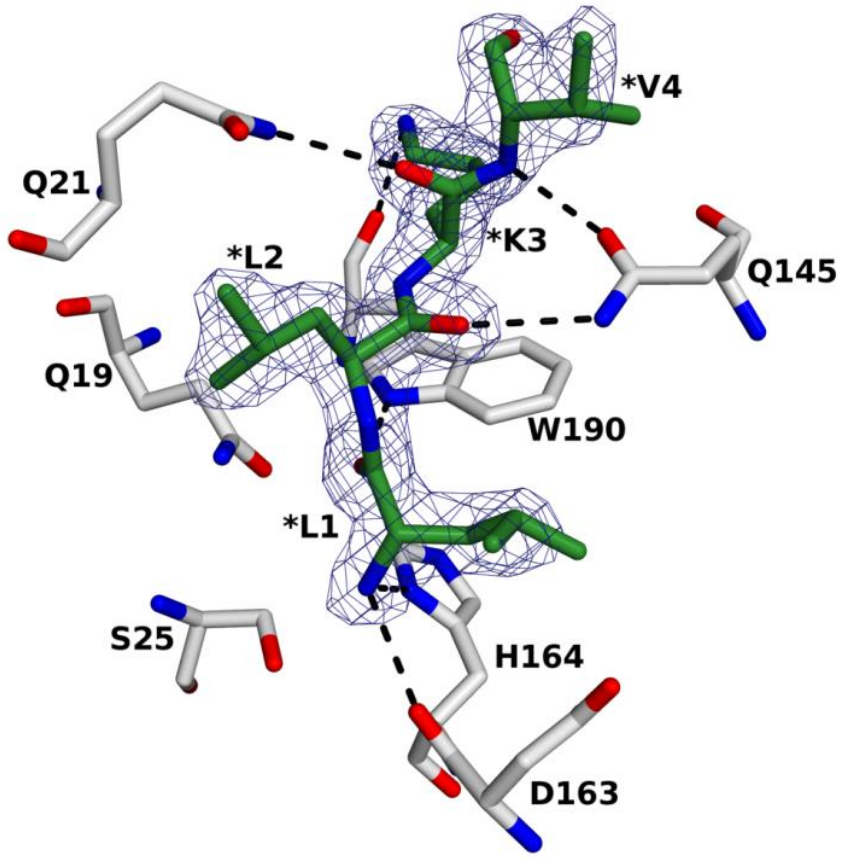


Figure 4.4: Structural insight to cathepsin V-C25S complex_3_soak. The LLKVAL peptide and protein residues are presented with sticks. The L1-L6 peptide and protein residue carbon atoms are colored dark green, light orange and white, respectively. Oxygen and nitrogen atoms are colored red and blue, respectively. The peptide and protein residues are labelled. The L1-L6 peptide residues are additionally marked with *. Electron density map around peptide is shown as a blue mesh contoured at 1.0 σ. The picture was prepared using the Open-source PyMOL™ v.1.7.2.1. (The PyMOL Molecular Graphics System, Version 1.7.2 Schrödinger, LLC).

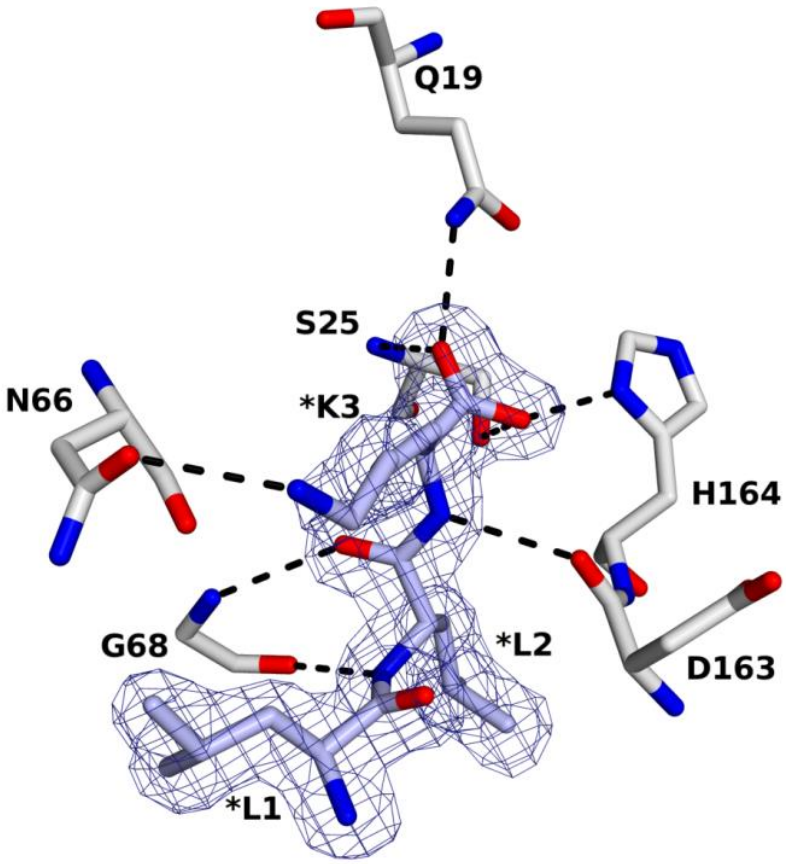


Figure 4.5: Structural insight to cathepsin V-C25S complex_4_cocrys. The LLKVAL peptide and protein residues are presented with sticks. The L1-L6 peptide and protein carbon atoms are colored light blue and white, respectively. Oxygen and nitrogen atoms are colored red and blue, respectively. The peptide and protein residues are labelled. The L1-L6 peptide residues are additionally marked with *. Electron density around peptide is shown as a blue mesh contoured at 1.0 σ. The picture was prepared using the Open-source PyMOL™ v.1.7.2.1. (The PyMOL Molecular Graphics System, Version 1.7.2 Schrödinger, LLC).

Determined structures show peptides either at non-primed or at the primed active site regions. Complete amino acid residue chains of molecules A and B were unambiguously resolved with several residues built with alternate conformations.

The complex_4_cocrys peptide binds to the S3-S1 subsites as in the EYS peptide in native complex. In the complex_4_cocrys the three first residues L1, L2 and K3 were built in the S3-S1 binding sites, respectively, while for the other three residues the corresponding electron density map was not observed. In this complex in the primed active site region two MPD molecules were built.

In three other complexes (complex_1_soak, complex_2_soak and complex_3_soak) the peptides were however built into the primed active site region. The Peptide's N-terminal first residue was built in the S1' subsite and the following residues were built onwards. In addition, a chlorine atom and MPD molecule were built in the oxyanion hole and S2 subsite of molecule A, respectively. An exception is the complex_3_soak where the EY peptide was built instead of the MPD molecules. In all structures apart from the

complex_2_soak, the molecule B contained the EYS sequence as in the native complex. In the complex_2_soak structure we build the whole peptide from A1-Q6.

4.1.1 Residue binding to cathepsin V-C25S active sites and beyond

In this paragraph results of residues binding sites are described from S3 to S3'. Protein residues are numbered as in the structures of the cathepsin V enzyme.

Two structures (native complex and complex_4_cocrys) have amino acid residues bound to S3, S2 and S1 binding sites. Three structures (complex_1_soak, complex_2_soak and complex_3_soak) reveal binding of amino acid residues into the S1' and S2' binding sites and beyond.

4.1.1.1 P3 binding to S3 binding site

Two residues were determined to bind to S3 area, L1 (P3) from complex_4_soak peptide and native complex E1 (P3) from the propeptide part and are presented in Figure 4.6 A. Both residues do not make main chain atoms interactions with the enzyme. The side chain NZ atom of the K3 forms a hydrogen bond with main chain oxygen atom of G61. Superimposition of both structures did not indicate any significant differences within the S3 binding site apart from the F69 side chain, which is in the complex_4_cocrys structure positioned ~1 Å away from the P3 residue compared to the native complex structure.

4.1.1.2 P2 binding to S2 binding site

Residues in two structures: complex_4_soak and native complex bind to S2 site, L2 (P2) from LLKVAL peptide and Y2 (P2) from propeptide part (Figure 4.6 B). Main chain atoms from both residues form antiparallel hydrogen bond ladder with the cathepsin V G68 main chain atoms. The amino acid side chains bind into the S2 pocket utilizing hydrophobic interactions.

Superimposition of both structures did not indicate significant residue repositions within the S2 binding site.

4.1.1.3 P1 binding to S1 binding site

K3 and S3 from the structures of the complex_4_cocrys and the native complex were observed bound into the S1 site from the propeptide part (Figure 4.6 C). Both residues have free carboxy group at C-terminus. The main chain nitrogen atoms of these two residues form a hydrogen bond with the cathepsin V D163 main chain oxygen. While the carboxylic group forms hydrogen bonds with the nitrogen atoms of Q19, S25 and H164 side chain atoms. The P1 residue side chain points towards the solvent region as previously described in 1.2.1. The K3 (P1) side chain NZ atom from the complex_4_cocrys forms a hydrogen bond with cathepsin V N66 side chain OD1 atom, whereas the S3 (P1) side chain from the native complex does not interact with cathepsin V.

Superimposition of both structures did not show significant residue repositions within the S1 binding site.

4.1.1.4 P1' binding to S1' binding site

Two different residues A1 and L1 bind into the S1' subsite of the complex_1_soak and complex_2_soak complex_3_soak peptide structures (Figure 4.6 D). The P1' residues have a free amine group at N terminal. The P1' main chain nitrogen atom form hydrogen

bonds with cathepsin V H164 side chain ND1 atom and D163 main chain oxygen atom. The P1' main chain oxygen atoms form hydrogen bonds with cathepsin V Q19 and Y190 side chain nitrogen atoms. The L1 (P1') side chain faces away from cathepsin V Q145 side chain towards D163 side chain.

Superimposition of all three structures did not show significant residue repositions within the S1' binding site.

4.1.1.5 P2' binding to S2' binding site

Two different residues L2 and V2 bind into the S2' subsite, of the structures of the complex_1_soak and complex_3_soak of both molecules A and B of complex_2_soak peptide (Figure 4.6 E). The main chain oxygen atoms form hydrogen bond with cathepsin V Q145 side chain nitrogen NE2 atom. Three out of the four determined interactions within the S2' binding site were confined by the cathepsin V Q21 side chain which forms a hydrogen bond with the carbonyl oxygen main chain atom of the P3' residues. This interaction forces the L2 and V2 side chains to be turned towards the S1 binding site. The fourth interaction was found in the molecule B of complex_2_soak, where the cathepsin V Q21 side chain is turned by ~180° compared to other described here complexes and leads to alternate conformations of the V2 residue.

Besides the described ones, the superimposition of these structures did not show other significantly different positions of residues within the S2' binding site.

4.1.1.6 P3' binding to S3' position

Two different residues A3 from complex_1_soak, complex_2_soak and K3 complex_3_soak structures bind into the S3' area (Figure 4.6 F). The main chain oxygen atoms form hydrogen bonds with cathepsin V Q21 side chain nitrogen NE2. Their side chains are turned towards the area between the cathepsin V K20, Q145 and F146 side chains and over W190 and W194. In addition, the K3 side chain NZ atom forms a hydrogen bond with cathepsin V W190 main chain oxygen atom.

Superimposition of structures did not show significant differences between positions of residues within the S3' binding site.

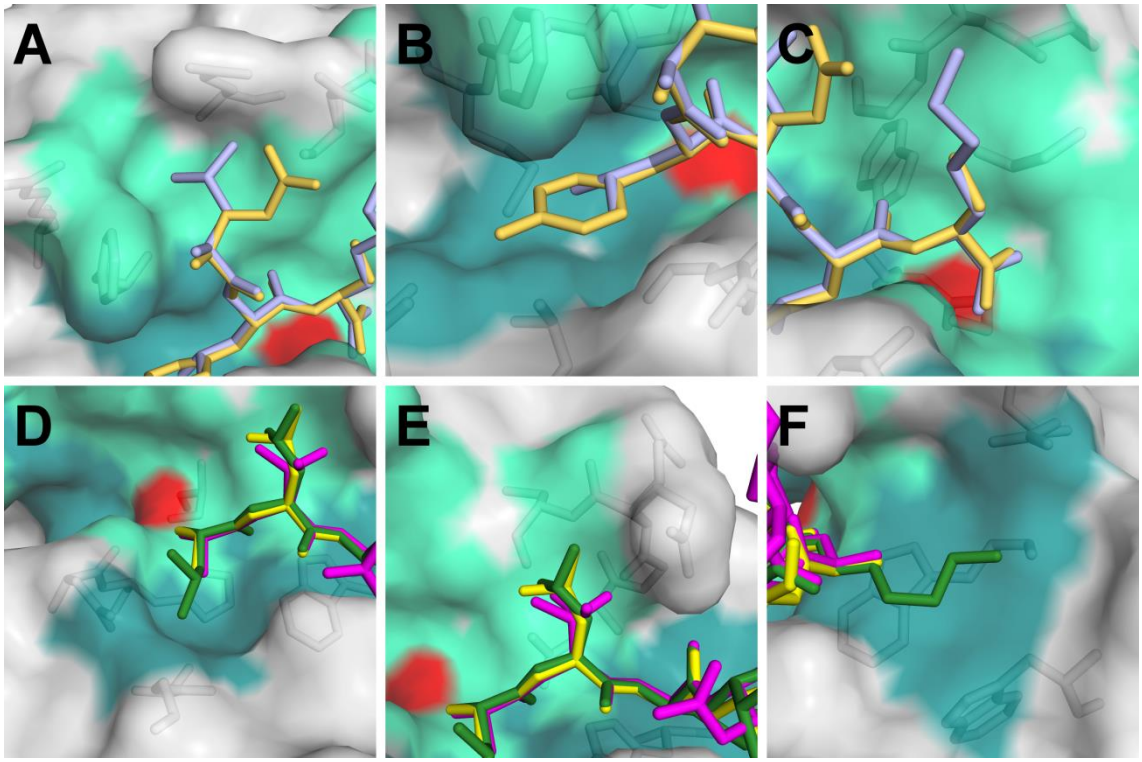


Figure 4.6: Cathepsin V-C25S structural insight to binding sites with peptides. Residues binding within S3, S2, S1, S1', S2' and S3' binding sites are presented on cathepsin V-C25S surface in figures A, B, C, D, E and F, respectively. Peptides and interacting residues are presented with sticks. Peptides from complex_4_soak and native complex are colored light blue and gold. Peptides from complex_1_soak, complex_2_soak and complex_3_soak are colored yellow, magenta and dark green, respectively. The active site subsites are colored in different green hue alternately. The picture was prepared using the Open-source PyMOL™ v.1.7.2.1. (The PyMOL Molecular Graphics System, Version 1.7.2 Schrödinger, LLC).

4.1.1.7 P4' substrate position in cathepsin V-C25S

Three different residues bind within the S4' substrate binding area (Figure 4.7). The A4, E4 and V4 side chains from the complex_1_soak, complex_2_soak and complex_3_soak structures, respectively, are facing towards the region formed by the cathepsin V S142, Q145 and F146 side chains. In addition, the P4' main chain nitrogen atoms form a hydrogen bond with the cathepsin V Q145 side chain OE1 atom. This is the only observed interaction of the P4' residue with cathepsin V.

Superimposition of structures did not show significant differences between positions of residue near the S4' subsite.

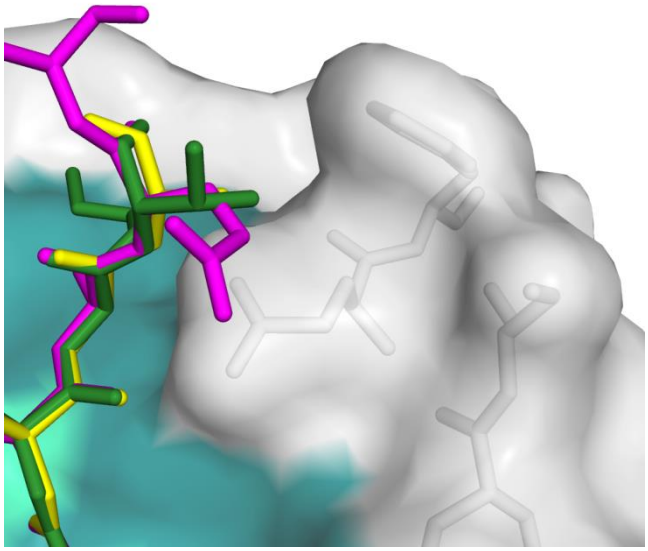


Figure 4.7: Cathepsin V-C25S structural insight to the P4' residue binding on cathepsin V-C25S grey surface. Substrate P4' residues and enzyme residues are presented with sticks. Peptides from complex_1_soak, complex_2_soak and complex_3_soak are colored yellow, green and magenta, respectively. The picture was prepared using the Open-source PyMOL™ v.1.7.2.1. (The PyMOL Molecular Graphics System, Version 1.7.2 Schrödinger, LLC).

4.1.1.8 P5' and P6' substrate position in cathepsin V-C25S

The peptide residues S5 and S6 from complex_1_soak, K5 and Q6 from molecule A of complex_2_soak and A5 and L6 from complex_3_soak correspond to the P5' and P6' positions. They are not involved in any specific interactions with cathepsin V. The electron density maps do not clearly reveal positions of the P5' and P6' residues in neither of the structures (complex_1_soak, complex_2_soak and complex_3_soak). The peptide residues are oriented towards the solvent region.

The only complete peptide from molecule B of the complex_2_soak structure is unambiguously defined by the electron density map. However, in this case the P5' and P6' residues are fixed by interactions with a symmetry related cathepsin V molecule positioned over the active site primed region of molecule B. In addition, the C-terminal peptide residues form hydrogen bonds with the residues from the neighboring molecule.

4.2 Cathepsin L-C25S Complex

Cathepsin L was our second target for studying the binding interactions between cathepsins and substrates. We crystallized cathepsin L-C25S and determined its crystal structure. To our surprise, resulting crystals were inadequate for studying peptide interactions. The determined crystal structure contains cathepsin L-C25S in complex with cleaved cathepsin L bound to the active site.

The cathepsin L-C25S crystallized in the orthorhombic space group P212121 with unit cell dimensions a = 39.97 Å, b = 67.36 Å, c = 77.39 Å. The crystals diffracted to

1.42 Å resolution and contained one molecule in the asymmetric unit. The crystal structure of cathepsin L-C25S mutant contains sequences from A1 to V220 using the mature enzyme nomenclature, with additional residue E-1 at the N-terminus. The presence of E-1 residue was confirmed by the N-terminal sequencing of the protein. The chromatogram presented in Figure 1 shows two possible N-terminal sequences: EAPRS and APRSV, which differ by one residue. They both correspond to the native cathepsin L sequence. In the crystal structure, the residues from A1 to S105 and from D109 to V220 were unambiguously resolved by the electron density maps, whereas the residues V106, A107, and N108 could not be positioned and are therefore absent from the model. The E-1 residue was partially fitted to the electron density. Several residues were built in alternate conformations. The asymmetric unit of the crystal also contains five SO42- ions and a molecule of glycerol.

Cathepsin L's structure reveals the fold of the papain-like enzyme composed of two domains named left (L-) and right (R-). The L-domain is α-helical, whereas the R-domain forms a β-barrel fold, as shown in the pink molecule in Figure 4.8. Between the two domains, there is a V-shaped active site cleft on which the L- and R-domain catalytic residues C25 and H163 are positioned. In this structure, the C25 residue was mutated to serine to abolish its activity.

To our surprise, we found that the cathepsin L molecule in the crystal was cleaved between the residues S105 and V106 within the region located at the surface of the L-domain and that the crystal structure actually contains a part of the peptidyl substrate bound to the non-primed side of the active site. The new carboxylic end (S105) of the N-terminal part of the cleaved peptide is located in the active site of the neighboring molecule related by the ½+x, -½-y, -z symmetry operator. Its residues are marked with an asterisk (*). The C-terminal part of the cleaved peptide is resolved from D109 onwards. It points away from the protein surface. The zig-zag pattern of the N-terminal region of the cleaved peptides linking cathepsin L-C25S molecules in the crystal packing is shown in orange in Figure 2. The peptides are positioned alternatively in front and at the back of the pattern. Hence, the crystal structure of cathepsin L C25S mutant is a complex between the molecule being cleaved and the neighboring molecule performing the cleavage. Because each molecule is cleaved and has the active site performing the cleavage, each molecule simultaneously plays two roles: that of an enzyme and that of a substrate.

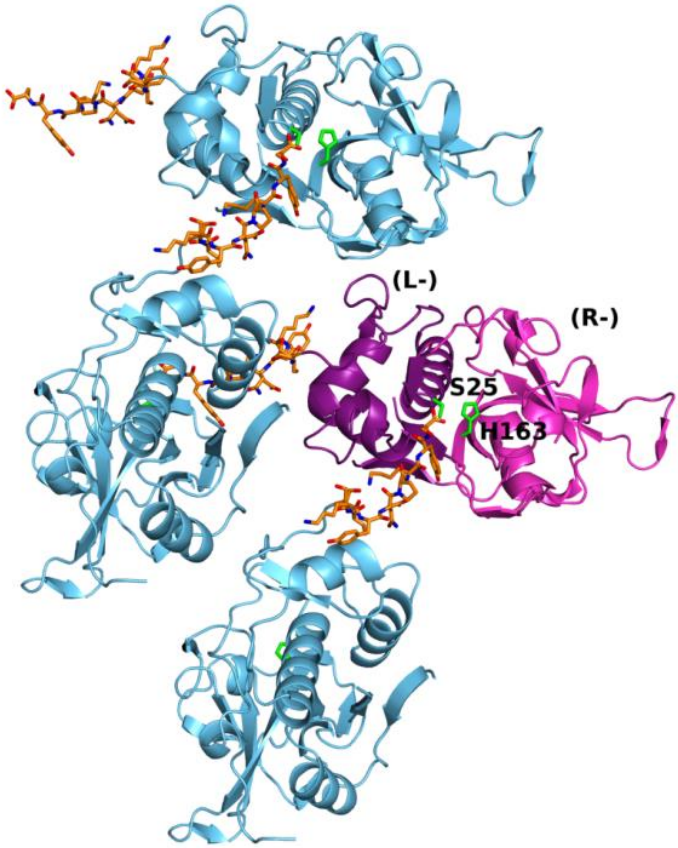


Figure 4.8: The crystal packing of cathepsin L-C25S exhibits zig-zag alignment: Each molecule represents an asymmetric unit. The fold is shown as blue and violet ribbons, while the cleaved peptide reaching into the neighboring molecules is shown as orange sticks. Right and left domains are shown in different violet color hue. The active site residues are labelled and colored in cyan. Left and right domains are marked (L-) and (R-), respectively. The picture was prepared using the Open-source PyMOL™ v.1.7.2.1. (The PyMOL Molecular Graphics System, Version 1.7.2 Schrödinger, LLC).

4.2.1 Residues binding to cathepsin L-C25S active sites and beyond

The structure of the cleaved peptide and its interaction with the neighboring cathepsin L-C25S molecule in detail are shown in Figure 4.9. The cleaved peptide residues K103, Y104 and S105, at positions P3, P2 and P1, respectively (using Schechter and Berger nomenclature [13]), are positioned within the S3, S2 and S1 binding sites of the neighboring molecule, respectively. The C-terminal residue S105 (P1) forms three hydrogen bonds with main chain and side chain atoms of residues Q19*, S25* and H163*. The interatomic distances between S105 carbonyl oxygens and Q19* atom NE1, S25* atom OG and H163* atom ND1 side chains are 2.95 Å, 2.75 Å and 2.8 Å, respectively. The Y104 (substrate residue P2) fills the hydrophobic S2 binding site between A135*, L69*, and D162*. The Y104 hydroxylic group does not make any additional interactions. The K103 side chain (substrate residue P3) is located in the S3 binding site. The K103 side chain amino group stabilizes the binding of the cleaved peptide by hydrogen bond interactions with three oxygen atoms: the S97 side chain atom OG and the C98 and N101 main chain carbonyls. The residues before K103 (P3) are not involved in binding to

the neighboring molecule. Their structure is stabilized within the region of the cleaved peptide itself. The P102 residue is located over the K103, protecting the lysine side chain from solvent. The N101 is present in two conformations; in one, the side chain forms a hydrogen bond with the K103 main chain oxygen atom, whereas in the other, it is oriented ~90° away from the first and it weakly interacts with the Y100 main chain nitrogen atom. The Y100 is positioned parallel to the K99 and forms a hydrogen bond with the D84 oxygen OD2 atom. The K99 side chain is positioned next to the Y89 and does not interact with the enzyme residues. Hence, the three residues of the N-terminal part of the cleaved peptide are stabilized within the active site of the neighboring molecule, whereas the structure of the preceding residues is stabilized within the environment of their own molecules.

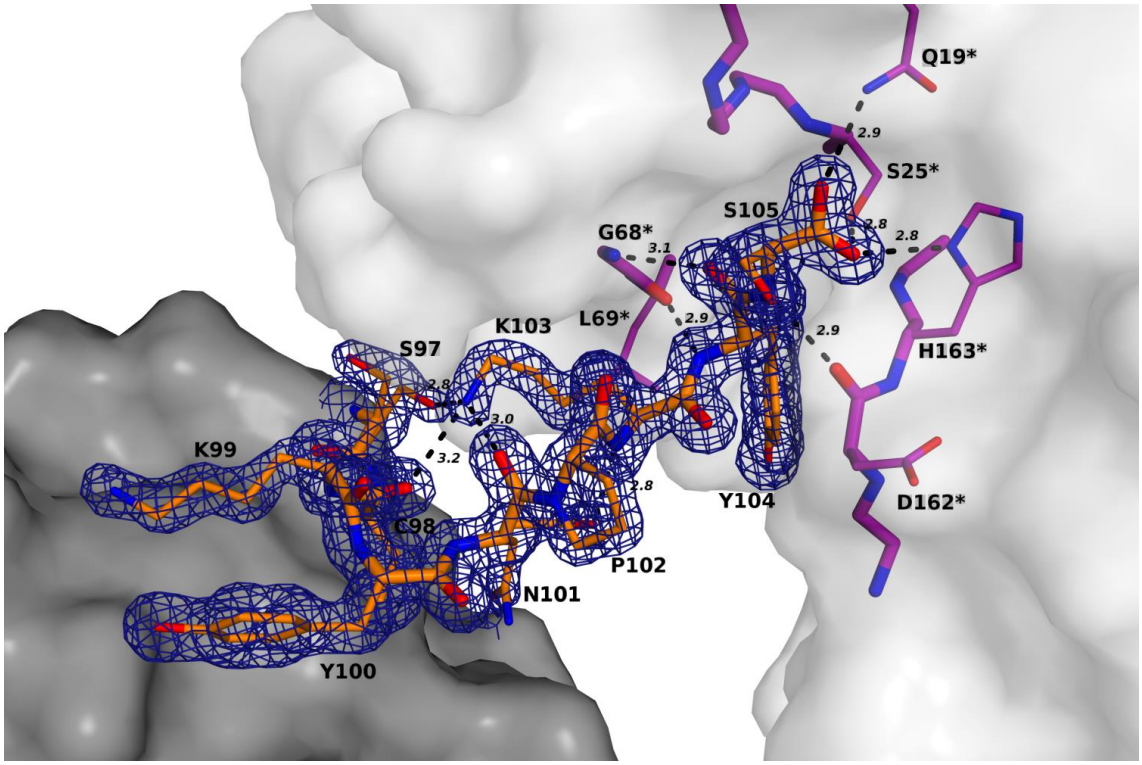


Figure 4.9: Cleaved peptide binding into the active site cleft of the neighboring molecule. The cleaved peptide and residues in contact of the neighboring molecule are shown in stick presentation on the background of white and grey surfaces of cathepsin L molecules. The thicker and thinner sticks represent the main and side chain bonds, respectively. Parts of bonds belonging to oxygen, nitrogen, and sulphur atoms are colored red, blue, and yellow, respectively, whereas the cleaved peptide carbons are shown in orange and those of the attached molecule in pink. Electron density map contoured at 1.0 σ around the peptide is presented as a blue mesh. Hydrogen bonds between the peptide and underlying cathepsin L molecules are presented with dashed lines and their length is shown. The picture was prepared using the Open-source PyMOL™ v.1.7.2.1. (The PyMOL Molecular Graphics System, Version 1.7.2 Schrödinger, LLC).

To confirm that the cathepsin L-C25S was auto-cleaved in two parts also by biochemical analysis, crystals were dissolved and analyzed using SDS-PAGE electrophoresis. Figure 4.10 shows the two bands from the cathepsin L-C25S from dissolved crystals with the apparent weights of 11.5 kDa and 12.6 kDa. They are clearly distinct from the single band of cathepsin L sample before crystallization with the

apparent weight of 24.1 kDa. The weight of two bands corresponds to the parts resulting from the cleavage observed in the crystal structure.

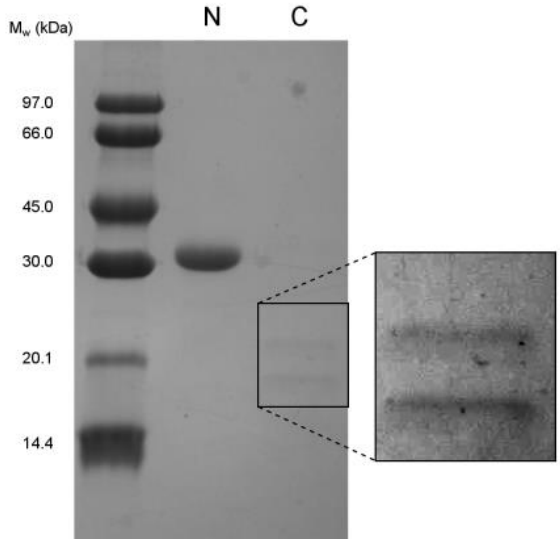


Figure 4.10: Auto-processing of human cathepsin L-C25S during crystallization presented on 15% SDS-PAGE gel colored by Coomassie brilliant blue. The two bands of cathepsin L after the cleavage are enlarged and contrasted for better visualization. Mw, N and C stands for molecular weight marker, native protein before crystallization and crystallized protein, respectively.

To validate the importance of N108 glycosylation for autocatalytic cleavage of procathepsin L was incubated in 100 mM sodium acetate, pH 5.5, 5 mM DTT, 5 mM EDTA buffer at 37°C. On the third day, the sample was degraded, revealed no intermediates, thereby exposing no particular biological relevance of this glycosylation against the cleavage. This suggests that the observed cleavage intermediate can be trapped only under the crystallization conditions.

Determined structure of cathepsin L-C25S shows unique interaction between two cathepsin L molecules trapped in the crystal and indicates that the C25S mutation does not abolish the enzyme activity.

4.3 Cathepsin K with Nitrile Inhibitors

Three structurally similar nitrile inhibitors, with a different benzoyl group, were successfully soaked into the cathepsin K. For each of these three nitrile inhibitors, the best diffracting protein crystals grew in different conditions. The structures show similar way of binding of nitrile inhibitors to the active site of cathepsin K in spite of crystals with different space groups. The complex of cathepsin K with N-(2-Amino-4-chlorobenzoyl)homocycloleucyl-glycinonitrile (Chloro-benzoyl) crystallized in the monoclinic space group P 21, whereas the complexes of cathepsin K with N-(2-Amino-4-bromobenzoyl)homocycloleucyl-glycinonitrile (Bromo-benzoyl) and N-(3-Fluoro-4-aminobenzoyl)homocycloleucyl-glycinonitrile (Fluoro-benzoyl) crystallized in the orthorhombic space group P212121. All three complexes have one molecule of cathepsin K in the asymmetric unit. In all three structures the protein residues as well as the inhibitors are unambiguously defined by the electron density maps. The crystals of 3-

Fluoro-benzoyl complex diffracted to the atomic resolution of 1 Å. They bind into the active site of cathepsin K in the nonprimed region occupying the S1, S2, and S3 binding sites and are characterized by several H-bonds (Figure 4.11). They all form H-bonds with the main chain carbonyl of Asn161, the amide group of Gly66, and the edge of Trp26.

Even thou all three nitrile inhibitors share the same scaffold, the nitrile group of the Chloro-benzoyl formed a covalent bond with the reactive site Sγ atom of Cys25, whereas Bromo-benzoyl did not. In the latter case, the nitrile group points away from the Cys25 reactive site. Interestingly, Fluoro-benzoyl exhibited a dual behavior. Most of the Fluoro-benzoyl (79%) formed a covalent bond with the reactive site SG atom, whereas the rest did not; its nitrile group is clearly chemically unaltered and is positioned similarly to that of Bromo-benzoyl. According to the atomic distances (3.0 and 3.1 Å) the covalently bound nitrile groups form hydrogen bonds with the O atom of Gln19 of cathepsin K, indicating a flip of the Gln19 side chain and the change in the oxyanion hole electrostatic potential.

The cyclohexyl groups of all three inhibitors bind in the S2 pocket in the same orientation, yet the covalently and noncovalently bound nitrile inhibitors exhibit the same positional shift as the glycinonitrile part. The carbonyl O atom of the homocycloleucyl residue is oriented toward the peptide bond N atom of Gly66 and the edge of the indole ring of Trp26.

The benzoyl groups are positioned in the S3 pocket differently. In the Chloro-benzoyl, the interaction between the N atom at the meta position and Asp61 is mediated by a solvent molecule, while the F atom at the meta position in Fluoro-benzoyl is only 2.5 Å away from the carboxylic O atom of Asp61.

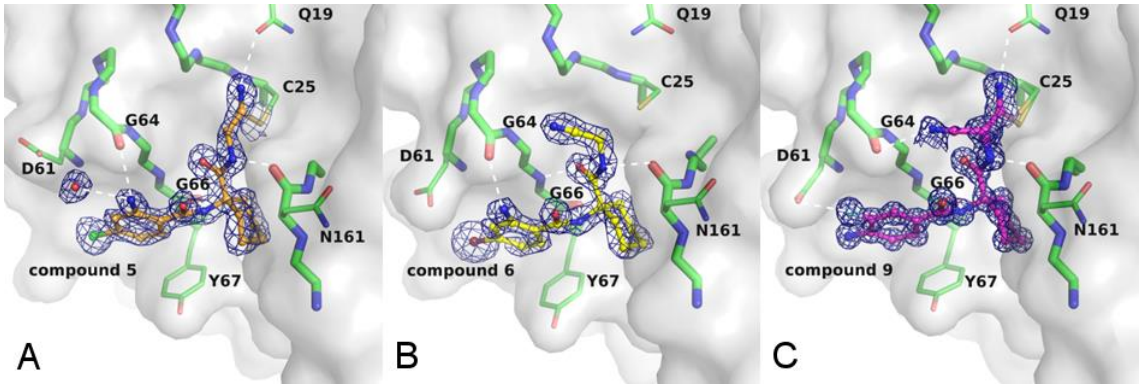


Figure 4.11: Human cathepsin K-inhibitor complex. (A) The blue electron density map around N-(2-Amino-4-chlorobenzoyl)homocycloleucyl-glycinonitrile is contoured at 2.0σ. (B) The blue electron density map around N-(2-Amino-4-bromobenzoyl)homocycloleucyl-glycinonitrile is contoured at 1.0σ. (C) The blue electron density map around N-(3-Fluoro-4-aminobenzoyl)homocycloleucyl-glycinonitrile is contoured at 1.5σ. The picture was prepared using the Open-source PyMOL™ v.1.7.2.1. (The PyMOL Molecular Graphics System, Version 1.7.2 Schrödinger, LLC).

Chapter 5

Discussion

The system for studying interactions between a putatively inactive enzymes, cathepsin V and L with the catalytic site mutations, turned out to be an intriguing model. In contrast to expectations the peptides bound to unexpected positions in the active site, given enough time with cocrystallization the peptide even bound to a different position than during the soaking experiments. And last but not least, the catalytic mutant exhibited low amounts of catalytic activity not observed by the enzyme kinetic studies using fluorogenic substrates.

Cathepsin K complexes with nitrile inhibitors exhibited unexpected behavior in respect to the interaction between the nitrile group and active site cysteine too.

5.1 Evaluation of Experimental Model for Substrate Subsite Studies

Currently there are three available structures of cathepsin V. Those structures are complexes of active cathepsin V with vinyl sulfone inhibitor [136], cathepsin V clitopycin complex [133] and the unreduced cathepsin V complex with stefin A [137]. In those structures the active site is blocked by small or protein inhibitors which prevent peptide interaction studies. In contrast to these structures, which all contained compounds preventing binding along the whole active site cleft, we succeeded in preparation of a crystal form suitable for interaction studies with peptide, by getting the structure of cathepsin V mutant C25S. The structure is characterized with putatively inactive catalytic residue mutation and two protein molecules A and B in the asymmetric unit. Molecules differ in the space available for binding of peptidyl ligands. In the molecule A the active site is in contact with the solvent region, whereas in the molecule B the active site is in part restrained by the residues of the neighboring molecule in the crystal. As the structure of the complex_2_soak demonstrates the crystal packing extended binding surface enabled us to extend the insight into the primed binding region of the active site. The structures of complexes indicated that we were able to establish a model for studies of cathepsin V peptidyl substrate interactions. Four new structures with peptides bound to the active site were determined (complex_1_soak, complex_2_soak, complex_3_soak, complex_4_cocrys). These structures contain bound peptides with amino acid residues which correspond to the positions from P3 to P4'. Structures of these peptides superimpose on each other with small deviations (Figure 5.1). Main chain atoms of the peptides are pinned to the surface of cathepsin V with seven hydrogen bonds and the side chains of all residues positioned in the same orientations apart from the terminal residues, whose positioning exhibits slightly larger variability. The P3-P3' residues lie

along the active site in extended conformation with side chains alternating to the left and right as predicted before [32]. The side chains of residues at positions P3, P1 and P2' interact with the L- domain, whereas the side chains of residues at positions P2, P1' interact with the R-domain residues ([32]).

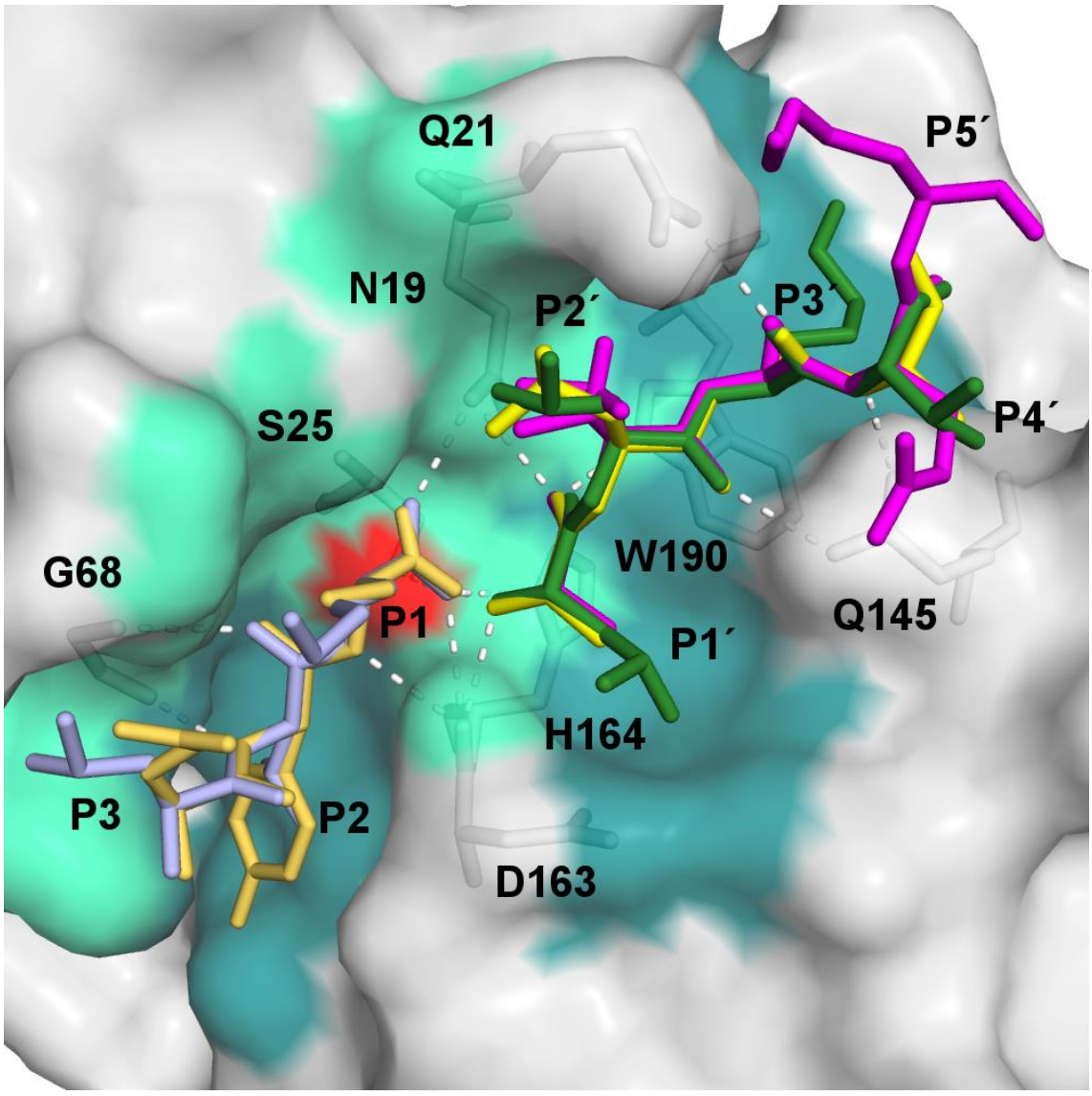


Figure 5.1: Superimposition of peptides binding into the active site of cathepsin V-C25S shown on the white protein surface. The peptides and interacting protein residues are shown as sticks. The peptides: EYS, ALAA(SS), AVAEK(Q), LLKV(AL) and LLK(VAL) from the structures are colored dark yellow, light yellow, pink, green and light blue, respectively. Peptide residues in brackets were not noticed in the structures. The cathepsin V residues forming hydrogen bonds with peptide main chain atoms are labeled. Hydrogen bonds are presented as dashed, white lines. The surface of the substrate binding subsites is colored in different green hue alternately. The picture was prepared using the Open-source PyMOL™ v.1.7.2.1. (The PyMOL Molecular Graphics System, Version 1.7.2 Schrödinger, LLC).

The complex structures however provide no evidence for interactions between the main chain attachment of peptides to the surface of cathepsin V in the primed site beyond the S2' binding site. This suggests that the S3' and the S4' binding sites do not really exist. The presence of interactions however suggests that the term binding area introduced in the work of Turk et al. (1998) is the more appropriate term for describing them [32]. Figure 5.1 shows that peptides in cathepsin V complexes at these two subsites are involved in interactions with the non-conserved side chain of Q145. In cathepsin V the Q145 side forms a hydrogen bond with the peptide P4' main chain atoms. In cathepsin L, however, the equipositioned side chain of L144 can only provide hydrophobic surface. As a consequence, cathepsin L substrates either do not bind into the S3' area or they are engaged in different interactions. The side chain of residue at the P3' position lies between the domains. The side chain at P4' is exposed to the solvent region above the cathepsin V Q21 residue.

Besides, cathepsin L and V differentiate at residue position 145. Cathepsin L has L145, while cathepsin V has Q145 (Figure 5.2). This difference likely affects the substrate binding. While cathepsin V forms hydrogen bonds with main chain of a substrate by the Q145 side chain, this hydrogen bond cannot be established in cathepsin L structure. Moreover, the Q145 side chain (in cathepsin V) turns by 90° downwards to the protein surface (Figure 5.2 B and C), interacts with the peptide and makes space for the P4' side chain binding. This comparison indicates that the substrate's main chain may be positioned differently in different cathepsins.

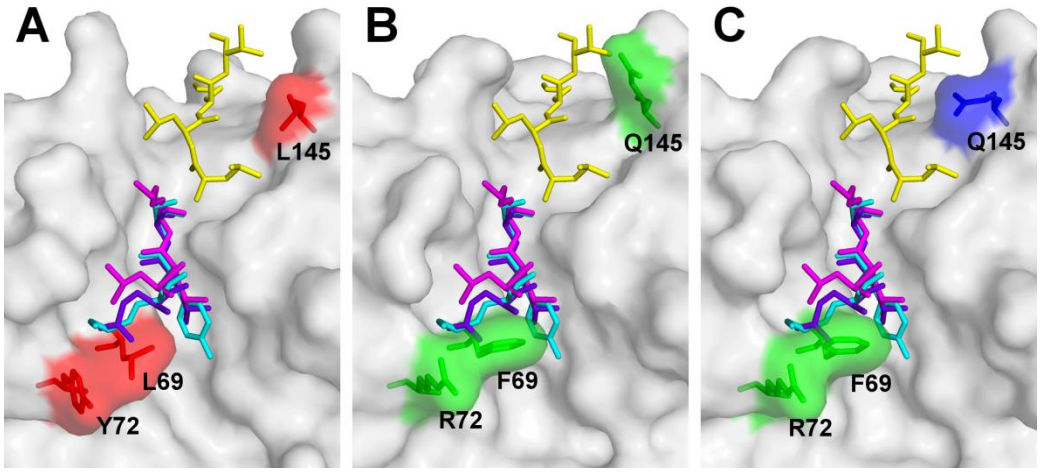


Figure 5.2: Binding of peptides presented on the background of the surface of cathepsin L and V. The binding of four peptides: QLA peptide from cathepsin L-C25A mutant (colored violet) [44], peptidyl substrate from cathepsin L-C25S mutant (colored cyan), the LLK peptide from complex_4_cocrys (colored pink) and the LLKVAL peptide from complex_3_soak (colored yellow) shown at the surfaces of A) cathepsin L-C25S, B) cathepsin V with no peptide bound to the primed region and C) cathepsin V with a peptide bound to the primed active site region. Peptide and protein residues are shown as sticks. Residues which are different in the active sites of cathepsin L and V are colored red and green. The blue colored residue represents the change after the peptide is bound to the primed region of cathepsin V-C25S. Different protein residues are labelled. The picture was prepared using the Open-source PyMOL™ v.1.7.2.1. (The PyMOL Molecular Graphics System, Version 1.7.2 Schrödinger, LLC).

The crystal structures complex_1_soak, complex_2_soak and complex_3_soak provide information regarding the specificity at primed substrate positions. The alanine

at P1' position is well ordered in the S1' subsite. The structures of the soaked complexes also confirm that the leucine binds well to the S1' subsite. These data are consistent with a comparative substrate specificity analysis of cathepsins L and V with fluorogenic peptides [138] which revealed a significant specificity of cathepsin V for valine and isoleucine residues and presumably high specificity for leucine at P2' position. However, the observed cathepsin V interactions are not entirely consistent with the proteomic data (Figure 1.3). The leucine at P1' and P2' positions is described as not preferable residue, while our structures clearly show leucine bound to these two subsites. This suggests that considering the specificity at the subsite binding sites independently may not be the most appropriate approach.

With the choice of peptide sequences we attempted to address a potential cooperativity effects between the side chain binding in the S1 and S2' subsites, by exploring the variances in the pattern large-small at this positions by pairs A3-S5, A3-K5, K3-A5, and R3-R5. The distance between Cβ atoms of residues at P1 and P2' is longer than 6 Å. The side chains of K3 and E3 at P1 from complex_4_cocrys and native complex, respectively, and V2 and L2 at P2' from complex_2_soak and complex_1_soak, respectively, are parallel with each other, with no noticeable interactions between each other. However, it seems that there is not enough space for large residues to bind to both positions simultaneously, since they may sterically hinder binding of the other residue as indicated by the unsuccessful attempts to determine the structure of the complex with AVRARL peptide with the R3-R5 pair.

Taken together, these structures now provide structural evidence for interactions of seven amino acid residues long peptides with the active site of a cysteine cathepsin, which is in agreement with work of Schechter and Berger back in 1968 [13] who studied degradation of polyalanine peptides in the depends of their length that only seven sites contribute to the kinetics of peptide degradation with papain. Insight in the geometry of binding in the S3' and S4' areas is an important extension of previous work as back in 1998 when the analysis of the substrate binding sites was revised the structural evidence for the existence of binding interactions beyond the S2' was lacking. In addition, these data strongly suggest that interactions between peptidyl substrates and a cysteine cathepsin are common to the whole group of cysteine cathepsins as well as papain-like proteases exhibiting endopeptidase activities.

Peptides bound only to the cathepsin V active site primed regions in the complex_1_soak, complex_2_soak and complex_3_soak structures and into the non-primed region as revealed by the native complex and the complex_4_cocrys structure. None of the structures revealed binding of a peptide along the whole length of the active site cleft with amino acid residues bridging the catalytic region. The answer to this conundrum was however unexpectedly provided by the cathepsin L C25S structure.

5.2 Unique Complex of Cathepsin L-C25S

The crystal structure of cathepsin L surprisingly provided insight into the proteolysis of a protein substrate by revealing the interaction between a cysteine cathepsin and its protein substrate, in which the N-terminal part of the cleaved peptide (from K103 to S105) remained bound to the active site cleft of the neighboring molecule in the crystal. To gain insight into the cleavage event, the structure of the cleaved region from our structure was superimposed with four other known cathepsin L structures: the structure of cathepsin L isolated from human kidney (khCL) [139]; the recombinant human cathepsin L structure (rhCL_1) [140] with the same sequence (T110A mutation) as that presented here, which is the full structure with exception of the catalytic site C25; the

recombinant human cathepsin L (rhCL_2) structure [44] with a sequence that enables the glycosylation of the N108 residue and imbeds the catalytic site C25A mutation; and the highest resolution recombinant human cathepsin L covalently linked with a nitrile inhibitor (rhCL_3) [141]. Our structure of cathepsin L C25S is superimposed with the structure of cathepsin L khCL, rhCL_1, rhCL_2 and rhCL_3, with a RMSD of the Cα-atoms of 0.57 Å, 0.4 Å, 0.57 Å and 1 Å, respectively (Table 5.1). The superimposition of structures showed that the regions of the cleaved peptide from K99 to A110 in the uncleaved molecules share a similar conformation and that the conformation and positioning of the residues K103, Y104 and S105 in the structure of our cathepsin L C25S mutant was clearly different from that observed in uncleaved molecules (Figure 5). The structures of the uncleaved molecules thus represent the structure before cleavage. Before cleavage, this region was folded as a short α-helix. Together, these comparisons demonstrate that the cleaved region must first be pulled out of the conformation observed in the uncleaved form into the conformation that allowed cleavage to occur. After cleavage took place, the cleaved molecules must have rotated approximately 90° away from the position in which they were cleaved to reach the position observed in the crystal. This movement is reminiscent of the change observed in the trypsin-antitrypsin complex, where after cleavage, the enzyme molecule is transferred to the other side of antitrypsin [142]. However, in the trypsin-antitrypsin structure, the cleaved peptide remained attached to the catalytic serine residue by a covalent bond, whereas in our case, this movement indicated a strong non-covalent binding affinity between the C-terminus of the cleaved peptide and the non-primed part of the active site cleft of cathepsin L.

Table 5.1: Data Calculated statistics of determined cathepsin L-C25S aligned with khCL, rhCL_1, rhCL_2 and rh_CL3. Q scoring - Quality of alignment which takes both the alignment length and RMSD into account. Cα RMSD – Root Mean Square Deviation, calculated between Cα-atoms. N-align - number of matched residues. Data were acquired using PDBeFold [143].

Model/PDB ID	Q scoring	Cα RMSD (Å)	N-align
khCL/1ICF	0.86	0.58	206
rhCL_1/3HHA	0.86	0.4	204
rhCL_2/3K24	0.86	0.57	206
rhCL_3/2XU3	0.88	1	209

The structure of cathepsin C25S mutant enabled us to address the question of whether the binding of small synthetic inhibitors [140] mimics the substrate binding in cysteine cathepsins. The positioning and orientation of the K103 to S105 region peptide is tightly superimposed with the main and side chains in the S1 and S2 binding sites (Figure 5.2), whereas in the S3 binding sites, due to the differences in chemical composition, the positioning of the ligands begins to diverge.

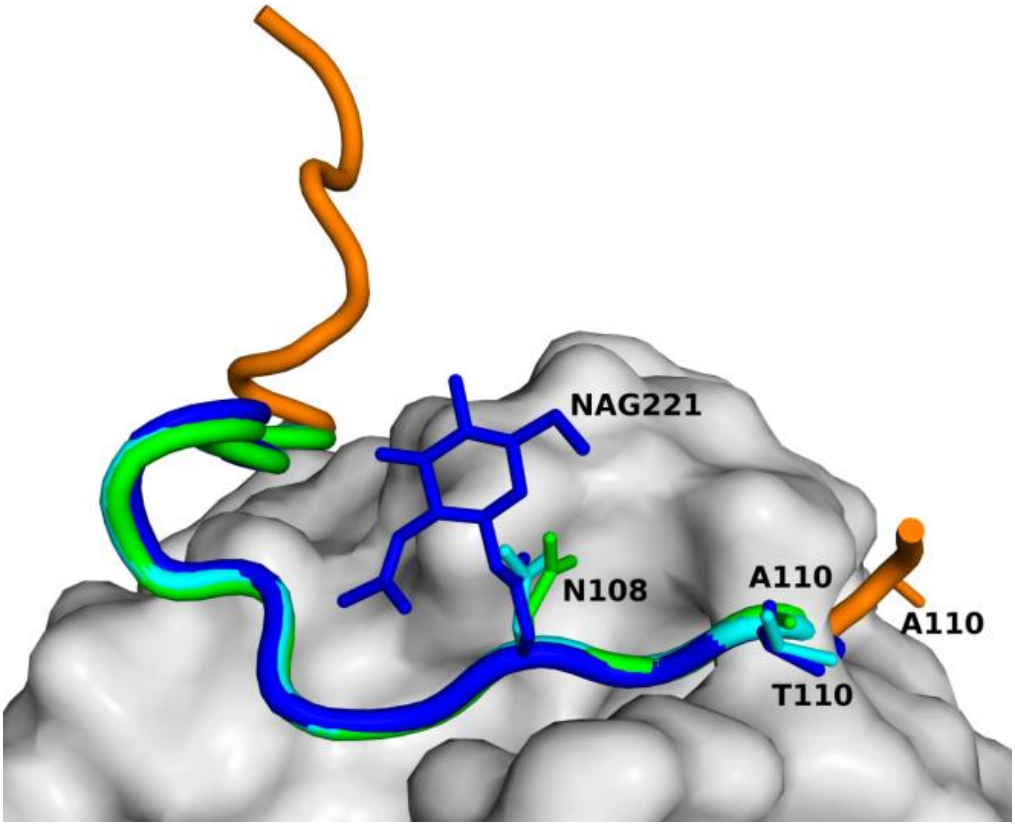


Figure 5.3: The cathepsin L-C25S cleaved region superimposed on the same region of the rhCL_1, rhCL_2 and rhCL_3 structures. The region of the cleaved peptide is located on the white surface of our cathepsin L-C25S structure. The protein main chains within the region are presented as ribbons. The regions of rhCL_1, rhCL_2, rhCL_3 and our structure are colored green, blue, cyan and orange, respectively. The glycosylated N108 and the NAG221 from rhCL_2 and mutated amino acids within the region are labeled and presented as sticks. The figure was prepared using the open source PyMOL™ v.1.7.2.1.

5.3 Cathepsin L C25S Mutant Partial Activity

Our cathepsin L molecule was cleaved, whereas in other cathepsin L structures cathepsin L remained intact. Prevented access to the active site by inhibitors including the propeptide explains the crystal structures of khCL complex with the p41 fragment [139] and the small molecule complexes rhCL_1 and rhCL_3 [140, 141] and the proenzyme structure [144]. However, among the remaining two other structures of cathepsin L, rhCL_2 and ours, ours was cleaved and the other one not. Both structures have the peptidyl ligand visible only in the non-primed region, which suggests that also the rhCL_2 C25A mutant likely retained low catalytic activity [44]. The other difference to look into was the cleaved region. Namely, the N108 is in the rhCL_2 structure glycosylated, whereas in ours the glycosylation was prevented by the T110A mutation, which suggests that glycosylation of N108 may be responsible for the difference in behavior. To address the biological significance of protective role of N108 glycosylation, we performed time dependent autoprocessing assay, which did not expose its relevance.

Nevertheless, this experiment does not exclude a possible protective role of the glycosylation of N108 against cleavage by other proteases.

When comparing our results with Adams-Cioaba's [44] QLA peptide (part of H319-33 substrate) bound to the human cathepsin L-C25A, similar conclusions can be drawn. In their work, they were not able to see any electron densities within the primed region nor did they identify cathepsin L-C25A partial activity. On the other side, their structure was obtained by soaking crystals of human cathepsin L-C25A with a substrate for only 30 min at 4°C, compared to our two weeks crystallization. We assume that the time length of soaking or cocrystallization does not exclude the possibility of partial activity of C25A mutants. It merits mention that we did not observe a peptide bound across the catalytic site from S3 to S2' binding sites in any of the cathepsin V-C25S structures. Therefore we are unable to confirm that 30-minute peptide soaking is enough for our peptides to be cleaved by cathepsin V. However, the observation of the cleavage and binding of the cleaved peptides only in the non-primed region suggests that both C25S and C25A mutants still exhibit low levels of catalytic activity which is not detectable with fluorogenic substrates.

5.4 Cathepsin K with Nitrile Inhibitors

Superimposition of determined cathepsin K complex structures revealed that the positions of the pair of reactive site Cys25 residues forming the covalent interaction Chloro-benzoyl and 79% of Fluoro-benzoyl and the pair that does not form it Bromo-benzoyl and 21% of Fluoro-benzoyl superimpose tightly and are distinct from each other (the catalytic sulfur atoms are positioned approximately 0.5 Å apart (Figure 5.4.). Due to the high resolution of the diffraction data we can conclude that these differences are real. In addition, the crystal structures of cathepsin K complexes showed similar way of binding to the active site as the OLB inhibitor. The superimposition of the nitrile inhibitors with the OLB inhibitor is shown in Figure 5.4.

The differences and similarities in the interaction patterns of Chloro-benzoyl, Bromo-benzoyl and Fluoro-benzoyl offer no explanation why Asp61 has been observed in two different conformations (Figure 4.11). It seems more likely that it is the change in the crystal lattice which leads to a different interaction pattern with the Asp61 side chain. Namely, the position of Asp61 is such that in the orthorhombic crystals it is tightly packed between two asymmetric molecules of cathepsin K, whereas in the 1monoclinic space group it is not.

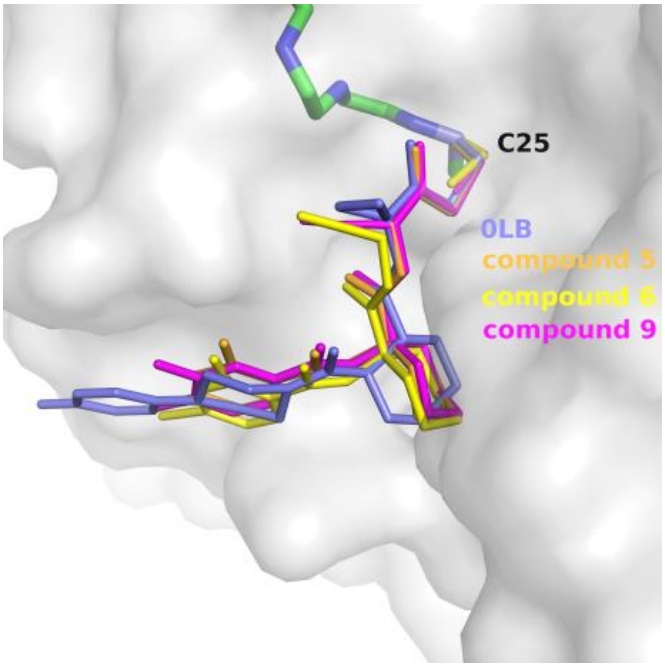


Figure 5.4: Superimposition of crystal structures of nitrile inhibitors on OLB inhibitor (PDB ID: 4DMX). Compound 5 is the N-(2-Amino-4-chlorobenzoyl)homocycloleucyl-glycinonitrile. Compound 6 is the N-(2-Amino-4-bromobenzoyl)homocycloleucyl-glycinonitrile. Compound 9 is the N-(3-Fluoro-4-aminobenzoyl)homocycloleucyl-glycinonitrile

Chapter 6

Conclusions

Our attempt to establish a system to study interactions between peptidyl substrates and cysteine endopeptidases by determining the crystal structures of presumably catalytically inactive mutants of cathepsins L and V resulted in unexpected novel outcomes. The crystal structures of cathepsin V C25S mutant in complex with three different peptides showed that peptides can bind at unexpected locations within the active site cleft. The binding specificity appears consistent with the specificity studies for cathepsin V [138, 145]. It is not only the length of the main chain which primarily defines the binding, but the amino acid sequence and time too.

Bound peptides enable us to gain insight beyond the defined subsite binding in the geometry of binding of peptides at positions of P3 to P4' and therefore completed quest for identifying the lacking interactions within the primed active site region [32]. Interestingly, the insight into peptide binding through the active site does not originate from a single peptide, but is a result of combined data from superimposed structures. The cathepsins V and L C25S mutant structures showed that when a peptide binds into the nonprimed region, it gets cleaved and the amino acid residues binding in the S3, S2 and S1 side remain bound after the cleavage as it was shown in cathepsin L complex with histone H3 peptide [44].

The insight into the geometry of binding in the primed region was however only made possible by the structures of the peptides which were not cleaved. The most intriguing are the crystal structures of complexes with the LLKVAL peptide obtained by soaking and cocrystallization. Whereas the soaking approach revealed binding into the primed region of the active site, the crystal structure of the complex obtained by cocrystallization revealed that given enough time the LLKVAL peptide was able to displace the presumably remaining part of the propeptide bound to the nonprimed region of the active site and get cleaved by preserving only the LLK part bound to the structure.

The undisputable evidence for the remaining catalytic activity of cathepsins' active site cysteine mutants came from the structure of cathepsin L C25S, which was cleaved in two parts during the two-week crystallization process. That work showed that the catalytic mutant only reduced the potency of the enzymes to a level that is not detectable by a fluorogenic substrate. The crystal structure caught the cathepsin L protease cathepsin L substrate complex "in the act" after cleavage and before the final separation of the two molecules. The evidence suggests that the C25S mutant and the C25A active site mutant can hydrolyze peptide bonds. In this respect, papain-like cysteine proteases behave similarly to serine proteases for which the presence of an oxyanion hole assures the low catalytic activity of the protein [146].

The crystal structures of cathepsin K with nitrile inhibitors show specific interactions of the inhibitors with the residues at the S subsites of cathepsin K and show that structurally similar nitrile inhibitors either do or do not form a covalent bond with the active site C25 residue. The different halides at the benzoyl group do not affect the inhibitor's binding at the S subsites of the cathepsin K, because in both cases the nitrile inhibitors are tightly bound to the S subsites of the cathepsin K active site. Therefore, the inhibitor specificity is most likely not dependent on the inhibitor's way of binding to the C25 active site residue.

References

[1] G. H. Reference. (2016, 01.03.2016). *Handbook: help me understand genetics [Online].* Available: https://ghr.nlm.nih.gov/handbook.pdf

[2] R. Olby, "The macromolecular concept and the origins of molecular biology," *J Chem Educ,* vol. 47, pp. 168-74, Mar 1970.

[3] H. Lodish, *Molecular Cell Biology*: W. H. Freeman, 2004.

[4] G. M. Cooper, *The Cell: A Molecular Approach*: ASM Press, 2000.

[5] A. Radzicka and R. Wolfenden, "A proficient enzyme," *Science,* vol. 267, pp. 90-3, Jan 6 1995.

[6] I. U. o. Biochemistry, M. B. N. Committee, and E. C. Webb, *Enzyme nomenclature 1992: recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes*: Published for the International Union of Biochemistry and Molecular Biology by Academic Press, 1992.

[7] N. D. Rawlings, M. Waller, A. J. Barrett, and A. Bateman, "MEROPS: the database of proteolytic enzymes, their substrates and inhibitors," *Nucleic Acids Research,* vol. 42, pp. D503-D509, Jan 2014.

[8] H. Neurath and K. A. Walsh, "Role of proteolytic enzymes in biological regulation (a review)," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 73, pp. 3825-3832, 1976.

[9] A. R. Khan and M. N. James, "Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes," *Protein Sci,* vol. 7, pp. 815-36, Apr 1998.

[10] T. J. Berg JM, Stryer L. (2002, 10.03.2016). *Biochemistry. 5th edition.* Available: http://www.ncbi.nlm.nih.gov/books/NBK22589/

[11] M. Drag and G. S. Salvesen, "Emerging principles in protease-based drug discovery," *Nat Rev Drug Discov,* vol. 9, pp. 690-701, Sep 2010.

[12] V. Turk, V. Stoka, O. Vasiljeva, M. Renko, T. Sun, B. Turk, *et al.*, "Cysteine cathepsins: from structure, function and regulation to new frontiers," *Biochim Biophys Acta,* vol. 1824, pp. 68-88, Jan 2012.

[13] I. Schechter and A. Berger, "On the size of the active site in proteases. I. Papain," *Biochem Biophys Res Commun,* vol. 27, pp. 157-62, Apr 20 1967.

[14] N. D. Rawlings and A. J. Barrett, "Evolutionary families of peptidases," *Biochem J,* vol. 290 (Pt 1), pp. 205-18, Feb 15 1993.

[15] L. C. Alves, R. L. Melo, S. J. Sanderson, J. C. Mottram, G. H. Coombs, G. Caliendo, *et al.*, "S1 subsite specificity of a recombinant cysteine proteinase, CPB, of Leishmania mexicana compared with cruzain, human cathepsin L and papain

using substrates containing non-natural basic amino acids," *Eur J Biochem,* vol. 268, pp. 1206-12, Mar 2001.

[16] S. Hasnain, T. Hirama, C. P. Huber, P. Mason, and J. S. Mort, "Characterization of cathepsin B specificity by site-directed mutagenesis. Importance of Glu245 in the S2-P2 specificity for arginine and its role in transition state stabilization," *J Biol Chem,* vol. 268, pp. 235-40, Jan 5 1993.

[17] J. C. Krupa, S. Hasnain, D. K. Nagler, R. Menard, and J. S. Mort, "S2' substrate specificity and the role of His110 and His111 in the exopeptidase activity of human cathepsin B," *Biochem J,* vol. 361, pp. 613-9, Feb 1 2002.

[18] F. Lecaille, E. Authie, T. Moreau, C. Serveau, F. Gauthier, and G. Lalmanach, "Subsite specificity of trypanosomal cathepsin L-like cysteine proteases. Probing the S2 pocket with phenylalanine-derived amino acids," *Eur J Biochem,* vol. 268, pp. 2733-41, May 2001.

[19] R. Menard, E. Carmona, C. Plouffe, D. Bromme, Y. Konishi, J. Lefebvre*, et al.*, "The specificity of the S1' subsite of cysteine proteases," *FEBS Lett,* vol. 328, pp. 107-10, Aug 9 1993.

[20] B. C. Bohrer, Y. F. Li, J. P. Reilly, D. E. Clemmer, R. D. DiMarchi, P. Radivojac*, et al.*, "Combinatorial Libraries of Synthetic Peptides as a Model for Shotgun Proteomics," *Analytical Chemistry,* vol. 82, pp. 6559-6568, 2010/08/01 2010.

[21] Y. Du, J. Zhou, J. Fan, Z. Shen, and X. Chen, "A streamline proteomic approach for characterizing protein-protein interaction network in a RAD52 protein complex," *Journal of proteome research,* vol. 8, pp. 2211-2217, 2009.

[22] D. N. Perkins, D. J. C. Pappin, D. M. Creasy, and J. S. Cottrell, "Probability-based protein identification by searching sequence databases using mass spectrometry data," *ELECTROPHORESIS,* vol. 20, pp. 3551-3567, 1999.

[23] O. Schilling and C. M. Overall, "Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites," *Nat Biotech,* vol. 26, pp. 685-694, 06//print 2008.

[24] I. G. Kamphuis, K. H. Kalk, M. B. Swarte, and J. Drenth, "Structure of papain refined at 1.65 A resolution," *J Mol Biol,* vol. 179, pp. 233-56, Oct 25 1984.

[25] B. Asboth and L. Polgar, "Transition-state stabilization at the oxyanion binding sites of serine and thiol proteinases: hydrolyses of thiono and oxygen esters," *Biochemistry,* vol. 22, pp. 117-122, 1983/01/01 1983.

[26] B. Turk, V. Turk, and D. Turk, "Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors," *Biol Chem,* vol. 378, pp. 141-50, Mar-Apr 1997.

[27] V. Turk, B. Turk, and D. Turk, "Lysosomal cysteine proteases: facts and opportunities," *EMBO J,* vol. 20, pp. 4629-33, Sep 3 2001.

[28] J. Drenth, K. H. Kalk, and H. M. Swen, "Binding of chloromethyl ketone substrate analogues to crystalline papain," *Biochemistry,* vol. 15, pp. 3731-8, Aug 24 1976.

[29] E. N. Baker, "Structure of actinidin, after refinement at 1.7 A resolution," *J Mol Biol,* vol. 141, pp. 441-84, Aug 25 1980.

[30] W. M. Wolf, J. Bajorath, A. Muller, S. Raghunathan, T. P. Singh, W. Hinrichs*, et al.*, "Inhibition of proteinase K by methoxysuccinyl-Ala-Ala-Pro-Ala-chloromethyl ketone. An x-ray study at 2.2-A resolution," *J Biol Chem,* vol. 266, pp. 17695-9, Sep 15 1991.

[31] D. Turk, M. Podobnik, T. Popovic, N. Katunuma, W. Bode, R. Huber*, et al.*, "Crystal structure of cathepsin B inhibited with CA030 at 2.0-A resolution: A basis for the design of specific epoxysuccinyl inhibitors," *Biochemistry,* vol. 34, pp. 4791-7, Apr 11 1995.

[32] D. Turk, G. Guncar, M. Podobnik, and B. Turk, "Revised definition of substrate binding sites of papain-like cysteine proteases," *Biol Chem,* vol. 379, pp. 137-47, Feb 1998.

[33] G. Lowe and Y. Yuthavong, "Kinetic specificity in papain-catalysed hydrolyses," *Biochem J,* vol. 124, pp. 107-15, Aug 1971.

[34] H. Kirschke, A. J. Barrett, and N. D. Rawlings, "Proteinases 1: lysosomal cysteine proteinases," *Protein Profile,* vol. 2, pp. 1581-643, 1995.

[35] B. Asboth, Z. Majer, and L. Polgar, "Cysteine proteases: the S2P2 hydrogen bond is more important for catalysis than is the analogous S1P1 bond," *FEBS Lett,* vol. 233, pp. 339-41, Jun 20 1988.

[36] C. Garcia-Echeverria and D. H. Rich, "Effect of P2' substituents on kinetic constants for hydrolysis by cysteine proteinases," *Biochem Biophys Res Commun,* vol. 187, pp. 615-9, Sep 16 1992.

[37] F. C. Portaro, A. B. Santos, M. H. Cezari, M. A. Juliano, L. Juliano, and E. Carmona, "Probing the specificity of cysteine proteinases at subsites remote from the active site: analysis of P4, P3, P2' and P3' variations in extended substrates," *Biochem J,* vol. 347 Pt 1, pp. 123-9, Apr 1 2000.

[38] W. Bode and R. Huber, "Natural protein proteinase inhibitors and their interaction with proteinases," *European Journal of Biochemistry,* vol. 204, pp. 433-451, 1992.

[39] A. Wlodawer and A. Gustchina, "Structural and biochemical studies of retroviral proteases," *Biochim Biophys Acta,* vol. 1477, pp. 16-34, Mar 7 2000.

[40] D. Musil, D. Zucic, D. Turk, R. A. Engh, I. Mayr, R. Huber*, et al.*, "The refined 2.15 A X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity," *EMBO J,* vol. 10, pp. 2321-30, Sep 1991.

[41] D. Turk, V. Janjic, I. Stern, M. Podobnik, D. Lamba, S. W. Dahl*, et al.*, "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,* vol. 20, pp. 6570-82, Dec 3 2001.

[42] G. Guncar, M. Podobnik, J. Pungercar, B. Strukelj, V. Turk, and D. Turk, "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,* vol. 6, pp. 51-61, Jan 15 1998.

[43] G. Guncar, I. Klemencic, B. Turk, V. Turk, A. Karaoglanovic-Carmona, L. Juliano*, et al.*, "Crystal structure of cathepsin X: a flip-flop of the ring of His23

allows carboxy-monopeptidase and carboxy-dipeptidase activity of the protease," *Structure,* vol. 8, pp. 305-13, Mar 15 2000.

[44] M. A. Adams-Cioaba, J. C. Krupa, C. Xu, J. S. Mort, and J. Min, "Structural basis for the recognition and cleavage of histone H3 by cathepsin L," *Nat Commun,* vol. 2, p. 197, 2011.

[45] D. Bromme, K. Okamoto, B. B. Wang, and S. Biroc, "Human cathepsin O2, a matrix protein-degrading cysteine protease expressed in osteoclasts. Functional expression of human cathepsin O2 in Spodoptera frugiperda and characterization of the enzyme," *J Biol Chem,* vol. 271, pp. 2126-32, Jan 26 1996.

[46] Z. Li, Y. Yasuda, W. Li, M. Bogyo, N. Katz, R. E. Gordon*, et al.*, "Regulation of collagenase activities of human cathepsins by glycosaminoglycans," *J Biol Chem,* vol. 279, pp. 5470-9, Feb 13 2004.

[47] D. Turk, B. Turk, and V. Turk, "Papain-like lysosomal cysteine proteases and their inhibitors: drug discovery targets?," *Biochem Soc Symp,* pp. 15-30, 2003.

[48] Y. Yasuda, Z. Li, D. Greenbaum, M. Bogyo, E. Weber, and D. Bromme, "Cathepsin V, a novel and potent elastolytic activity expressed in activated macrophages," *J Biol Chem,* vol. 279, pp. 36761-70, Aug 27 2004.

[49] R. Itoh, S. Kawamoto, W. Adachi, S. Kinoshita, and K. Okubo, "Genomic organization and chromosomal localization of the human cathepsin L2 gene," *DNA Res,* vol. 6, pp. 137-40, Apr 30 1999.

[50] D. Bromme, Z. Li, M. Barnes, and E. Mehler, "Human cathepsin V functional expression, tissue distribution, electrostatic surface potential, enzymatic characterization, and chromosomal localization," *Biochemistry,* vol. 38, pp. 2377-85, Feb 23 1999.

[51] I. Santamaria, G. Velasco, M. Cazorla, A. Fueyo, E. Campo, and C. Lopez-Otin, "Cathepsin L2, a novel human cysteine proteinase produced by breast and colorectal carcinomas," *Cancer Res,* vol. 58, pp. 1624-30, Apr 15 1998.

[52] C. H. Tang, J. W. Lee, M. G. Galvez, L. Robillard, S. E. Mole, and H. A. Chapman, "Murine cathepsin F deficiency causes neuronal lipofuscinosis and late-onset neurological disease," *Mol Cell Biol,* vol. 26, pp. 2309-16, Mar 2006.

[53] B. Goulet, A. Baruch, N. S. Moon, M. Poirier, L. L. Sansregret, A. Erickson*, et al.*, "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,* vol. 14, pp. 207-19, Apr 23 2004.

[54] A. H. Aguda, P. Panwar, X. Du, N. T. Nguyen, G. D. Brayer, and D. Bromme, "Structural basis of collagen fiber degradation by cathepsin K," *Proc Natl Acad Sci U S A,* vol. 111, pp. 17474-9, Dec 9 2014.

[55] D. Bromme and F. Lecaille, "Cathepsin K inhibitors for osteoporosis and potential off-target effects," *Expert Opin Investig Drugs,* vol. 18, pp. 585-600, May 2009.

[56] X. W. Cheng, G. P. Shi, M. Kuzuya, T. Sasaki, K. Okumura, and T. Murohara, "Role for cysteine protease cathepsins in heart disease: focus on biology and mechanisms with clinical implication," *Circulation,* vol. 125, pp. 1551-62, Mar 27 2012.

[57] Y. Hua and S. Nair, "Proteases in cardiometabolic diseases: Pathophysiology, molecular mechanisms and clinical applications," *Biochim Biophys Acta,* vol. 1852, pp. 195-208, Feb 2015.

[58] Y. Hua, Y. Zhang, J. Dolence, G. P. Shi, J. Ren, and S. Nair, "Cathepsin K knockout mitigates high-fat diet-induced cardiac hypertrophy and contractile dysfunction," *Diabetes,* vol. 62, pp. 498-509, Feb 2013.

[59] S. Nagase, M. Ohyama, Y. Hashimoto, M. Small, J. Sharpe, J. Manako*, et al.*, "Bone turnover markers and pharmacokinetics of a new sustained-release formulation of the cathepsin K inhibitor, ONO-5334, in healthy post-menopausal women," *J Bone Miner Metab,* vol. 33, pp. 93-100, Jan 2015.

[60] C. Vassalle and G. Iervasi, "Cathepsin K--a classical bone biomarker in cardiovascular disease: the heart is not alone anymore," *Atherosclerosis,* vol. 228, pp. 36-7, May 2013.

[61] K. Brixen, R. Chapurlat, A. M. Cheung, T. M. Keaveny, T. Fuerst, K. Engelke*, et al.*, "Bone density, turnover, and estimated strength in postmenopausal women treated with odanacatib: a randomized trial," *J Clin Endocrinol Metab,* vol. 98, pp. 571-80, Feb 2013.

[62] R. D. Chapurlat, "Treatment of postmenopausal osteoporosis with odanacatib," *Expert Opin Pharmacother,* vol. 15, pp. 559-64, Mar 2014.

[63] L. T. Duong, G. A. Wesolowski, P. Leung, R. Oballa, and M. Pickarski, "Efficacy of a cathepsin K inhibitor in a preclinical model for prevention and treatment of breast cancer bone metastasis," *Mol Cancer Ther,* vol. 13, pp. 2898-909, Dec 2014.

[64] K. Engelke, S. Nagase, T. Fuerst, M. Small, T. Kuwayama, S. Deacon*, et al.*, "The effect of the cathepsin K inhibitor ONO-5334 on trabecular and cortical bone in postmenopausal osteoporosis: the OCEAN study," *J Bone Miner Res,* vol. 29, pp. 629-38, Mar 2014.

[65] S. Nagase, Y. Hashimoto, M. Small, M. Ohyama, T. Kuwayama, and S. Deacon, "Serum and urine bone resorption markers and pharmacokinetics of the cathepsin K inhibitor ONO-5334 after ascending single doses in post menopausal women," *Br J Clin Pharmacol,* vol. 74, pp. 959-70, Dec 2012.

[66] T. Nakamura, M. Shiraki, M. Fukunaga, T. Tomomitsu, A. C. Santora, R. Tsai*, et al.*, "Effect of the cathepsin K inhibitor odanacatib administered once weekly on bone mineral density in Japanese patients with osteoporosis--a double-blind, randomized, dose-finding study," *Osteoporos Int,* vol. 25, pp. 367-76, Jan 2014.

[67] K. W. Ng, "Potential role of odanacatib in the treatment of osteoporosis," *Clin Interv Aging,* vol. 7, pp. 235-47, 2012.

[68] Y. Ochi, H. Yamada, H. Mori, Y. Nakanishi, S. Nishikawa, R. Kayasuga*, et al.*, "Effects of ONO-5334, a novel orally-active inhibitor of cathepsin K, on bone metabolism," *Bone,* vol. 49, pp. 1351-6, Dec 2011.

[69] S. A. Stoch, S. Zajic, J. A. Stone, D. L. Miller, L. van Bortel, K. C. Lasseter*, et al.*, "Odanacatib, a selective cathepsin K inhibitor to treat osteoporosis: safety, tolerability, pharmacokinetics and pharmacodynamics--results from single oral

dose studies in healthy volunteers," *Br J Clin Pharmacol,* vol. 75, pp. 1240-54, May 2013.

[70] T. M. Runger, S. Adami, C. L. Benhamou, E. Czerwinski, J. Farrerons, D. L. Kendler*, et al.*, "Morphea-like skin reactions in patients treated with the cathepsin K inhibitor balicatib," *J Am Acad Dermatol,* vol. 66, pp. e89-96, Mar 2012.

[71] J. Borisek, M. Vizovisek, P. Sosnowski, B. Turk, D. Turk, B. Mohar*, et al.*, "Development of N-(Functionalized benzoyl)-homocycloleucyl-glycinonitriles as Potent Cathepsin K Inhibitors," *J Med Chem,* vol. 58, pp. 6928-37, Sep 10 2015.

[72] M. Monti, S. Orru, D. Pagnozzi, and P. Pucci, "Functional proteomics," *Clin Chim Acta,* vol. 357, pp. 140-50, Jul 24 2005.

[73] M. Monti, S. Orru, D. Pagnozzi, and P. Pucci, "Interaction proteomics," *Biosci Rep,* vol. 25, pp. 45-56, Feb-Apr 2005.

[74] Y. Perez-Riverol, R. Wang, H. Hermjakob, M. Muller, V. Vesada, and J. A. Vizcaino, "Open source libraries and frameworks for mass spectrometry based proteomics: a developer's perspective," *Biochim Biophys Acta,* vol. 1844, pp. 63-76, Jan 2014.

[75] J. Song, H. Tan, A. J. Perry, T. Akutsu, G. I. Webb, J. C. Whisstock*, et al.*, "PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites," *PLoS One,* vol. 7, p. e50300, 2012.

[76] (13.3.2016). *Peptide libraries.* Available: http://www.pbcpeptide.com/Peptide%20Library.htm

[77] O. Schilling, P. F. Huesgen, O. Barre, U. auf dem Keller, and C. M. Overall, "Characterization of the prime and non-prime active site specificities of proteases by proteome-derived peptide libraries and tandem mass spectrometry," *Nat. Protocols,* vol. 6, pp. 111-120, 01//print 2011.

[78] K. Sreekrishna, R. H. Potenz, J. A. Cruze, W. R. McCombie, K. A. Parker, L. Nelles*, et al.*, "High level expression of heterologous proteins in methylotrophic yeast Pichia pastoris," *J Basic Microbiol,* vol. 28, pp. 265-78, 1988.

[79] J. M. Cregg, K. R. Madden, K. J. Barringer, G. P. Thill, and C. A. Stillman, "Functional-Characterization of the 2 Alcohol Oxidase Genes from the Yeast Pichia-Pastoris," *Molecular and Cellular Biology,* vol. 9, pp. 1316-1323, Mar 1989.

[80] L. Technologies, *Pichia Expression Kit*: Life Technologies, 2014.

[81] H. Dobeli and F. Hochuli, "Metal chelate resins," ed: Google Patents, 1991.

[82] G. H. Lathe and C. R. Ruthven, "The separation of substances on the basis of their molecular weights, using columns of starch and water," *Biochem J,* vol. 60, p. xxxiv, Aug 1955.

[83] H. Small, *Ion Chromatography*: Springer, 1989.

[84] G. Healthcare, *Ion Exchange Chromatography: Principles and Methods*. Uppsala: eneral Electric Company, 2016.

[85] B. Turk, D. Turk, and V. Turk, "Lysosomal cysteine proteases: more than scavengers," *Biochim Biophys Acta,* vol. 1477, pp. 98-111, Mar 7 2000.

[86] M. Mihelic, A. Dobersek, G. Guncar, and D. Turk, "Inhibitory fragment from the p41 form of invariant chain can regulate activity of cysteine cathepsins in antigen presentation," *J Biol Chem,* vol. 283, pp. 14453-60, May 23 2008.

[87] F. H. Niesen, H. Berglund, and M. Vedadi, "The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability," *Nat Protoc,* vol. 2, pp. 2212-21, 2007.

[88] N. J. Greenfield, "Using circular dichroism spectra to estimate protein secondary structure," *Nat. Protocols,* vol. 1, pp. 2876-2890, 01//print 2007.

[89] U. B. Ericsson, B. M. Hallberg, G. T. Detitta, N. Dekker, and P. Nordlund, "Thermofluor-based high-throughput stability optimization of proteins for structural studies," *Anal Biochem,* vol. 357, pp. 289-98, Oct 15 2006.

[90] K. Phillips and A. H. de la Pena, "The combined use of the Thermofluor assay and ThermoQ analytical software for the determination of protein stability and buffer optimization as an aid in protein crystallization," *Curr Protoc Mol Biol,* vol. Chapter 10, p. Unit10 28, Apr 2011.

[91] N. E. Chayen, "Methods for separating nucleation and growth in protein crystallisation," *Prog Biophys Mol Biol,* vol. 88, pp. 329-37, Jul 2005.

[92] N. Asherie, "Protein crystallization and phase diagrams," *Methods,* vol. 34, pp. 266-72, Nov 2004.

[93] N. E. Chayen and E. Saridakis, "Protein crystallization: from purified protein to diffraction-quality crystal," *Nat Methods,* vol. 5, pp. 147-53, Feb 2008.

[94] N. E. Chayen, "Comparative studies of protein crystallization by vapour-diffusion and microbatch techniques," *Acta Crystallogr D Biol Crystallogr,* vol. 54, pp. 8-15, Jan 1 1998.

[95] E. L. Forsythe, D. L. Maxwell, and M. Pusey, "Vapor diffusion, nucleation rates and the reservoir to crystallization volume ratio," *Acta Crystallogr D Biol Crystallogr,* vol. 58, pp. 1601-5, Oct 2002.

[96] A. McPherson, "Current approaches to macromolecular crystallization," *Eur J Biochem,* vol. 189, pp. 1-23, Apr 20 1990.

[97] (2016, 26.02.2016). *Vapour diffusion crystallization overview, [Online].* Available: http://people.mbi.ucla.edu/sawaya/m230d/Crystallization/tipsandtricks.html

[98] A. V. Yadav, A. S. Shete, A. P. Dabke, P. V. Kulkarni, and S. S. Sakhare, "Co-Crystals: A Novel Approach to Modify Physicochemical Properties of Active Pharmaceutical Ingredients," *Indian Journal of Pharmaceutical Sciences,* vol. 71, pp. 359-370, Jul-Aug

04/15/received

08/01/revised

08/03/accepted 2009.

[99] A. M. Hassell, G. An, R. K. Bledsoe, J. M. Bynum, H. L. Carter, S.-J. J. Deng, *et al.*, "Crystallization of protein–ligand complexes," *Acta Crystallographica Section D: Biological Crystallography,* vol. 63, pp. 72-79, 12/13

04/04/received

11/07/accepted 2007.

[100] A. M. Hassell, G. An, R. K. Bledsoe, J. M. Bynum, H. L. Carter, 3rd, S. J. Deng, *et al.*, "Crystallization of protein-ligand complexes," *Acta Crystallogr D Biol Crystallogr,* vol. 63, pp. 72-9, Jan 2007.

[101] G. Taylor, "The phase problem," *Acta Crystallogr D Biol Crystallogr,* vol. 59, pp. 1881-90, Nov 2003.

[102] (2016, 26.02.2016). *Representation of several Miller indices h, k, l, [Online].* Available: http://81.161.252.57/ipci/courses/technology/index_59.htm.

[103] *International tables for crystallography* vol. F. Dordrecht/Boston/London: Kluwer Academic Publishers, 2001.

[104] (2016, 26.02.2016). *Representation of X-ray diffraction and Bragg's law, [Online].* Available: http://l-esperimento-piu-bello-della-fisica.bo.imm.cnr.it/english/history/figuredett1.html#bragg

[105] R. J. Read. (2009, 14.3.2016). *Fourier transformants: structure factors, phases and electron density.* Available: http://www-structmed.cimr.cam.ac.uk/Course/Fourier/Fourier.html

[106] K. Cowtan. (2014, 14.3.2016). *Kevin Cowtan's Picture Book of Fourier Transforms.* Available: http://www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html

[107] T. G. Battye, L. Kontogiannis, O. Johnson, H. R. Powell, and A. G. Leslie, "iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM," *Acta Crystallogr D Biol Crystallogr,* vol. 67, pp. 271-81, Apr 2011.

[108] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, *et al.*, "Overview of the CCP4 suite and current developments," *Acta Crystallogr D Biol Crystallogr,* vol. 67, pp. 235-42, Apr 2011.

[109] M. Krug, M. S. Weiss, U. Heinemann, and U. Mueller, "XDSAPP: a graphical user interface for the convenient processing of diffraction data using XDS," *Journal of Applied Crystallography,* vol. 45, pp. 568-572, Jun 2012.

[110] Z. Otwinowski and W. Minor, "Processing of X-ray diffraction data collected in oscillation mode," *Macromolecular Crystallography, Pt A,* vol. 276, pp. 307-326, 1997.

[111] F. Fish. (2009, 14.3.2016). *Phase determination.* Available: http://my.yetnet.ch/dergutemensch/crystallography/phasedetermination.htm

[112] J. Bella and M. G. Rossmann, "A general phasing algorithm for multiple MAD and MIR data," *Acta Crystallogr D Biol Crystallogr,* vol. 54, pp. 159-74, Mar 1 1998.

[113] Z. Dauter, M. Dauter, and E. Dodson, "Jolly SAD," *Acta Crystallogr D Biol Crystallogr,* vol. 58, pp. 494-506, Mar 2002.

[114] C. M. Ogata, "MAD phasing grows up," *Nat Struct Biol,* vol. 5 Suppl, pp. 638-40, Aug 1998.

[115] A. Pahler, J. L. Smith, and W. A. Hendrickson, "A probability representation for phase information from multiwavelength anomalous dispersion," *Acta Crystallogr A,* vol. 46 (Pt 7), pp. 537-40, Jul 1 1990.

[116] M. G. Rossmann, "The molecular replacement method," *Acta Crystallogr A,* vol. 46 (Pt 2), pp. 73-82, Feb 1 1990.

[117] T. C. Terwilliger, "Reciprocal-space solvent flattening," *Acta Crystallographica Section D: Biological Crystallography,* vol. 55, pp. 1863-1871, 11/01
05/17/received
07/22/accepted 1999.

[118] Y. P. Nieh and K. Y. Zhang, "A two-dimensional histogram-matching method for protein phase refinement and extension," *Acta Crystallogr D Biol Crystallogr,* vol. 55, pp. 1893-900, Nov 1999.

[119] T. C. Terwilliger, "Finding non-crystallographic symmetry in density maps of macromolecular structures," *Journal of Structural and Functional Genomics,* vol. 14, pp. 91-95, 07/24
04/10/received
07/15/accepted 2013.

[120] P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols*, et al.*, "PHENIX: a comprehensive Python-based system for macromolecular structure solution," *Acta Crystallogr D Biol Crystallogr,* vol. 66, pp. 213-21, Feb 2010.

[121] G. G. Langer, S. X. Cohen, V. S. Lamzin, and A. Perrakis, "Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7," *Nature protocols,* vol. 3, pp. 1171-1179, 2008.

[122] K. Cowtan, "The Buccaneer software for automated model building. 1. Tracing protein chains," *Acta Crystallogr D Biol Crystallogr,* vol. 62, pp. 1002-11, Sep 2006.

[123] D. Turk, "MAIN software for density averaging, model building, structure refinement and validation," *Acta Crystallogr D Biol Crystallogr,* vol. 69, pp. 1342-57, Aug 2013.

[124] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, "Features and development of Coot," *Acta Crystallogr D Biol Crystallogr,* vol. 66, pp. 486-501, Apr 2010.

[125] A. Wlodawer, W. Minor, Z. Dauter, and M. Jaskolski, "Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures," *FEBS J,* vol. 275, pp. 1-21, Jan 2008.

[126] R. J. Read. (2009, 10.03.2016). *Likelihood; theory and application to structure refinement.* Available: http://www-structmed.cimr.cam.ac.uk/Course/Likelihood/likelihood.html

[127] A. T. Brunger, "Free R value: a novel statistical quantity for assessing the accuracy of crystal structures," *Nature,* vol. 355, pp. 472-5, Jan 30 1992.

[128] J. Praznikar and D. Turk, "Free kick instead of cross-validation in maximum-likelihood refinement of macromolecular crystal structures," *Acta Crystallogr D Biol Crystallogr,* vol. 70, pp. 3124-34, Dec 1 2014.

[129] K. Cowtan. (2008, 10.03.2016). *Atomic displacement parameter.* Available: http://ccp4wiki.org/~ccp4wiki/wiki/index.php?title=Temperature_factor

[130] U. Mueller, N. Darowski, M. R. Fuchs, R. Forster, M. Hellmig, K. S. Paithankar*, et al.*, "Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin," *J Synchrotron Radiat,* vol. 19, pp. 442-9, May 2012.

[131] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read, "Phaser crystallographic software," *J Appl Crystallogr,* vol. 40, pp. 658-674, Aug 1 2007.

[132] G. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls*, et al.*, "REFMAC5 for the refinement of macromolecular crystal structures," *Acta Crystallogr D Biol Crystallogr,* vol. 67, pp. 355-67, Apr 2011.

[133] M. Renko, J. Sabotic, M. Mihelic, J. Brzin, J. Kos, and D. Turk, "Versatile loops in mycocypins inhibit three protease families," *J Biol Chem,* vol. 285, pp. 308-16, Jan 1 2010.

[134] K. A. Kantardjieff and B. Rupp, "Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals," *Protein Sci,* vol. 12, pp. 1865-71, Sep 2003.

[135] A. G. Dossetter, H. Beeley, J. Bowyer, C. R. Cook, J. J. Crawford, J. E. Finlayson*, et al.*, "(1R,2R)-N-(1-cyanocyclopropyl)-2-(6-methoxy-1,3,4,5-tetrahydropyrido[4,3-b]indole -2-carbonyl)cyclohexanecarboxamide (AZD4996): a potent and highly selective cathepsin K inhibitor for the treatment of osteoarthritis," *J Med Chem,* vol. 55, pp. 6363-74, Jul 26 2012.

[136] J. R. Somoza, H. Zhan, K. K. Bowman, L. Yu, K. D. Mortara, J. T. Palmer*, et al.*, "Crystal structure of human cathepsin V," *Biochemistry,* vol. 39, pp. 12543-51, Oct 17 2000.

[137] M. Renko, Turk, D. (29.02.2016). *Unreduced cathepsin V in complex with stefin A [Online].* Available: http://www.rcsb.org/pdb/explore.do?structureId=3KFQ

[138] L. Puzer, S. S. Cotrin, M. F. Alves, T. Egborge, M. S. Araujo, M. A. Juliano*, et al.*, "Comparative substrate specificity analysis of recombinant human cathepsin V and cathepsin L," *Arch Biochem Biophys,* vol. 430, pp. 274-83, Oct 15 2004.

[139] G. Guncar, G. Pungercic, I. Klemencic, V. Turk, and D. Turk, "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,* vol. 18, pp. 793-803, Feb 15 1999.

[140] N. Asaad, P. A. Bethel, M. D. Coulson, J. E. Dawson, S. J. Ford, S. Gerhardt*, et al.*, "Dipeptidyl nitrile inhibitors of Cathepsin L," *Bioorg Med Chem Lett,* vol. 19, pp. 4280-3, Aug 1 2009.

[141] L. A. Hardegger, B. Kuhn, B. Spinnler, L. Anselm, R. Ecabert, M. Stihle*, et al.*, "Systematic investigation of halogen bonding in protein-ligand interactions," *Angew Chem Int Ed Engl,* vol. 50, pp. 314-8, Jan 3 2011.

[142] J. A. Huntington, R. J. Read, and R. W. Carrell, "Structure of a serpin-protease complex shows inhibition by deformation," *Nature,* vol. 407, pp. 923-6, Oct 19 2000.

[143] E. B. Krissinel, M. D. Winn, C. C. Ballard, A. W. Ashton, P. Patel, E. A. Potterton*, et al.*, "The new CCP4 Coordinate Library as a toolkit for the design of coordinate-related applications in protein crystallography," *Acta Crystallogr D Biol Crystallogr,* vol. 60, pp. 2250-5, Dec 2004.

[144] R. Coulombe, P. Grochulski, J. Sivaraman, R. Menard, J. S. Mort, and M. Cygler, "Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment," *EMBO J,* vol. 15, pp. 5492-503, Oct 15 1996.

[145] Y. Choe, F. Leonetti, D. C. Greenbaum, F. Lecaille, M. Bogyo, D. Bromme*, et al.*, "Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities," *Journal of Biological Chemistry,* vol. 281, pp. 12824-12832, May 5 2006.

[146] E. Krissinel, "On the relationship between sequence and structure similarities in proteomics," *Bioinformatics,* vol. 23, pp. 717-723, Mar 15 2007.

Bibliography

Publications Related to the Thesis

Sosnowski, P. and Turk, D., Caught in the Act: The Crystal Structure of cleaved Cathepsin L bound to the active site of Cathepsin L, 2016, FEBS Lett. Accepted Author Manuscript. doi:10.1002/1873-3468.12140

Borisek, J., Vizovisek, M., Sosnowski, P., Turk, B., Turk, D., Novic, M., Development of N-(Functionalized benzoyl)-homocycloleucyl-glycinonitriles as Potent Cathepsin K inhibitors, 2015, J Med Chem, 58(17): 6928-37, doi: 10.1021/acs.jmedchem.5b00746

Other Publications not related to thesis

Zurga, S., Pohleven, J., Renko, M., Bleuler-Martinez, S., Sosnowski, P., Turk, D., Kunzler, M., Kos, J., Sabotic, J., A novel beta-trefoil lectin from the parasol mushroom (Macrolepiota procera) is nematoxic, 2014, FEBS Journal, 281(15): 3489-3506, doi: 10.1111/febs.12875

Appendix A

Materials

A.1 Culture Media

10X D	···	20% Dextrose
10X GY	···	10% Glycerol
10X M	···	5% Methanol
10X YNB	···	13.4% Yeast Nitrogen Base with Ammonium Sulfate without amino acids
500X B	···	0.02% Biotin
liquid LB (lysogeny broth)	···	10 g casein
culture media	···	5 g yeast extract
	···	10 g NaCl
	···	until 1 L dH_2O
liquid LBA (lysogeny broth	···	LB media
with ampicillin) culture media	···	ampicillin (final concentration 0.1 mM/L)
liquid YPD (yeast extract	···	10 g yeast extract
peptone dextrose media	···	20 g peptone
	···	10 g yeast extract
	···	100 ml of 10X D
	···	until 1 L dH_2O
liquid BMGY (buffered	···	10 g yeast extract
glycerol-complex) media	···	20 g peptone
	···	100 mL of 1 M potassium phosphate, pH 6.0
	···	2 ml of 500X B

	...	100 of ml 10X GY
	...	until 1 L dH_2O
liquid BMMY (buffered	...	10 g yeast extract
methanol-complex) media	...	100 mL of 1 M potassium phosphate, pH 6.0
	...	2 ml of 500X B
	...	100 ml of 10X M
	...	until 1 L dH_2O
LBA (lysogeny broth with	...	LBA media
ampicillin) plates	...	20 g agar
MD (minimal dextrose)	...	1.34% YNB
plates	...	4 x 10^{-5}% biotin
	...	2% dextrose
	...	15 g agar
YPD (yeast extract peptone	...	YPD media
dextrose) plates	...	20 g agar

A.1.1 Purification buffers for affinity chromatography and activation buffer

AKTA Xpress washing buffer	...	30 mM TRIS, pH 7.5
	...	0.4 M NaCl
	...	10 mM Imidazole
AKTA Xpress elution buffer	...	30 mM TRIS, pH 7.5
	...	0.4 M NaCl
	...	0.4 M Imidazole
AKTA Xpress size exclusion	...	30 mM TRIS, pH 7.5
chromatography (SEC) buffer	...	0.4 M NaCl
Activation buffer	...	100 mM sodium acetate, pH 5.0
	...	1 mM EDTA
	...	5 mM DTT

A.1.2 Other chemicals

acetone	Serva
agar	Sigma
acetic acid	Sigma Alrdich
acrilamid/N,N-methylenbisakrilamid	Serva
ammonium nitrate	Fluka
ammonium sulfate	Merck
antifoam 204	Sigma
casein	Sigma
coomassie brilliant blue R250	Fluka
crystallization screens	Qiagen, Rigaku Reagents, Hampton Research
DMSO (dimethyl sulfoxide)	Merck
DTT (dithiothreitol)	Sigma
EDTA (ethylenediaminetetraacetic acid)	Serva
glycerol	AppliChem
HCl (hydrochloric acid)	Merck
low molecular weight protein ladder	Amersham Biosciences
methanol 100 %	Emshure
MPD (2-Methyl-2,4-pentanediol)	Sigma
NaOH (sodium hydroxide)	Merck
Page Ruler Prestained Protein Ladder	Fermentas
peptone	Sigma
potassium phosphate	Fluka
sodium acetate	Merck
sodium phosphate	Merck
SYPRO Orange Dye (5000x in DMSO)	Life Technologies
TEMED (tetramethylethylenediamine)	Serva
Tris (tris(hydroxymethyl)aminomethane)	Serva
Tris/HCl	Serva
yeast extract	Sigma

A.1.3 Laboratory equipment

-20°C freezer	Gorenje
-80°C freezer	Sanyo
Akta Xpress	GE Healthcare

Akta Xplorer	GE Healthcare
bioreactor Labfors	InFors HT
Gene Pulser II	Bio-RAD
centrifuge 5804R	Eppendorf
centrifuge 5415 R	Eppendorf
centrifuge Sorval RC6+	Sorvall
concentrator	PALL
CD Spectrometer model 62ADS	AVIV, LAKEWOOD
crystallization robot Gryphon	ARI-Art Robbins Instruments
ice producing machine AF 200	Scotsman
Imager RI182	Formulatrix
Imager RI1000	Formulatrix
incubator	Memmert
lamp for SDS PAGE gel Hobbylite	Kaiser
magnetic mixer	Tehtnica
mass spectrometry ESI-Q-TOF Q Star Elite	Applied Biosystems
NanoVue	GE Healthcare
PCR 2720 Thermal Cycler	Applied Biosystems
RT PCR Mx 3005 P	Stratagene
SDS-PAGE equipment	BioRad
shaker Innova4430	New Brunswick Scientific
shaker Innova44	New Brunswick Scientific
Ultraspec 8000	GE Healthcare
weight PG503-S DeltaRange	Mettler Toledo
X-RAY MicroStar	Bruker

A.1.4 Peptides

AVAEKQ	Biomatik
LLKVAL	Biomatik
ALAASS	Biomatik
AVRARL	Biomatik

Appendix B

Published articles:
ACCEPTED FOR PUBLICATION
Sosnowski P., Turk D., Caught in the Act: The Crystal Structure of cleaved Cathepsin L bound to the active site of Cathepsin L (To be published)

Borisek, J., Vizovisek, M., Sosnowski, P., Turk, B., Turk, D., Novic, M., Development of N-(Functionalized benzoyl)-homocycloleucyl-glycinonitriles as Potent Cathepsin K inhibitors, 2015, J Med Chem, 58(17): 6928-37, doi: 10.1021/acs.jmedchem.5b00746

Caught in the Act: The Crystal Structure of cleaved Cathepsin L bound to the active site of Cathepsin L

Piotr Sosnowski[1,2], Dušan Turk[1,2,3*]

[1]Centre of Excellence for Integrated Approaches in Chemistry and Biology of Proteins, Slovenia
[2]International Postgraduate School Jozef Stefan, Jamova 39, 1000 Ljubljana
[3]Jozef Stefan Institute, 1000 Ljubljana, Slovenia

*To whom correspondence should be addressed: Jozef Stefan Institute, 1000 Ljubljana, Slovenia, Phone: +386 1 477 3857, Fax: +386 1 477 3984, email: Dusan.Turk@ijs.si

Keywords:

cathepsin, cysteine cathepsin, substrate interaction

Abbreviations:

C25S – point mutation at 25th cysteine to serine using mature cathepsin L numbering

GlcNAc – *N*-acetyl-D-glucosamine

MHC-II - (major histocompatibility complex) class II molecules

SDS-PAGE - sodium dodecyl sulfate polyacrylamide gel electrophoresis

Highlights:

Crystal structure shows a protease substrate complex between two cathepsin L molecules.

After cleavage cathepsin L remains bound to the S subsites of the neighbor molecule.

Cathepsin L active site mutants can still exhibit low levels of catalytic activity.

Abstract

Cathepsin L is a ubiquitously expressed papain-like cysteine protease involved in the endosomal degradation of proteins and has numerous roles in physiological and pathological processes, such as arthritis, osteoporosis and cancer. Insight into the specificity of cathepsin L is important for elucidating its physiological roles and drug discovery. To study interactions with synthetic ligands, we prepared a presumably inactive mutant and crystallized it. Unexpectedly, the crystal structure determined at 1.4 Å revealed that the cathepsin L molecule is cleaved, with the cleaved region trapped in the active site cleft of the neighboring molecule. Hence, the catalytic mutant demonstrated low levels of catalytic activity.

Introduction

Cathepsin L is a member of the family of papain-like cysteine proteases. In humans, two enzymes - cathepsins L and V - correspond to the mouse homologue cathepsin L. Historically, cysteine cathepsins were described as endosomal/lysosomal proteases. However, in this millennium, the view of cysteine cathepsins as lysosomal proteases has changed due to a growing body of evidence of their localization in other cellular compartments (Reviewed in [1]). In addition to their role in endolysosomal protein degradation machinery, they participate in processes such as endosomal antigen presentation and are important regulators and signaling molecules of a large number of biological processes [2], such as thyroid hormone processing [3]. They are also involved in diseases, such as cancer and rheumatoid arthritis [4, 5]. In contrast to other cysteine cathepsins with a rather uneven tissue-specific expression pattern, cathepsin L is ubiquitously expressed and thereby involved in large number of physiological processes, such as elastin degradation [6-8]. In mice, it was found to be localized in the nucleus, where it is involved in the processing of histone proteins [9]. Questions remain elucidation of underlying properties which define a protein as a substrate of cathepsin L and the interplay and roles of protein localization and substrate selectivity. Here, we focus on selectivity.

The crystal structures of the human cathepsin L and the rat proenzyme provided insight into the active site cleft, which was unrestricted at both sides, enabling binding of rather long stretches of substrate chains – a characteristic shape for endopeptidases [10-12]. From structural studies of substrate-like inhibitors, we learned that they bind in the active site cleft in extended conformations, where they interact within strictly confined binding sites S2, S1 and S1' and areas beyond, where the compounds adapt to the best matching contact surface [13]. Studies showed that this general understanding is valid for cathepsin L as well [14, 15]. Recently, the groups of Min and Mort made attempts to gain insight into the interactions between a part of the physiological substrate histone H3 and human cathepsin L by crystallizing the complex with a catalytic mutant of the enzyme and elucidating its structure [16]. Refinement confirmed the positioning of only three amino acids from the histone $H3_{19-33}$ peptide. Our crystal packing analysis suggested that there is not enough space to accommodate all 14 peptide residues in the available space in the active site cleft. Nevertheless, this work encouraged us to address the substrate selectivity by setting up a platform that would enable us to determine crystal structures

of complexes of cathepsin L with a variety of peptides. Therefore, the presumably inactive catalytic mutant C25S of cathepsin L was expressed and crystals in the absence of a peptide were grown. To our surprise, the crystals contained cleaved cathepsin L molecules with a part of the cleaved region still attached to the active site of the neighboring molecule. Hence, the structure provided rare insight in the intermediate state of the reaction between two protein molecules immediately after the cleavage, when the N-terminal part of the substrate still remained bound to the non-primed binding side and the C-terminal part already left the primed binding sites.

Materials and Methods

Expression of human cathepsin L-C25S mutant.

Human procathepsin L cDNA was obtained from Deutsches Ressourcenzentrum fuer Genomforschung GmbH. The cDNA of human procathepsin L was amplified by PCR and the specific mutation C25S (mature cathepsin numbering) was introduced to express the inactive protein. It was prepared as described [17]. Additional site-specific mutations were introduced at T110A to prevent glycosylation. PCR products were inserted into the *Pichia pastoris* expression vector pPIC9 from Invitrogen between the NotI-HF and AvrII restriction sites. Recombinant plasmids were inserted into *P. pastoris* strain GS115 (Invitrogen) by electroporation. The isolation of a clone with the best protein expression was performed in accordance with the Invitrogen manual. The protein was expressed and isolated as described [18]. Fractions of human pro-cathepsin L-C25S were joined and dialyzed against activation buffer (100 mM sodium acetate, pH 5.0, 1 mM EDTA, 5 mM DTT, [18]). Propeptide was cleaved by adding pepsin to a final concentration of 0.03 mg/ml and incubated for 3 hours at 37 °C. Afterwards, the protein was dialyzed against 50 mM sodium acetate (pH 5.5) and applied to a previously equilibrated SP sepharose column. Pure fractions of human apo-cathepsin L C25S were eluted at the beginning of gradient elution from 0 M to 2 M NaCl. Fractions were joined and dialyzed against 20 mM sodium acetate, pH 4.5, 100 mM NaCl, 1 mM DTT, 5% glycerol, concentrated to 44 mg/ml, and stored at -20 °C. In addition, to confirm propeptide removal, mature cathepsin L-C25S was analyzed for N-terminal sequencing using a Procise protein sequencing system 492A (PE Applied Biosystems) at the Department of Molecular and Biomedical Sciences, Jozef Stefan Institute.

Crystallization and structure determination of human cathepsin LC25S mutant.

Conditions for the crystallization of human cathepsin L-C25S were established using QIAGEN and Hampton Crystallization screens. Two conditions delivering protein crystals were further optimized and the diffracting crystals were acquired. Rod-shaped crystals grew within two weeks in 0.1 M BisTRIS/HCl, pH 5.5, 2 M ammonium sulfate. Cube-shaped crystals grew within two weeks in 0.11 M HEPES/NaOH, pH 7.5, 2.14 M ammonium sulfate. Crystals were harvested, cryo-protected in the crystallization solution with 25% glycerol and flash cooled in liquid nitrogen. Data sets were collected at BESSY BEAMLINE 14.1 in Berlin on a PILATUS 6M detector [19]. Crystals from both conditions diffracted and after analysis revealed the same structure. The description of the structure is based on the best data set from the cubic shaped crystal, which diffracted to resolution of 1.42 Å. The data were collected at 1.0 Å wavelength. The data set was processed with the XDSAPP [20] and scaled in the P212121 space group. The Matthews coefficient analysis [21] unambiguously defined one molecule in the asymmetric unit. The structure of cathepsin L-C25S was determined using PHASER [22] with the cathepsin L structure (PDB ID 2YJ2) [23] as the search model. Model building was performed using *Coot* [24] and *MAIN* [25]. Refinement started using *REFMAC5* [26] and was completed with *MAIN* [25] using the free kick maximum likelihood target, which allows the use of all data in refinement by the simulation of the model error. The data statistics are presented in Table 1. The completeness in highest resolution shell is at 69%, because it only includes reflections from the corners of a square detector. The detector was set to collect a complete data set at up to 1.6 Å resolution as a consequence of on our initial estimate of diffraction limit of 1.7 Å.

Coordinates of cathepsin L complex was deposited in the Protein Data Bank (http://www.pdb.org) [27], accession number 5I4H.

Results

Cathepsin L-C25S crystallized in the orthorhombic space group P212121 with unit cell dimensions a = 39.97 Å, b = 67.36 Å, c = 77.39 Å. The crystals diffracted to 1.42 Å resolution and contained one molecule in the asymmetric unit. The crystal structure of cathepsin L-C25S mutant contains sequences from A1 to V220 using the mature enzyme nomenclature, with additional residue E-1 at the N-terminus. The presence of E-1 residue was confirmed by the N-

terminal sequencing of the protein. The chromatogram presented in Figure 1 shows two possible N-terminal sequences: EAPRS and APRSV, which differ by one residue. They both correspond to the native cathepsin L sequence. In the crystal structure, the residues from A1 to S105 and from D109 to V220 were unambiguously resolved by the electron density maps, whereas the residues V106, A107, and N108 could not be positioned and are therefore absent from the model. The E-1 residue was partially fitted to the electron density. Several residues were built in alternate conformations. The asymmetric unit of the crystal also contains five SO_4^{2-} ions and a molecule of glycerol.

Cathepsin L's structure reveals the fold of the papain-like enzyme composed of two domains named left (L-) and right (R-). The L-domain is α-helical, whereas the R-domain forms a β-barrel fold, as shown in the pink molecule in Figure 2. Between the two domains, there is a V-shaped active site cleft on which the L- and R-domain catalytic residues C25 and H163 are positioned. In this structure, the C25 residue was mutated to serine to abolish its activity.

To our surprise, we found that the cathepsin L molecule in the crystal was cleaved between the residues S105 and V106 within the region located at the surface of the L-domain and that the crystal structure actually contains a part of the peptidyl substrate bound to the non-primed side of the active site. The new carboxylic end (S105) of the N-terminal part of the cleaved peptide is located in the active site of the neighboring molecule related by the ½+x, -½-y, -z symmetry operator. Its residues are marked with an asterisk (*). The C-terminal part of the cleaved peptide is resolved from D109 onwards. It points away from the protein surface. The zig-zag pattern of the N-terminal region of the cleaved peptides linking cathepsin L-C25S molecules in the crystal packing is shown in orange in Figure 2. The peptides are positioned alternatively in front and at the back of the pattern. Hence, the crystal structure of cathepsin L C25S mutant is a complex between the molecule being cleaved and the neighboring molecule performing the cleavage. Because each molecule is cleaved and has the active site performing the cleavage, each molecule simultaneously plays two roles: that of an enzyme and that of a substrate.

The structure of the cleaved peptide and its interaction with the neighboring cathepsin L-C25S molecule in detail are shown in Figure 3. The cleaved peptide residues K103, Y104 and S105, at positions P3, P2 and P1, respectively (using Schechter and Berger nomenclature [28]), are positioned within the S3, S2 and S1 binding sites of the neighboring molecule, respectively. The

C-terminal residue S105 (P1) forms three hydrogen bonds with main chain and side chain atoms of residues Q19*, S25* and H163*. The interatomic distances between S105 carbonyl oxygens and Q19* atom NE1, S25* atom OG and H163* atom ND1 side chains are 2.95 Å, 2.75 Å and 2.8 Å, respectively. The Y104 (substrate residue P2) fills the hydrophobic S2 binding site between A135*, L69*, and D162*. The Y104 hydroxylic group does not make any additional interactions. The K103 side chain (substrate residue P3) is located in the S3 binding site. The K103 side chain amino group stabilizes the binding of the cleaved peptide by hydrogen bond interactions with three oxygen atoms: the S97 side chain atom OG and the C98 and N101 main chain carbonyls. The residues before K103 (P3) are not involved in binding to the neighboring molecule. Their structure is stabilized within the region of the cleaved peptide itself. The P102 residue is located over the K103, protecting the lysine side chain from solvent. The N101 is present in two conformations; in one, the side chain forms a hydrogen bond with the K103 main chain oxygen atom, whereas in the other, it is oriented ~90° away from the first and it weakly interacts with the Y100 main chain nitrogen atom. The Y100 is positioned parallel to the K99 and forms a hydrogen bond with the D84 oxygen OD2 atom. The K99 side chain is positioned next the Y89 and does not interact with the enzyme residues. Hence, the three residues of the N-terminal part of the cleaved peptide are stabilized within the active site of the neighboring molecule, whereas the structure of the preceding residues is stabilized within the environment of their own molecules.

To confirm that the cathepsin L-C25S was auto-cleaved in two parts also by biochemical analysis, crystals were dissolved and analyzed using SDS-PAGE electrophoresis. Figure 4 show the two bands from the cathepsin L-C25S from dissolved crystals with the apparent weights of 11.5 kDa and 12.6 kDa. They are clearly distinct from the single band of cathepsin L sample before crystallization with the apparent weight of 24.1 kDa. The weight of two bands corresponds to the parts resulting from the cleavage observed in the crystal structure.

To validate the importance of N108 glycosylation for autocatalytic cleavage of procathepsin L, the procathepsin L was incubated in 100 mM sodium acetate, pH 5.5, 5 mM DTT, 5 mM EDTA buffer at 37 °C. On the third day, the sample was degraded without revealing any intermediates, thereby exposing no particular biological relevance of this glycosylation against the cleavage.

This suggests that the observed cleavage intermediate can be trapped only under the crystallization conditions.

Discussion

The crystal structure of cathepsin L surprisingly provided insight into the proteolysis of a protein substrate by revealing the interaction between a cysteine cathepsin and its protein substrate, in which the N-terminal part of the cleaved peptide (from K103 to S105) remained bound to the active site cleft of the neighboring molecule in the crystal. To gain insight into the cleavage event, the structure of the cleaved region from our structure was superimposed with four other known cathepsin L structures: the structure of cathepsin L isolated from human kidney [11] (khCL); the recombinant human cathepsin L structure (rhCL_1) [14] with the same sequence (T110A mutation) as that presented here, which is the full structure with exception of the catalytic site C25; the recombinant human cathepsin L (rhCL_2) structure [16] with a sequence that enables the glycosylation of the N108 residue and imbeds the catalytic site C25A mutation; and the highest resolution recombinant human cathepsin L covalently linked with a nitrile inhibitor [29] (rhCL_3). Our structure of cathepsin L C25S is superimposed with the structure of cathepsin L khCL, rhCL_1, rhCL_2 and rhCL_3, with a RMSD of the Cα-atoms of 0.57 Å, 0.4 Å, 0.57 Å and 1 Å, respectively (Table 2). The superimposition of structures showed that the regions of the cleaved peptide from K99 to A110 in the uncleaved molecules share a similar conformation and that the conformation and positioning of the residues K103, Y104 and S105 in the structure of our cathepsin L C25S mutant was clearly different from that observed in uncleaved molecules (Figure 5). The structures of the uncleaved molecules thus represent the structure before cleavage. Before cleavage, this region was folded as a short α-helix. Together, these comparisons demonstrate that the cleaved region must first be pulled out of the conformation observed in the uncleaved form into the conformation that allowed cleavage to occur. After cleavage took place, the cleaved molecules must have rotated approximately 90° away from the position in which they were cleaved to reach the position observed in the crystal. This movement is reminiscent of the change observed in the trypsin-antitrypsin complex, where after cleavage, the enzyme molecule is transferred to the other side of antitrypsin [30]. However, in the trypsin-antitrypsin structure, the cleaved peptide remained attached to the catalytic serine residue by a covalent bond, whereas in our case, this movement indicated a strong non-covalent

binding affinity between the C-terminus of the cleaved peptide and the non-primed part of the active site cleft of cathepsin L.

Our structure enabled us to address the question of whether the binding of small synthetic inhibitors [14] mimics the substrate binding in cysteine cathepsins. The positioning and orientation of the K103 to S105 region peptide is tightly superimposed with the main and side chains in the S1 and S2 binding sites (Figure 6), whereas in the S3 binding sites, due to the differences in chemical composition, the positioning of the peptidyl ligands begins to diverge.

The remaining question is why our cathepsin L molecule was cleaved, whereas in other cathepsin L structures, cathepsin L remains intact. Access to the active site was prevented by inhibitors including the propeptide, which explains the crystal structures of khCL complex with the p41 fragment [11], the small molecule complexes rhCL_1 and rhCL_3 [14, 29] and the proenzyme structure [10]. However, among the remaining two other structures of cathepsin L, rhCL_2 and ours, ours was cleaved and the other was not. In both structures, the peptidyl ligand was visible only in the non-primed region, which suggests that also the rhCL_2 C25A mutant likely retained low catalytic activity [16]. The other difference was observed in the cleaved region. Namely, the N108 is in the rhCL_2 structure glycosylated, whereas in ours, glycosylation was prevented by the T110A mutation, which suggests that the glycosylation of N108 may be responsible for the difference in behavior. To address the biological significance of protective role of N108 glycosylation, we performed a timed dependent autoprocessing assay, which did not expose its relevance. Nevertheless, this experiment does not exclude a possible protective role of the glycosylation of N108 against cleavage by other proteases.

In conclusion, the cleavage of cathepsin L C25S in two parts during the two-week crystallization process clearly shows that the cathepsin L active site mutation C25S does not completely inactivate the enzyme. It only reduces its potency to a level that is not detectable by fluorogenic substrates. The crystal structure caught the cathepsin L protease – cathepsin L substrate complex "in the act" after cleavage and before the final separation of the two molecules. We anticipate that the C25S mutant and the C25A active site mutant can hydrolyze peptide bonds. In this respect, papain-like cysteine proteases behave similarly to serine proteases for which the presence of an oxyanion hole assures the low catalytic activity of the protein [31]. Therefore, another approach

will have to be applied to determine the structure of the peptidyl substrate that is bound along the whole length of the active site of cathepsin L and other papain-like proteases.

Acknowledgements

Ajda Taler-Verčič is acknowledged for careful proofreading of the manuscript. This work was supported by grants from the Slovenian Research Agency (P-00480), the young researcher stipend and CIPKeBiP infrastructural founds.

References: (max 70)

1. Turk, V., Stoka, V., Vasiljeva, O., Renko, M., Sun, T., Turk, B. & Turk, D. (2012) Cysteine cathepsins: from structure, function and regulation to new frontiers, *Biochimica et biophysica acta.* **1824**, 68-88.
2. Turk, B., Turk, D. & Turk, V. (2012) Protease signalling: the cutting edge, *The EMBO journal.* **31**, 1630-43.
3. Hitzel, C., Kanzler, H., Konig, A., Kummer, M. P., Brix, K., Herzog, V. & Koch, N. (2000) Thyroglobulin type-I-like domains in invariant chain fusion proteins mediate resistance to cathepsin L digestion, *FEBS letters.* **485**, 67-70.
4. Olson, O. C. & Joyce, J. A. (2015) Cysteine cathepsin proteases: regulators of cancer progression and therapeutic response, *Nature reviews Cancer.* **15**, 712-29.
5. 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, *Current pharmaceutical design.* **13**, 387-403.
6. Bromme, D., Okamoto, K., Wang, B. B. & Biroc, S. (1996) Human cathepsin O2, a matrix protein-degrading cysteine protease expressed in osteoclasts. Functional expression of human cathepsin O2 in Spodoptera frugiperda and characterization of the enzyme, *The Journal of biological chemistry.* **271**, 2126-32.
7. Li, Z., Yasuda, Y., Li, W., Bogyo, M., Katz, N., Gordon, R. E., Fields, G. B. & Bromme, D. (2004) Regulation of collagenase activities of human cathepsins by glycosaminoglycans, *The Journal of biological chemistry.* **279**, 5470-9.
8. Yasuda, Y., Li, Z., Greenbaum, D., Bogyo, M., Weber, E. & Bromme, D. (2004) Cathepsin V, a novel and potent elastolytic activity expressed in activated macrophages, *The Journal of biological chemistry.* **279**, 36761-70.
9. 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.
10. 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, *The EMBO journal.* **15**, 5492-503.
11. 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, *The EMBO journal.* **18**, 793-803.

12. 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 letters.* **384**, 211-4.
13. Turk, B., Turk, V. & Turk, D. (1997) Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors, *Biological chemistry.* **378**, 141-50.
14. Asaad, N., Bethel, P. A., Coulson, M. D., Dawson, J. E., Ford, S. J., Gerhardt, S., Grist, M., Hamlin, G. A., James, M. J., Jones, E. V., Karoutchi, G. I., Kenny, P. W., Morley, A. D., Oldham, K., Rankine, N., Ryan, D., Wells, S. L., Wood, L., Augustin, M., Krapp, S., Simader, H. & Steinbacher, S. (2009) Dipeptidyl nitrile inhibitors of Cathepsin L, *Bioorganic & medicinal chemistry letters.* **19**, 4280-3.
15. Chowdhury, S. F., Sivaraman, J., Wang, J., Devanathan, G., Lachance, P., Qi, H., Menard, R., Lefebvre, J., Konishi, Y., Cygler, M., Sulea, T. & Purisima, E. O. (2002) Design of noncovalent inhibitors of human cathepsin L. From the 96-residue proregion to optimized tripeptides, *Journal of medicinal chemistry.* **45**, 5321-9.
16. Adams-Cioaba, M. A., Krupa, J. C., Xu, C., Mort, J. S. & Min, J. (2011) Structural basis for the recognition and cleavage of histone H3 by cathepsin L, *Nature communications.* **2**, 197.
17. Sambrook, J. & Russell, D. W. (2001) *Molecular cloning : a laboratory manual*, 3rd edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
18. Mihelic, M., Dobersek, A., Guncar, G. & Turk, D. (2008) Inhibitory fragment from the p41 form of invariant chain can regulate activity of cysteine cathepsins in antigen presentation, *The Journal of biological chemistry.* **283**, 14453-60.
19. Mueller, U., Darowski, N., Fuchs, M. R., Forster, R., Hellmig, M., Paithankar, K. S., Puhringer, S., Steffien, M., Zocher, G. & Weiss, M. S. (2012) Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin, *Journal of synchrotron radiation.* **19**, 442-9.
20. Krug, M., Weiss, M. S., Heinemann, U. & Mueller, U. (2012) XDSAPP: a graphical user interface for the convenient processing of diffraction data using XDS, *J Appl Crystallogr.* **45**, 568-572.
21. Kantardjieff, K. A. & Rupp, B. (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals, *Protein science : a publication of the Protein Society.* **12**, 1865-71.
22. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007) Phaser crystallographic software, *J Appl Crystallogr.* **40**, 658-674.
23. Hardegger, L. A., Kuhn, B., Spinnler, B., Anselm, L., Ecabert, R., Stihle, M., Gsell, B., Thoma, R., Diez, J., Benz, J., Plancher, J. M., Hartmann, G., Isshiki, Y., Morikami, K., Shimma, N., Haap, W., Banner, D. W. & Diederich, F. (2011) Halogen bonding at the active sites of human cathepsin L and MEK1 kinase: efficient interactions in different environments, *ChemMedChem.* **6**, 2048-54.
24. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010) Features and development of Coot, *Acta crystallographica Section D, Biological crystallography.* **66**, 486-501.
25. Turk, D. (2013) MAIN software for density averaging, model building, structure refinement and validation, *Acta crystallographica Section D, Biological crystallography.* **69**, 1342-57.
26. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures, *Acta crystallographica Section D, Biological crystallography.* **67**, 355-67.
27. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Res.* **28**, 235-42.

28. Schechter, I. & Berger, A. (1967) On the size of the active site in proteases. I. Papain, *Biochemical and biophysical research communications.* **27**, 157-62.
29. Hardegger, L. A., Kuhn, B., Spinnler, B., Anselm, L., Ecabert, R., Stihle, M., Gsell, B., Thoma, R., Diez, J., Benz, J., Plancher, J. M., Hartmann, G., Banner, D. W., Haap, W. & Diederich, F. (2011) Systematic investigation of halogen bonding in protein-ligand interactions, *Angewandte Chemie.* **50**, 314-8.
30. Huntington, J. A., Read, R. J. & Carrell, R. W. (2000) Structure of a serpin-protease complex shows inhibition by deformation, *Nature.* **407**, 923-6.
31. Brändén, C. I. & Tooze, J. (1999) *Introduction to Protein Structure*, Garland Pub.
32. LLC., S. (2014) The PyMOL Molecular Graphics System in
33. Krissinel, E. (2007) On the relationship between sequence and structure similarities in proteomics, *Bioinformatics.* **23**, 717-723.

Figure legends

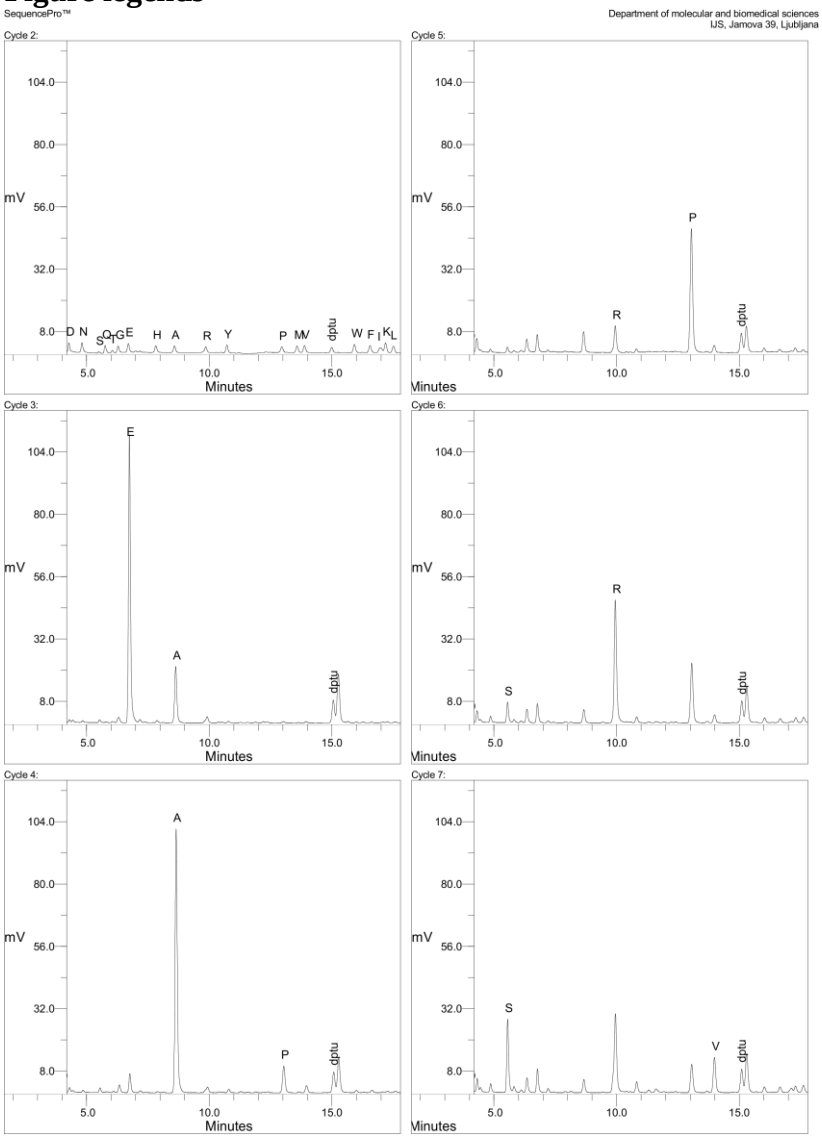


Figure 1. Chromatogram of human cathepsin L-C25S N-terminal amino acid separation and Edman degradation byproduct DPTU. Amino acids are marked with standard one-letter codes. Cycle 2 is the reference cycle for further cycles. Cycles 3-7 correspond to the N-terminal amino acid positions 1-5, respectively.

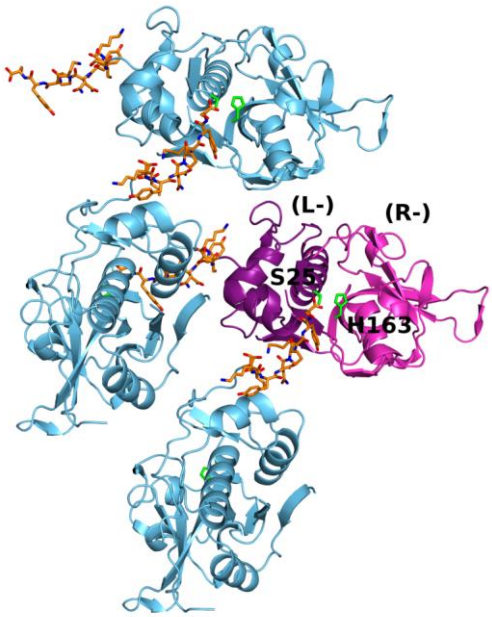


Figure 2. The crystal packing of cathepsin L-C25S exhibits a zig-zag alignment. Each molecule represents an asymmetric unit. The fold is shown in blue and violet ribbons, while the N-terminal part of the cleaved peptide reaching into the neighboring molecules are shown as orange sticks. R- and L-domains of the violet molecule are shown in different purple hues. The active site residues C25 and H163 are labeled and colored in green. The picture was prepared using the open source PyMOL™ v.1.7.2.1 [32].

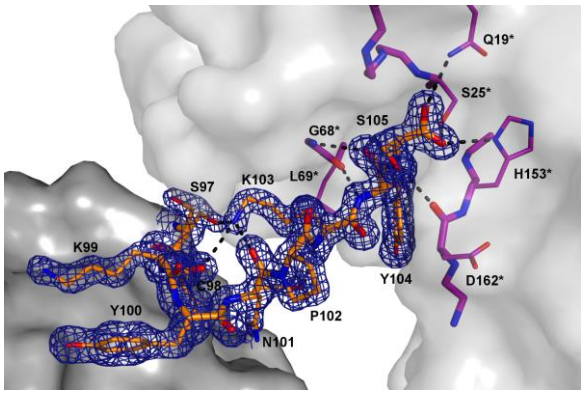


Figure 3. Binding of the N-terminal part of the cleaved peptide. The cleaved peptide and the residues in contact with the neighboring molecule are shown as sticks on the background of the white and grey surfaces of the two neighboring cathepsin L molecules. The main and side chain bonds are presented as the thicker and thinner sticks, respectively. Parts of the bonds belonging to oxygen, nitrogen and sulfur atoms are colored red, blue and yellow, respectively, whereas the cleaved peptide carbons are shown in orange and those of the attached molecule in pink. Electron density map around the N-terminal part of the cleaved peptide is contoured at 1.0 sigma and is presented as a blue mesh. Hydrogen bonds between the peptide and the underlying cathepsin L molecules are presented as dashed lines. The figure was prepared using the open source PyMOLTM v.1.7.2.1 [32].

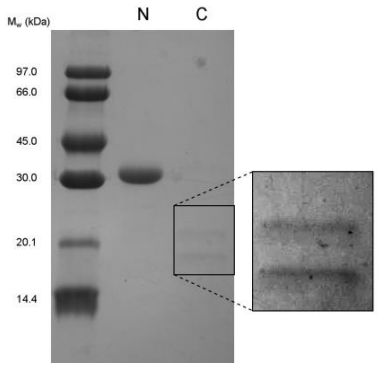


Figure 4. Auto-processing of human cathepsin L-C25S before and after crystallization. The black and white picture of the material before and after crystallization are presented on the 15% SDS-PAGE gel colored by Coomassie brilliant blue. The two bands of cathepsin L after the cleavage are enlarged and contrasted for better visualization. Mw, N and C stand for molecular weight marker, native protein before crystallization and crystallized protein, respectively.

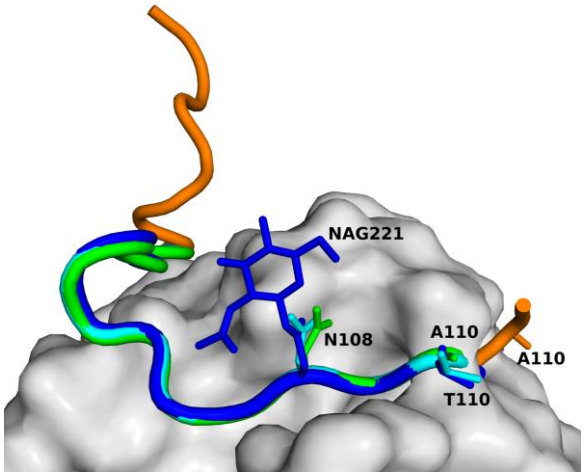


Figure 5. The cathepsin L-C25S cleaved region superimposed on the same region of the rhCL_1, rhCL_2 and rhCL_3 structures. The region of the cleaved peptide is located on the white surface of our cathepsin L-C25S structure. The protein main chains within the region are presented as ribbons. The regions of rhCL_1, rhCL_2, rhCL_3 and our structure are colored green, blue, cyan and orange, respectively. The glycosylated N108 and the NAG221 from rhCL_2 and mutated amino acids within the region are labeled and presented as sticks. The figure was prepared using the open source PyMOLTM v.1.7.2.1 [32].

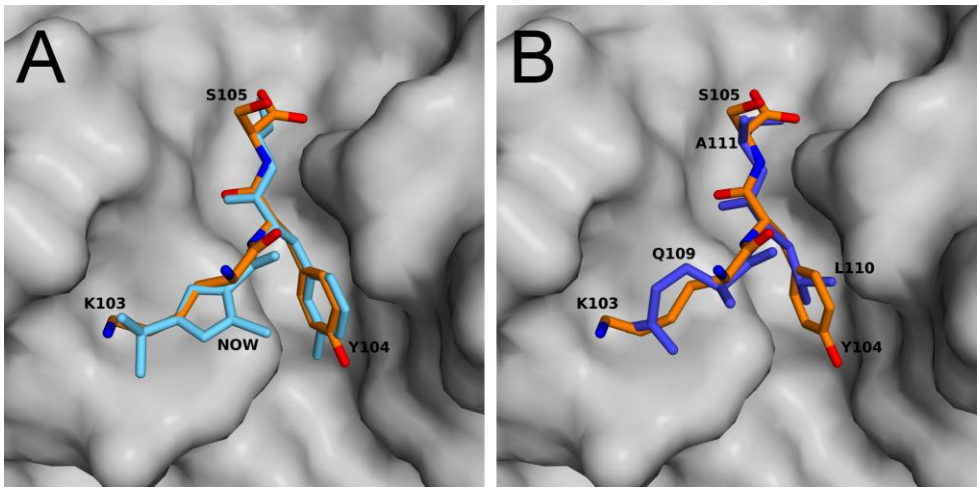


Figure 6. Superimposition of peptidyl ligands within the non-primed region of the active site. Peptidyl ligands are shown on the background of the grey cathepsin C25S surface. The three residues of the N-terminal part of the cleaved peptide are shown in orange. They are superimposed on A) the NOW inhibitor from rhCL_1 structure [14] and B) the QLA peptide from rhCL_2 structure [16]. The NOW and QLA peptide are colored cyan and blue, respectively. The picture was prepared using the open source PyMOLTM v.1.7.2.1 [32].

Data collection	Cathepsins L-C25S
PDB ID	5I4H
Wavelength (Å)	1.0
Resolution range (Å)	38.7 - 1.42 (1.47 - 1.419)
Space group	P 21 21 21
Unit cell	a=39.97 Å b=67.36 Å c=77.39 Å α=β=γ=90°
Reflections	
Unique reflections	38421 (2741)
Completeness (%)	95.39 (69.23)
I/σ(I)	20.46 (2.56)
R-merge	0.05287 (0.4368)
R-meas	9.1
Refinement	
Reflections	38401
Resolution (Å)	38.7 – 1.42
R-work	0.1626 (0.2636)
R-free-kick	0.1841 (0.2885)
Atom numbers	
Non-hydrogen atoms	1963
Macromolecules	1685
Ions and other molecules	32
Water	246
Protein residues	220
RMSD from target geometry	
Bond lengths (Å)	0.022
Bond angles (°)	2.03
Ramachandran plot (%)	
Favored	98
Allowed	2
Outliers	0
Average B-factor	*17.6*
Macromolecules	15.7
Ions and other molecules	35.9
Solvent	29.7

Table 1 Human cathepsin L-C25S data and refinement statistics. Values in brackets correspond to the highest resolution shell 10 for data processing 20 for refinement (1.42-1.44).

Model/ PDB ID	Q scoring	Cα RMSD (Å)	N-align
khCL/ 1ICF	0.86	0.58	206
rhCL_1/ 3HHA	0.86	0.4	204
rhCL_2/ 3K24	0.86	0.57	206
rhCL_3/ 2XU3	0.88	1	209

Table 2 Calculated statistics of cathepsin L-C25S aligned with khCL, rhCL_1, rhCL_2 and rh_CL3. Q scoring - Quality of alignment, which took both the alignment length and RMSD into account. Cα RMSD – Root Mean Square Deviation, calculated between Cα-atoms. N-align - number of matched residues. Data were acquired using PDBeFold [33].

Journal of Medicinal Chemistry

Article

pubs.acs.org/jmc

Development of *N*-(Functionalized benzoyl)-homocycloleucyl-glycinonitriles as Potent Cathepsin K Inhibitors

Jure Borišek,[†] Matej Vizovišek,[‡] Piotr Sosnowski,[‡] Boris Turk,*[,‡,§] Dušan Turk,*[,‡,§,∥] Barbara Mohar,[†] and Marjana Novič*[,†]

[†]National Institute of Chemistry, Hajdrihova 19, SI-1001 Ljubljana, Slovenia

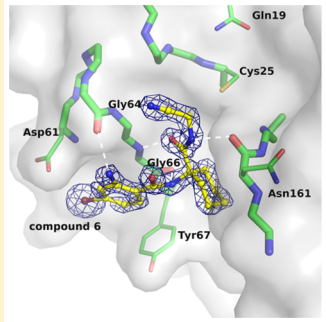
[‡]Department of Biochemistry, Molecular and Structural Biology, Jozef Stefan Institute, Jamova cesta 39, SI-1000 Ljubljana, Slovenia

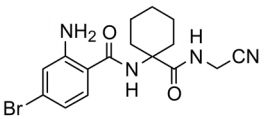
[§]Centre of Excellence for Integrated Approaches in Chemistry and Biology of Proteins, Jamova cesta 39, SI-1000 Ljubljana, Slovenia

[∥]Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, SI-1000 Ljubljana, Slovenia

Supporting Information

ABSTRACT: Cathepsin K is a major drug target for osteoporosis and related-bone disorders. Using a combination of virtual combinatorial chemistry, QSAR modeling, and molecular docking studies, a series of cathepsin K inhibitors based on *N*-(functionalized benzoyl)-homocycloleucyl-glycinonitrile scaffold was developed. In order to avoid previous problems of cathepsin K inhibitors associated with lysosomotropism of compounds with basic character that resulted in off-target effects, a weakly- to nonbasic moiety was incorporated into the P3 position. Compounds **5**, **6**, and **9** were highly selective for cathepsin K when compared with cathepsins L and S, with the K_i values in the 10−30 nM range. The kinetic studies revealed that the new compounds exhibited reversible tight binding to cathepsin K, while the X-ray structural studies showed covalent and noncovalent binding between the nitrile group and the catalytic cysteine (Cys25) site.





compound **6**

K_i (Cat K) = 10 nM
K_i (Cat L) = 11000 nM
K_i (Cat S) = no inhibition

INTRODUCTION

Cathepsin K (Cat K)[1−7] is a lysosomal cysteine protease, which has a major role in osteoporosis and other bone-related pathologies, diabetes, obesity, cancer, and atherosclerosis,[8−14] rendering it an attractive target for the development of new therapeutics. For the almost two past decades, pharmaceutical industry has been focusing on the development of reversible covalent Cat K inhibitors for osteoporosis treatment based on nitrile or carbonyl warheads.[15,16] The most advanced drug candidates, such as ONO-5334 and Odanacatib, are currently in clinical trials with Odanacatib successfully concluding the first phase III in postmenopausal women with osteoporosis (Figure 1a).[17−25] In contrast, the phase II clinical trials with *N*-[4-(4-propyl-piperazino)benzoyl]-homocycloleucyl-glycinonitrile Balicatib were discontinued due to unexpected off-target effects.[26] The side effect of Balicatib causing morphea-like skin hardening was ascribed to its lysosomotrophism, a tendency of the drug with a basic character to accumulate in the acidic lysosomal environment.[27] In order to avoid such side effects, the design of less basic *N*-(functionalized benzoyl)-homocycloleucyl-glycinonitriles with high biological activity and selectivity is of great interest.

Cat K binding pocket involves three main binding sites: the S1 site lies entirely in the L-domain. The P1 residue carbonyl group is placed into the oxyanion hole formed on the N-terminus of the central helix between the Cys25 amide proton and the Gln19 side chain amide, whereas its side chain points upward into the solvent regions. The binding site for the P1 side chain is formed between the Asn21, Cys22, and Gly23 loop, which continues down into the central helix on one side and a broad turn formed by the main-chain atoms of Cys63, Gly64, and Gly65 on the other side. Both chains are cross-linked at the top with the Cys22−Cys63 disulfide bond. Hence, a P1 residue side chain would point along the S1 site binding surface away from the bottom of the active site cleft toward the solvent. The hydrophobic S2 site with main chain amide and carbonyl groups of Gly66 as hydrogen donors and acceptors of the main chain P2 residue and hydrophobic moiety of the S2 site spanned between Tyr67, Ala134, and Leu209, and the S3 area, which is relatively shallow and solvent exposed (Figure 1b).[16,28,29]

In order to develop *N*-("low basic" functionalized benzoyl)-homocycloleucyl-glycinonitriles as Cat K inhibitors, we first combined several computer-assisted drug design (CADD) tools, including virtual combinatorial chemistry, quantitative structure−activity relationship (QSAR) modeling, and molecular docking, which were already described as a successful approach in drug lead discovery and optimization.[30,31] The applied QSAR model is based on our published work related to

Received: May 15, 2015
Published: August 17, 2015

ACS Publications © 2015 American Chemical Society

DOI: 10.1021/acs.jmedchem.5b00746
J. Med. Chem. 2015, 58, 6928−6937

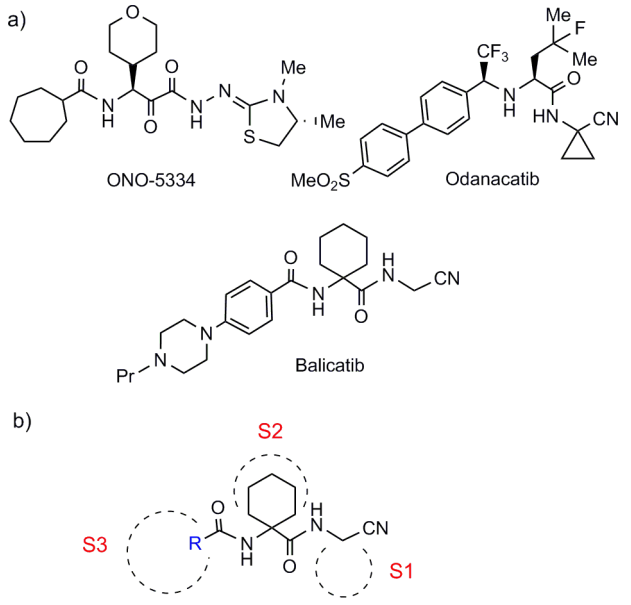


Figure 1. (a) Representative Cat K inhibitors. (b) Binding of *N*-(functionalized benzoyl)-homocycloleucyl-glycinonitriles within the Cat K active site indicated by S1−S3 subsites.

the prediction of biological activity of a given organic compound.[32] Next, a selection of the most promising inhibitor candidates were prepared and evaluated for their selectivity against cysteine cathepsins K, S, and L.

RESULTS AND DISCUSSION

Virtual Combinatorial Chemistry. First, a virtual combinatorial library of *N*-(functionalized benzoyl)-homocycloleucyl-glycinonitriles was generated by amidification of homocycloleucyl-glycinonitrile (**3**) using a database of readily accessible functionalized benzoic acids. For this, a commercial "Enamine fragment database simple-fl"[33] was filtered by a rule-of-three[34] (molecular weight < 300 Da, calculated Log *P* < 3, number of H-bond donor < 3) and was applied in the virtual combinatorial synthesis generating a 270-membered set of compounds. Applying the Veber filter[35] for oral bioavailability resulted in a 254-membered virtual combinatorial chemistry series, referred to as the combinatorial data set (see the Supporting Information, Table S1).

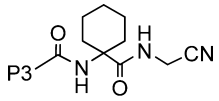
QSAR Modeling. The QSAR model, Counter-Propagation-Artificial Neural Network (CP-ANN), which was developed by some of us,[32] was used to predict the biological activity ($K_{i\text{-pred}}$) of the combinatorial data set within a nM range (Table 1).

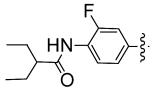
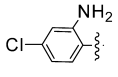
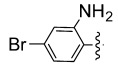
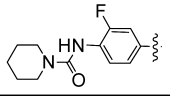
Molecular Docking. The experimentally determined structure of the Cat K-"0LB ligand" complex (PDB ID: 4DMX, Figure S1)[36] was used to carry out covalent molecular docking with our 254-membered combinatorial data set. For validation of this method, a 52-membered series data set, referred to as the experimental data set, composed of known *N*-(aroyl)-homocycloleucyl-glycinonitriles[37] was exploited.

Next, 4DMX was used to visually assess the fitness and orientation of the docked compounds. Redocking ligand validation gave satisfactory results as GOLD calculation reproduced the binding pose of 0LB within Cat K binding site. The ChemScore scoring function (CSF) yielded the lowest root-mean-square deviation (rmsd) value of 0.19 Å for the redocking and it was chosen for the assessment of the binding affinities of the combinatorial (Table 1) and experimental data sets.

Postdocking analysis was carried out in two steps. In the first step, ligand (combinatorial and experimental data sets members) fitness and orientation were assessed and marked with either "T" (true) or "F" (false), depending on the experimentally determined poses of the functional groups of 0LB ligand reference. When both orientation and fitness were correct (T), then the ligand was analyzed in the second step. Thus, visual inspection yielded 113 compounds out of our 254-

Table 1. Most Suitable Virtual Combinatorial *N*-(Functionalized benzoyl)-homocycloleucyl-glycinonitrile Compounds after Combining QSAR Modeling and Molecular Docking



Compound	P3	$K_{i\text{-pred}}$ (nM)	CSF	$d_{4DMX\text{-}1}$	$d_{4DMX\text{-}2}$	$d_{4DMX\text{-}3}$
4		66.1	35.0	3.25	2.91	2.80
5		6.85	32.4	3.23	2.75	2.98
6		66.1	33.6	3.29	2.83	2.92
10		1.40	33.3	3.25	2.76	2.95

DOI: 10.1021/acs.jmedchem.5b00746
J. Med. Chem. 2015, 58, 6928−6937

membered combinatorial data set, and 14 compounds out of the 52-membered experimental data set (see Supporting Information, Table S2).

In the second step of postdocking analysis, the key distances (d) between the features of the combinatorial and experimental ligands (derived from the first analysis step) and the key amino acid residues in 4DMX were calculated. Relying on a published data in the field,[37] three H-bond interactions between each ligand and 4DMX were considered: (R^1-N**H**-CH_2-CN)-(Asn161-C**O**) [d_{4DMX-1} = 3.28 Å], (R^2-C**O**-NH−CH_2-CN)-(Gly66-N**H**) [d_{4DMX-2} = 3.22 Å], and (R^3-Ph−CO-N**H**-R′)-(Gly66-C**O**) [d_{4DMX-3} = 2.9 Å]. Due to the dynamic nature of biological systems, a tolerance of ±1 Å was applied: (R^1-N**H**-CH_2-CN)-(Asn161-C**O**) [d_{4DMX-1} = 2.28−4.28 Å], (R^2-C**O**-NH-CH_2-CN)-(Gly66-N**H**) [d_{4DMX-2} = 2.22−4.22 Å], and (R^3-Ph-CO-N**H**-R′)-(Gly66-C**O**) [d_{4DMX-3} = 1.9−3.9 Å] (Table 1). Interatomic distance calculations yielded 12 valid compounds (or ligands) out of the retained 113-membered combinatorial data set and 7 valid compounds out of the retained 14-membered experimental data set.

Consensus Results. The approach combining QSAR modeling and molecular docking analysis resulted in 4 finally valid compounds (out of our original 254-membered combinatorial data set), which were then synthesized and subjected to kinetic studies. In addition to their predicted biological activities ($K_{i\text{-pred}}$) within the nM range, these compounds presented a correct visual orientation and fitness, and appropriate key interatomic distances between their features and the key amino acid residues in 4DMX (Figure 2).

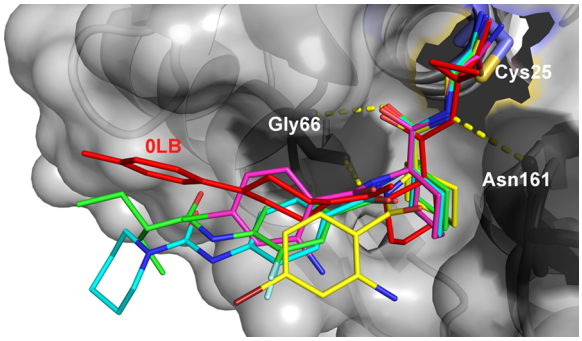


Figure 2. Close view of the natively present and GOLD-calculated covalently bound combinatorial Cat K ligands within 4DMX. The 0LB ligand conformation is depicted in red, and the molecular docking-calculated binding poses for the 4 valid combinatorial compounds are colored as following: compound **4** in green, compound **5** in magenta, compound **6** in yellow, and compound **10** in cyan. The key attachment point is Cys25 and the amino acid residues Gly66 and Asn161 are important for the stabilization of the ligand through H-bonding.

In comparison with the basic piperazino-substituted P3 side chain of Balicatib, the final four selected compounds possess significantly weaker basic P3 side chain, which was expected to reduce the off-target effects of these compounds during their application.

Synthesis of the Selected *N*-(Functionalized benzoyl)-homocycloleucyl-glycinonitriles. These compounds derive from homocycloleucyl-glycinonitrile (**3**; Scheme 1). This was prepared in 51% overall yield via the *N*-carboxylic acid cyclic anhydride **2**[38,39] derived from commercial homocycloleucine hydrochloride.

Three strategies were applied for amidification of **3**, depending on the benzoyl unit. Thus, compound **4** was prepared by coupling **3** with 4-(2-ethyl-butyramido)-3-fluorobenzoyl chloride in 58% yield under standard conditions of amide bond formation. Resorting to the corresponding 4-halo-isatoic anhydrides proved to yield **5** and **6** in 26−28% yield after simple trituration. Similarly, compound **7** was obtained in 73% yield using 7-chloro-2-trifluoromethyl-4*H*-3,1-benzoxazin-4-one. The latter was prepared from 4-chloroanthranilic acid upon treatment with trifluoroacetic anhydride.

However, the synthesis of the *N*-arylurea **10** necessitated a two-step conjugation as the preparation of **11** was unsuccessful. Thus, 3-fluoro-4-nitrobenzoyl chloride was coupled to **3** in 72% yield followed by Bechamp reduction of the nitro group into amino group which reacted in turn with 1-piperidinecarbonyl chloride to afford **10** (16% overall yield). Noteworthy, the reaction of 4-amino-3-fluorobenzoic acid with 1-piperidinecarbonyl chloride led to the piperidino aryl anhydride **12** or *N*-aroyl-piperidine **13**, depending on the choice of a neutral or acidic workup, respectively (Scheme 2).

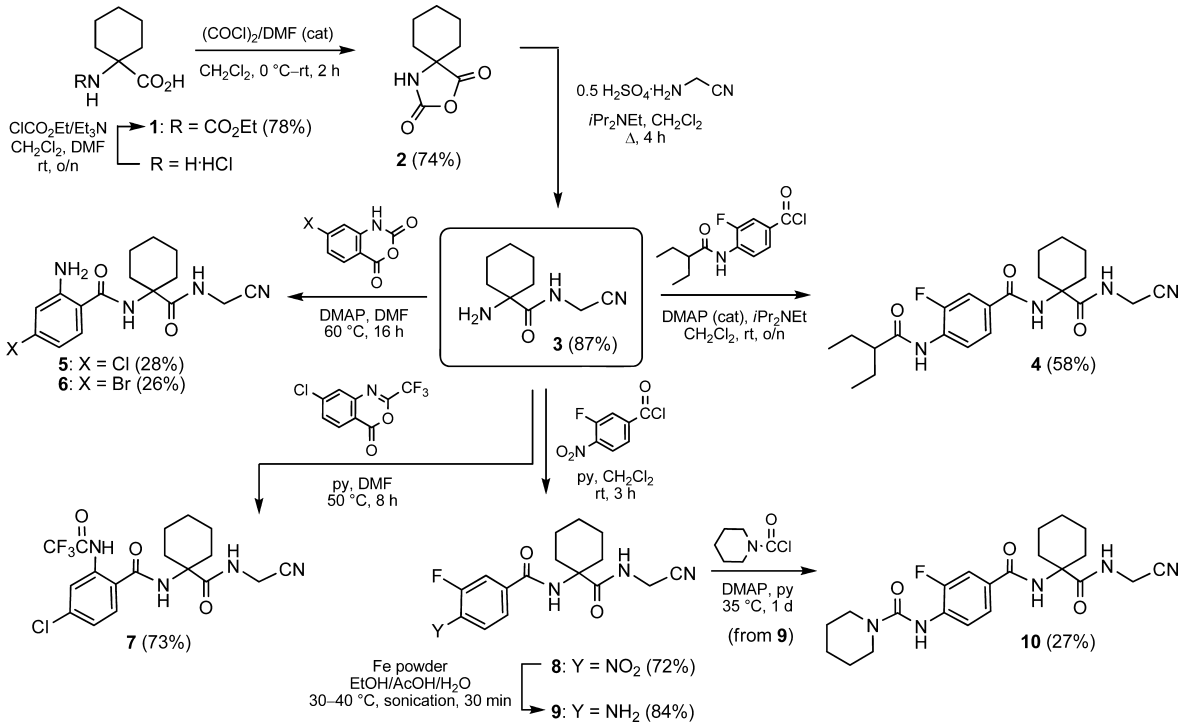
Biochemical Evaluation. The *N*-(functionalized benzoyl)-homocycloleucyl-glycinonitriles **4**−**10** (compounds **4**, **5**, **6**, and **10** were designated by the consensus model, and compounds **7**, **8**, and **9** correspond to their synthetic precursors) were tested for their binding to human Cat K, Cat L, and Cat S (Table 2). Compounds **5**, **6**, and **9** exhibited high affinity for Cat K with K_i values between 10 and 30 nM, and were highly selective against Cat L and Cat S. These compounds then underwent a more thorough kinetic investigation. It is worth mentioning that compound **6** has at least 3-fold greater selectivity ratio over Cat L compared to Balicatib (Table 2).[40]

As nitriles are known to be reversible inhibitors of cathepsins, we next checked the reversibility of Cat K inhibition by compounds **5**, **6**, and **9**, as described earlier for Cat K by rapid 100-fold dilution of the preformed enzyme−inhibitor complexes into the buffer solution containing the substrate.[41] The irreversible cysteine Cat inhibitor E-64[42] was used as a test control. As shown in Figure S2, upward curvatures were observed for all three compounds, characteristic for reversible inhibitors, but not for E-64, in agreement with previous findings. Finally, we monitored kinetics of binding of compounds **5**, **6**, and **9** to Cat K. As shown in Figure S3, all the curves exhibited a typical biphasic behavior slowly reaching the steady state, which is clearly seen at lower inhibitor concentrations (100−250 nM). Moreover, this also confirmed reversibility of the compounds. However, the rapid interaction between the inhibitors and Cat K precluded us from accurately determining the association (k_{on}) and dissociation (k_{off}) rate constants, as substantial parts of the initial portions of the curves were lost within the dead time of the method at inhibitor concentrations ≥250 nM. Therefore, only the curves obtained at 100 nM inhibitor concentration could have been used in the calculations, yielding estimates for the k_{on} value of ∼3×10^4 M^{-1} s^{-1} and for the k_{off} value of ∼1×10^{-4} s^{-1} for all three compounds. The slow binding kinetic of nitrile inhibitors to Cat K has already been observed previously for azadipeptide nitriles, although the k_{on} values for the latter were generally slightly higher (1×10^5 M^{-1} s^{-1} to 1×10^6 M^{-1} s^{-1}).[38]

Ligand−Cat K Structures. The most promising compounds **5**, **6**, and **9** were then subjected to structural analysis using X-ray crystallography. The complex of Cat K with compound **5** crystallized in the monoclinic space group *P*2(1), whereas the complexes of Cat K with compounds **6** and **9**

DOI: 10.1021/acs.jmedchem.5b00746
J. Med. Chem. 2015, 58, 6928−6937

Scheme 1. Preparation of *N*-(Functionalized benzoyl)-homocycloleucyl-glycinonitriles 4−10



Scheme 2. Attempt to Prepare Functionalized Benzoic Acid 11

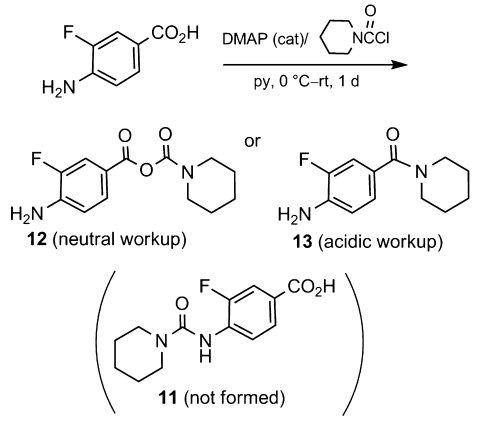


Table 2. Inhibition of Cats K, L, and S by *N*-(Functionalized benzoyl)-homocycloleucyl-glycinonitriles 4−10 at pH 6.0 and 37 °C[a]

	K_i (nM)		
compound	Cat K	Cat L	Cat S
4	70	n.i.[b]	4000
5	16 ± 2	10000	n.i.[b]
6	10 ± 3	11000	n.i.[b]
7	>500	1200	n.i.[b]
8	114	n.i.[b]	n.i.[b]
9	22 ± 1	720	5000
10	55	n.i.[b]	1000
Balicatib[c]	1.4	503	65000

[a]Experimental conditions were as described in the Experimental Section. [b]No inhibition observed even at 100 μM. [c]IC_{50} (nM) values are listed for Balicatib.

crystallized in the orthorhombic space group *P*2(1)2(1)2(1). All three complexes have one molecule of Cat K in the asymmetric unit. In all three structures the protein residues as well as the inhibitors are unambiguously defined by the electron density maps. The crystals of compound **9** complex diffracted to the atomic resolution of 1 Å. They bind into the active site of Cat K in the nonprimed region (Figure 3) occupying the S1, S2, and S3 binding sites.

Despite all three compounds have the same structure at the P1 region, the nitrile group of the compound **5** formed a covalent bond with the reactive site S gamma atom of Cys25, whereas compound **6** did not. In this latter case, the nitrile group points away from the Cys25 reactive site. Interestingly, compound **9** exhibited a dual behavior. Most of the compound **9** (79%) formed a covalent bond with the reactive site SG atom, whereas the rest did not; its nitrile group is clearly chemically unaltered and is positioned similarly to that of compound **6**. Superimposition of determined Cat K complex structures revealed that the positions of the pair of reactive site Cys25 residues forming the covalent interaction (compound **5** and 79% of compound **9**) and the pair that does not form it (compound **6** and 21% of compound **9**) superimpose tightly and are distinct from each other (the catalytic sulfur atoms are positioned approximately 0.5 Å apart; Figure 4). Due to the high resolution of the diffraction data we can conclude that these differences are real. Besides, according to the atomic distance (3.0 and 3.1 Å), the covalently bound nitrile groups form hydrogen bonds with the O atom of Gln19 of Cat K, indicating a flip of the Gln19 side chain and the change in the oxyanion hole electrostatic potential.

The cyclohexyl groups of all three inhibitors bind in the S2 pocket in the same orientation, yet the covalently and noncovalently bound compounds exhibit the same positional shift as the glycinonitrile part. The carbonyl O atom of the homocycloleucyl residue is oriented toward the peptide bond N atom of Gly66 and the edge of the indole ring of Trp26.

DOI: 10.1021/acs.jmedchem.5b00746
J. Med. Chem. 2015, 58, 6928−6937

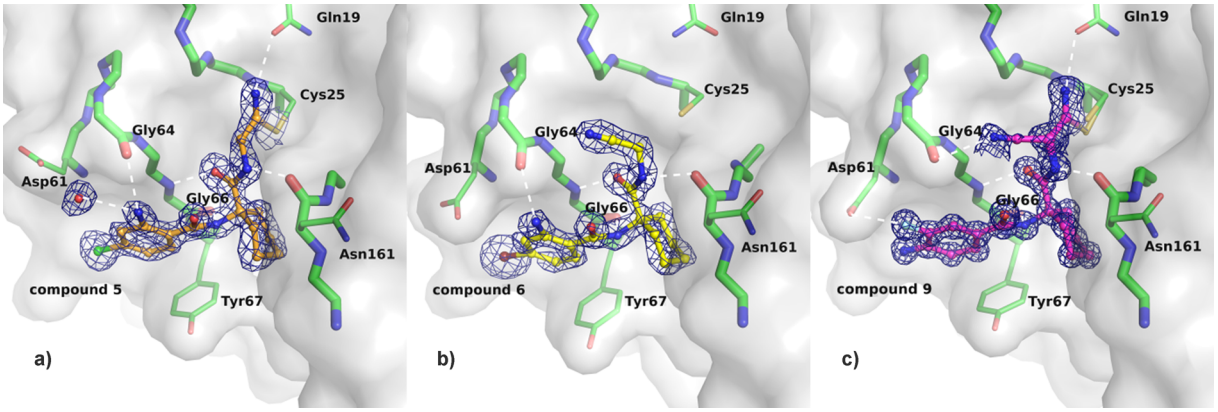


Figure 3. Human Cat K–inhibitor complexes. The blue electron density map is contoured: (a) around compound **5** at 2.0σ; (b) around compound **6** at 1.0σ; (c) around compound **9** at 1.5σ.

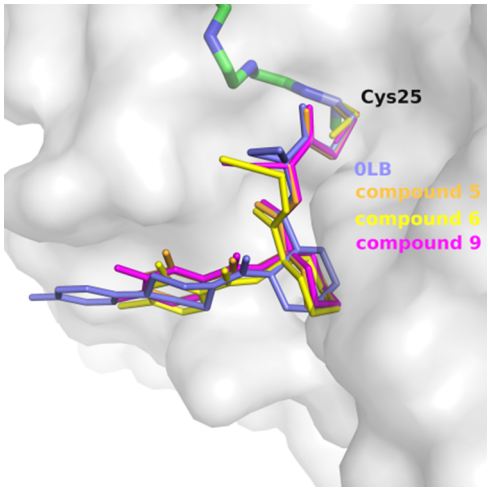


Figure 4. Superimposition of crystal structures of compounds **5**, **6**, and **9** on 0LB (PDB ID: 4DMX).

Each benzoyl group is positioned in the S3 pocket slightly differently. In the compound **5**, the interaction between the N atom at the meta position and Asp61 is mediated by a solvent molecule, while the F atom at the meta position in compound **9** is only 2.5 Å away from the carboxylic O atom of Asp61. The differences and similarities in the interaction patterns of compounds **5**, **6**, and **9** offer no explanation why Asp61 has been observed in two different conformations. It seems more likely that it is the change in the crystal lattice which leads to a different interaction pattern with the Asp61 side chain. Namely, the position of Asp61 is such that in the orthorhombic crystals it is tightly packed between two asymmetric molecules of Cat K, whereas in the monoclinic space group it is not.

The binding of inhibitors is characterized by several H-bonds (Figures 3 and S4). They all form H-bonds with the main chain carbonyl of Asn161, the amide group of Gly66, and the edge of Trp26. The covalently bound inhibitors form an additional interaction with the O atom of amide side chain of Gln19, whereas the *ortho* N atoms of the benzoyl rings of the compounds **5** and **6** form H-bonds with the carbonyl of Gly64, and the F atom in the compound **9** is positioned 2.5 Å away from the Asp61 carboxylic group.

The three crucial H-bonds between the inhibitor and Cat K Gly66 and Asn161 are consistent with the prediction made by GOLD molecular docking. The H-bond lengths between the inhibitor and the Cat K are listed in Table S3.

CONCLUSION

Integrated CADD approaches encompassing virtual combinatorial chemistry, QSAR modeling, and molecular docking proved to be powerful for the design of compounds with targeted properties. As a matter of fact, a selected 4-membered series of *N*-(functionalized benzoyl)-homocycloleucyl-glycinonitriles (compounds **5**, **6**, **9**, and **10**) containing a weakly to nonbasic P3 moiety, showed high selectivity for Cat K when compared with cathepsins L and S with the K_i values in the 10–30 nM range. The kinetic studies revealed reversible and tight binding of these compounds to Cat K, whereas X-ray structural studies of their complexes showed covalent and noncovalent type binding between the nitrile group of the compound and the catalytic cysteine (Cys25) site. From the presented data, it could be hypothesized in this case that the establishment of a covalent bond is not essential for an inhibitor's potency.

EXPERIMENTAL SECTION

Virtual Combinatorial Chemistry Generation. Pipeline Pilot's reaction enumeration module (BIOVIA, U.S.A.) was employed for combinatorial generation of *N*-(functionalized benzoyl)-homocycloleucyl-glycinonitriles. All structures with PSA > 140 Å^2 and number of rotatable bonds (nRB) > 10 were removed.

Molecular Docking Calculations. The covalent molecular docking studies were performed by GOLD Suite v5.1 docking package.[43] Binding site was defined around experimental coordinates of the 0LB[36] ligand in radius of 6 Å. In order to perform covalent docking, dummy sulfur linking atom was attached to each ligand. Subsequently, ligands were renumbered in a way that previously added sulfur atom number was set to 1. Cat K sulfur atom on Cys25 was then matched with dummy sulfur atom of ligand. Preparing ligands in this way was necessary for GOLD calculation to recognize and set the correct covalent bond between predefined ligands and Cat K linking atoms.[41,44,45] Redocking procedure was performed by docking reference ligand 0LB into the binding site with early termination allowed if the top three solutions were within 1.5 Å of the rmsd value in order to define optimal parameters and scoring function.[46] Finally, each ligand was docked into the binding site by the selected scoring function and applying the following parameters of the GOLD genetic algorithm: population size = 100, selection pressure = 1.1, No. of operations = 100 000, No. of islands = 5, niche size = 2, migrate = 10, mutate = 95, crossover = 95.[43]

General Synthetic Procedures. Reactions were conducted under an inert atmosphere using anhydrous solvents when required. Analytical (TLC) and preparative (PLC) layer chromatographies

DOI: 10.1021/acs.jmedchem.5b00746
J. Med. Chem. 2015, 58, 6928–6937

were performed on Silica Gel 60F_{254} plates. TLC R_f values are reported and visualization was accomplished with an UV lamp (254 nm) and staining with $KMnO_4$ solution. Filtration over a bed of silica gel was performed using Silica Gel 60 (40−63 μm). Melting points were determined on a Kofler apparatus and are uncorrected. ^{1}H (300 MHz, internal Me_4Si), ^{13}C (75 MHz, internal $CDCl_3$), and ^{19}F NMR (282 MHz, internal $CFCl_3$) were recorded as solutions in either $CDCl_3$ or DMSO-d_6. IR spectra were recorded using ATR cell equipped with a diamond crystal. HRMS measurements were obtained on a Waters Micromass Q-TOF Premier instrument equipped with an orthogonal Z-spray ESI interface. The chemical purity of all compounds **4**−**10** for biochemical evaluation was confirmed to be >95% based on elemental analysis which matched within ±0.4% of the calculated values.

Materials. Homocycloleucine hydrochloride, glycinonitrile sulfate, 4-amino-3-fluorobenzoic acid, 3-fluoro-4-nitrobenzoic acid, 4-chloroanthranilic acid, and 4-bromoanthranilic acid are commercially available.

*N-Ethoxycarbonyl-homocycloleucine (**1**).* To a cold (0 °C) mixture of homocycloleucine hydrochloride (5.00 g, 27.8 mmol) and Et_3N (13.5 mL, 96.9 mmol) in DMF (10 mL)/CH_2Cl_2 (90 mL) was added dropwise ethyl chloroformate (4.0 mL, 41.8 mmol). After stirring overnight, the mixture was concentrated, diluted with water (100 mL), acidified to pH 2 using 1 M HCl (ca. 8 mL), and extracted with EtOAc (2 × 100 mL). Filtration through a bed of silica gel/Na_2SO_4 and concentration (at 15 mbar followed by 0.1 mbar) afforded colorless syrup which was triturated with iPr_2O and the solidified product filtered, yielding a white powder (4.70 g, 78.5%). Mp 90−92 °C; ^{1}H NMR spectroscopy revealed two conformational isomers in 65:35 ratio. ^{1}H NMR (DMSO-d_6): δ 12.19 (s, 1H), 7.23 (s, 1H), 3.99 (m, 2H), 1.93 (m, 2H), 1.84−1.59 (m, 2H), 1.59−1.38 (m, 5H), 1.19 (m, 4H); ^{13}C NMR (DMSO-d_6): δ 176.5, 169.7, 155.8, 60.3, 59.79, 59.72, 32.2, 31.8, 25.4, 25.1, 21.4, 21.1, 15.0. HRMS (ESI): m/z calcd for $C_{10}H_{17}NO_4$ + H^+ [M + H^+], 216.1236; found, 216.1238.

*Homocycloleucine N-Carboxylic Acid Anhydride (**2**).* To a cold solution (0 °C) of *N*-EtOCO-homocycloleucine (**1**; 4.00 g, 18.6 mmol) in dry CH_2Cl_2 (60 mL) containing DMF (1 mL) was added dropwise oxalyl chloride (2.4 mL, 28.4 mmol). After stirring at rt for 2 h, the mixture was concentrated, dissolved in a small amount of CH_2Cl_2, and filtered through silica gel eluting with hexane/EtOAc (1:1). The residue was triturated with hexane, yielding white crystals (2.33 g, 74.2%). ^{1}H NMR: δ 7.62 (s, 1H), 2.05−1.60 (m, 8H), 1.60−1.34 (m, 2H). ^{1}H NMR data were consistent with those reported in the literature.[38,39]

*Homocycloleucyl-glycinonitrile (**3**).* To a solution of the anhydride **2** (4.00 g, 23.6 mmol) and iPr_2NEt (29 mL, 0.165 mol) in dry CH_2Cl_2 (100 mL) was added glycinonitrile sulfate (3.32 g, 15.8 mmol). After heating at reflux for 4 h, the mixture was concentrated and partitioned between EtOAc (100 mL)/water (100 mL). The water layer was re-extracted with EtOAc (2 × 100 mL) and the combined EtOAc phases were washed with satd aq $NaHCO_3$ (30 mL) and brine (100 mL), dried (Na_2SO_4), and concentrated. The oily residue was redissolved in a small amount of CH_2Cl_2 and filtered through silica gel eluting with CH_2Cl_2/MeOH (95:5) to afford a yellowish viscous syrup (3.73 g, 87.0%). ^{1}H NMR ($CDCl_3$): δ 8.37 (br s, 1H), 4.17 (d, J = 6.0 Hz, 2H), 2.02−1.92 (m, 2H), 1.83−1.57 (m, 3H), 1.57−1.10 (m, 7H). ^{1}H NMR data were consistent with those reported in the literature.[38]

4-Chloroisatoic Anhydride. To a solution of 4-chloroanthranilic acid (0.500 g, 2.91 mmol) in THF (10 mL) was added triphosgene (0.294 g, 0.99 mmol) under nitrogen. After stirring at rt overnight, the solid was filtered and washed with Et_2O (2 × 10 mL), affording white crystals (0.543 g, 94.4%). ^{1}H NMR (DMSO-d_6): δ 11.83 (s, 1H), 7.92 (d, J = 8.5 Hz, 1H), 7.29 (dd, J = 8.5 and 2.0 Hz, 1H), 7.14 (d, J = 1.9 Hz, 1H). ^{1}H NMR data were consistent with those reported in the literature.[47]

4-Bromoisatoic Anhydride. Prepared from 4-bromoanthranilic acid (0.629 g, 2.91 mmol) following a similar procedure as for 4-chloroisatoic anhydride, affording white fluffy crystals (0.574 g, 81.4%). ^{1}H NMR (DMSO-d_6): δ 11.81 (s, 1H), 7.83 (d, J = 8.4 Hz, 1H), 7.43 (dd, J = 8.4 and 1.8 Hz, 1H), 7.30 (d, J = 1.8 Hz, 1H). ^{1}H NMR data were consistent with those reported in the literature.[48]

7-Chloro-2-trifluoromethyl-4H-3,1-benzoxazin-4-one. To trifluoroacetic anhydride (5 mL) was added in portions 4-chloroanthranilic acid (0.343 g, 2.00 mmol). After stirring at rt for 3 h, the mixture was concentrated, dissolved in dry Et_2O (ca. 10 mL), and filtered under nitrogen to remove the insoluble yellow material. Concentration afforded the title compound as an off-white solid (0.505 g, 94.4%) in 95+% purity, as determined by ^{19}F NMR. Mp 53−54 °C (lit.:[49] 54.2−54.8 °C (*n*-hexane)). ^{1}H NMR: δ 8.22 (d, J = 8.5 Hz, 1H), 7.79 (d, J = 1.9 Hz, 1H), 7.67 (dd, J = 8.5 and 2.0 Hz, 1H). ^{19}F NMR: δ −65.3 (s).

N-Trifluoroacetyl-4-chloroanthranilic Acid. A mixture of 7-chloro-2-trifluoromethyl-4*H*-3,1-benzoxazin-4-one (50 mg, 0.20 mmol) in water (5 mL) was heated at 60 °C for 20 min. Upon cooling to rt, the formed precipitate was filtered and recrystallized from water/MeOH, affording white needles (0.490 g, 91.6%). Mp 165−166 °C (lit.:[49] 177.5−178.5 °C (benzene)). ^{1}H NMR: δ 12.05 (s, 1H), 8.77 (d, J = 2.0 Hz, 1H), 8.14 (d, J = 8.6 Hz, 1H), 7.28 (dd, J = 8.6 and 2.1 Hz, 1H). ^{13}C NMR: δ 171.1, 155.3 (q, J = 38 Hz), 142.6, 140.2, 133.1, 125.3, 120.9, 115.3 (q, J = 310 Hz), 113.6. ^{19}F NMR: δ −69.0 (d, J = 0.9 Hz). HRMS (ESI): m/z calcd for $C_9H_5ClF_3NO_3$ − H^- [M − H^-], 265.9832; found, 265.9840.

4-(2-Ethyl-butyramido)-3-fluorobenzoic Acid. To a cold (0 °C) solution of 4-amino-3-fluorobenzoic acid (0.50 g, 3.22 mmol) in dry pyridine (2.5 mL) was added dropwise 2-ethylbutyryl chloride (0.60 mL, 4.31 mmol). After stirring at rt overnight, water (5 mL) was added then acidified to pH 2 using *c.* HCl. The precipitate was filtered and recrystallized from MeCN affording an off-white powder (0.47 g, 57.6%). R_f 0.61 (EtOAc); mp 205−206 °C. ^{1}H NMR (DMSO-d_6): δ 13.08 (br s, 1H), 9.90 (s, 1H), 8.10 (t, J = 8.1 Hz, 1H), 7.76−7.67 (m, 2H), 1.70−1.31 (m, 4H), 0.97−0.75 (m, 6H). ^{13}C NMR (DMSO-d_6): δ 174.8, 166.0, 152.6 (d, J = 242 Hz), 130.5 (d, J = 12 Hz), 127.0 (d, J = 7 Hz), 125.7 (d, J = 3 Hz), 123.3, 116.0 (d, J = 21 Hz), 48.9, 25.2, 11.7. ^{19}F NMR (DMSO-d_6): δ −116.7 (m). HRMS (ESI): m/z calcd for $C_{13}H_{16}FNO_3$ − H^- [M − H^-], 252.1036; found, 252.1032.

*N-[4-(2-Ethyl-butyramido)-3-fluorobenzoyl]homocycloleucyl-glycinonitrile (**4**).* To a cold (0 °C) solution of 4-(2-ethyl-butyramido)-3-fluorobenzoic acid (0.385 g, 1.52 mmol) in CH_2Cl_2 (5 mL) containing DMF (1 mL) was added dropwise oxalyl chloride (0.17 mL, 2.01 mmol). After stirring at rt for 1 h, it was partially concentrated and added dropwise to a cold solution (0 °C) of homocycloleucyl-glycinonitrile (**3**; 0.286 g, 1.579 mmol), DMAP (a crystal), and iPr_2NEt (0.55 mL, 3.1 mmol) in CH_2Cl_2 (5 mL). The mixture was stirred at rt overnight and concentrated to dryness. The residue was partitioned between EtOAc (20 mL) containing few drops of MeOH and satd. aq. $NaHCO_3$ (15 mL). The organic phase was washed with brine (10 mL), filtered through a bed of silica gel/$MgSO_4$, and concentrated. The residue was recrystallized from EtOAc, affording an off-white powder (0.365 g, 57.7%). R_f 0.78 (EtOAc); mp 220−221 °C. ^{1}H NMR (DMSO-d_6): δ 9.83 (s, 1H), 8.21 (t, J = 5.7 Hz, 1H), 8.15−7.91 (m, 2H), 7.80 (m, 1H), 7.70 (m, 1H), 4.04 (d, J = 5.6 Hz, 2H), 2.13−2.09 (m, J = 12.8 Hz, 2H), 1.75−1.72 (m, J = 8.5 Hz, 2H), 1.66−1.37 (m, 9H), 1.29−1.25 (m, 1H), 0.87 (t, J = 7.4 Hz, 6H). ^{13}C NMR (DMSO-d_6): δ 174.8, 174.6, 164.7, 152.9 (d, J = 242 Hz), 131.2 (d, J = 6 Hz), 128.8 (d, J = 12 Hz), 124.0 (d, J = 3 Hz), 123.4, 117.8, 115.0 (d, J = 21 Hz), 59.4, 49.0, 31.7, 27.6, 25.2, 25.0, 21.2, 11.8. ^{19}F NMR (DMSO-d_6): δ −117.1 (m). HRMS (ESI): m/z calcd for $C_{22}H_{29}FN_4O_3$ + H^+ [M + H^+], 417.2302; found, 417.2300. Elem anal. Calcd for $C_{22}H_{29}FN_4O_3$ (%): C, 63.44; H, 7.02; N, 13.45. Found: C, 63.18; H, 6.78; N, 13.43.

*N-(2-Amino-4-chlorobenzoyl)homocycloleucyl-glycinonitrile (**5**).* To a solution of homocycloleucyl-glycinonitrile (**3**; 0.181 g, 1.00 mmol) in DMF (1 mL) was added DMAP (37 mg, 0.30 mmol) and 4-chloroisatoic anhydride (0.198 g, 1.00 mmol). After stirring at 60 °C for 16 h, the mixture was concentrated, dissolved in MeCN (1 mL), and added dropwise onto 5% aq $NaHCO_3$ (5 mL). The precipitated product was filtered, rinsed successively with water, Et_2O, and hexane, and then dried at 60 °C for 5 h, affording an off-white powder (0.095 g, 28.4%). Mp 219−220 °C. ^{1}H NMR (DMSO-d_6): δ 8.17 (t, J = 5.6 Hz, 1H), 7.83 (s, 1H), 7.56 (d, J = 8.4 Hz, 1H), 6.90 (d, J = 1.9 Hz, 1H), 6.68 (dd, J = 8.4 and 1.9 Hz, 1H), 6.47 (s, 2H), 4.05 (d, J = 5.6

DOI: 10.1021/acs.jmedchem.5b00746
J. Med. Chem. 2015, 58, 6928−6937

Hz, 2H), 2.07−2.03 (m, 2H), 1.73−1.70 (m, 2H), 1.51 (m, 5H), 1.28−1.25 (m, 1H). ^{13}C NMR (DMSO-d_6): δ 175.1, 168.0, 150.9, 136.2, 131.1, 117.8, 114.8, 114.0, 113.8, 59.0, 31.8, 27.6, 25.1, 21.3. HRMS (ESI): m/z calcd for $C_{16}H_{19}{}^{35}ClN_4O_2$ + H^+ [M + H^+], 335.1275; found, 335.1269. Elem anal. Calcd for $C_{16}H_{19}ClN_4O_2$ (%): C, 57.40; H, 5.72; N, 16.73. Found: C, 57.07; H, 5.38; N, 16.49.

*N-(2-Amino-4-bromobenzoyl)homocycloleucyl-glycinonitrile (**6**).* Prepared from homocycloleucyl-glycinonitrile (**3**; 0.181 g, 1.00 mmol) and 4-bromoisatoic anhydride (0.242 g, 1.00 mmol) following a similar procedure as for **5**, affording off-white powder (0.100 g, 26.4%). Mp 216−217 °C. ^{1}H NMR (DMSO-d_6): δ 8.16 (t, J = 5.7 Hz, 1H), 7.83 (s, 1H), 7.56 (d, J = 8.4 Hz, 1H), 6.89 (d, J = 2.0 Hz, 1H), 6.68 (dd, J = 8.4 and 2.0 Hz, 1H), 6.47 (s, 2H), 4.05 (d, J = 5.6 Hz, 2H), 2.07−2.03 (m, 2H), 1.73−1.70 (m, 2H), 1.51 (s, 5H), 1.28−1.25 (m, 1H). ^{13}C NMR (DMSO-d_6): δ 175.1, 168.1, 150.9, 131.2, 125.1, 117.79, 117.76, 116.8, 114.2, 59.0, 31.7, 27.6, 25.1, 21.3. HRMS (ESI): m/z calcd for $C_{16}H_{19}{}^{79}BrN_4O_2$ + H^+ [M + H^+], 379.0770; found, 379.0766. Elem anal. Calcd for $C_{16}H_{19}BrN_4O_2$ (%): C, 50.67; H, 5.05; N, 14.77. Found: C, 50.35; H, 4.65; N, 14.68.

*N-(2-Trifluoroacetamido-4-chlorobenzoyl)homocycloleucyl-glycinonitrile (**7**).* To a solution of homocycloleucyl-glycinonitrile (**3**; 0.180 g, 1.00 mmol) in DMF (1 mL) was added 7-chloro-2-trifluoromethyl-4*H*-3,1-benzoxazin-4-one (0.194 g, 0.78 mmol) and pyridine (0.100 mL, 1.24 mmol). After stirring at 50 °C for 8 h, the mixture was concentrated, suspended in EtOAc (10 mL), washed successively with satd aq NH_4Cl (2 × 10 mL), satd aq $NaHCO_3$ (10 mL), and brine (10 mL), and filtered through a bed of silica gel/Na_2SO_4. The residue was triturated with Et_2O and filtered, affording a white powder (0.245 g, 73.0%). Mp 191−192 °C. ^{1}H NMR (DMSO-d_6): δ 12.29 (s, 1H), 8.67 (s, 1H), 8.28 (t, J = 5.5 Hz, 1H), 8.20 (d, J = 1.8 Hz, 1H), 7.98 (d, J = 8.5 Hz, 1H), 7.50 (dd, J = 8.4 and 1.9 Hz, 1H), 4.02 (d, J = 5.6 Hz, 2H), 2.10−2.05 (m, 2H), 1.77−1.74 (m, 2H), 1.55 (m, 5H), 1.30−1.27 (m, 1H). ^{13}C NMR (DMSO-d_6): δ 174.3, 167.3, 154.56 (q, J = 37 Hz), 136.9, 136.3, 131.3, 125.0, 122.7, 121.3, 117.4, 115.5 (q, J = 292 Hz), 60.1, 31.5, 27.6, 25.0, 21.2. ^{19}F NMR (DMSO-d_6): δ −68.0 (s). ν_{max}: 3344, 2935, 2870, 1729, 1662, 1584, 1518, 1452, 1412 cm^{-1}. HRMS-ESI (m/z): calcd for $C_{18}H_{18}ClF_3N_4O_3$ + H^+ [M + H^+], 431.1098; found, 431.1091. Elem anal. Calcd for $C_{18}H_{18}ClF_3N_4O_3$ (%): C, 50.18; H, 4.21; N, 13.01. Found: C, 50.43; H, 3.81; N, 12.99.

*N-(3-Fluoro-4-nitrobenzoyl)homocycloleucyl-glycinonitrile (**8**).* To a cold (0 °C) solution of 3-fluoro-4-nitrobenzoic acid (0.500 g, 2.70 mmol) in CH_2Cl_2 (5 mL) containing DMF (0.2 mL) was added dropwise oxalyl chloride (0.34 mL, 4.05 mmol). After stirring at rt overnight, it was concentrated to about 1/5 of its volume, diluted with CH_2Cl_2 (5 mL), and added dropwise onto a cold (0 °C) solution of homocycloleucyl-glycinonitrile (**3**; 0.510 g, 2.81 mmol) in pyridine (2 mL). The mixture was stirred at rt for 3 h and poured onto water (30 mL). The precipitate was filtered, washed with water, dissolved in hot MeCN (30 mL), and filtered to remove the brown insoluble matter. Water (40 mL) was added, and the precipitate filtered and triturated with hexane/EtOAc (1:1), affording an off-white powder (0.675 g, 71.8%). R_f 0.71 (EtOAc); mp 216−217 °C. ^{1}H NMR (DMSO-d_6): δ 8.41 (s, 1H), 8.31 (t, J = 5.5 Hz, 1H), 8.25 (t, J = 8.1 Hz, 1H), 8.06 (dd, J = 12.1 and 1.3 Hz, 1H), 7.88 (d, J = 8.5 Hz, 1H), 4.06 (d, J = 5.5 Hz, 2H), 2.13−2.08 (m, 2H), 1.77−1.74 (m, 2H), 1.53 (m, 5H), 1.29−1.26 (m, 1H). ^{13}C NMR (DMSO-d_6): δ 174.4, 163.7, 154.1 (d, J = 257 Hz), 141.6 (d, J = 7 Hz), 138.2 (d, J = 8 Hz), 126.2 (d, J = 3 Hz), 124.6 (d, J = 4 Hz), 118.0, 117.7, 59.9, 31.6, 27.6, 25.0, 21.2. ^{19}F NMR (DMSO-d_6): δ −111.8 (dd, J = 12.0 and 7.7 Hz). HRMS (ESI): m/z calcd for for $C_{16}H_{17}FN_4O_4$ + H^+ [M + H^+], 349.1312; found, 349.1304. Elem anal. Calcd for $C_{16}H_{17}FN_4O_4$ (%): C, 55.17; H, 4.92; N, 16.08. Found: C, 55.48; H, 4.64; N, 16.10.

*N-(3-Fluoro-4-aminobenzoyl)homocycloleucyl-glycinonitrile (**9**).* To compound **8** (0.385 g, 1.105 mmol) in EtOH (4 mL), AcOH (4 mL), and water (2 mL) was added iron powder (0.370 g, 6.63 mmol), and the mixture was heated at 30−40 °C under ultrasound for 1 h. The red-brown mixture was poured onto satd aq $NaHCO_3$ (12 mL) and extracted with EtOAc (10 mL) containing MeOH (2 mL). The organic layer was filtered through a bed of silica gel/Na_2SO_4 and concentrated, affording off-white crystals (0.295 g, 83.8%). R_f 0.76 (EtOAc); mp 218−219 °C. ^{1}H NMR (DMSO-d_6): δ 8.13 (t, J = 5.6 Hz, 1H), 7.75−7.56 (m, 2H), 7.49 (dd, J = 8.3 and 1.9 Hz, 1H), 6.75 (t, J = 8.7 Hz, 1H), 5.70 (s, 2H), 4.03 (d, J = 5.6 Hz, 2H), 2.11−2.07 (m, 2H), 1.75−1.69 (m, 2H), 1.51 (m, 5H), 1.27−1.24 (m, 1H). ^{13}C NMR (DMSO-d_6): δ 175.2, 165.2, 149.3 (d, J = 232 Hz), 139.7 (d, J = 13 Hz), 125.0, 121.6 (d, J = 5 Hz), 117.8, 114.6 (d, J = 19 Hz), 114.4, 59.0, 31.7, 27.6, 25.1, 21.2. ^{19}F NMR (DMSO-d_6): δ −128.9 (m). HRMS (ESI): m/z calcd for $C_{16}H_{19}FN_4O_2$ + H^+ [M + H^+], 319.1570; found, 319.1566. Elem anal. Calcd for $C_{16}H_{19}FN_4O_2$ (%): C, 60.37; H, 6.02; N, 17.60. Found: C, 60.39; H, 5.65; N, 17.47.

*N-[3-Fluoro-4-(piperidinocarboxamido)benzoyl]-homocycloleucyl-glycinonitrile (**10**).* To a solution of compound **9** (50 mg, 0.157 mmol) and DMAP (37 mg, 0.30 mmol) in pyridine (1 mL) was added dropwise 1-piperidinecarbonyl chloride (0.10 mL, 0.80 mmol). After stirring at 35 °C for 1 day, the mixture was concentrated and partitioned between EtOAc (10 mL) containing a few drops of MeOH and 1 mM HCl (10 mL). The organic phase was washed with brine (10 mL), dried (Na_2SO_4), and concentrated. The residue was purified on a PLC plate eluting with EtOAc. The product was washed from silica gel using EtOAc/MeOH (10:1) and triturated with Et_2O, affording an off-white powder (18 mg, 26.8%). R_f 0.73 (EtOAc); mp 195−196 °C. ^{1}H NMR (DMSO-d_6): δ 8.39 (s, 1H), 8.20 (t, J = 5.4 Hz, 1H), 7.95 (s, 1H), 7.82−7.47 (m, 3H), 4.04 (d, J = 5.5 Hz, 2H), 3.41 (m, 4H), 2.13−2.08 (m, 2H), 1.74−1.72 (m, 2H), 1.72−1.30 (m, 11H), 1.29−1.26 (m, 1H). ^{13}C NMR (DMSO-d_6): δ 174.9, 164.8 (d, J = 2.0 Hz), 154.4, 153.7 (d, J = 241 Hz), 131.0 (d, J = 11 Hz), 130.1 (d, J = 6 Hz), 123.9 (d, J = 1 Hz), 123.7 (d, J = 2 Hz), 117.8, 114.9 (d, J = 21 Hz), 59.4, 44.8, 31.7, 27.6, 25.5, 25.1, 24.0, 21.2. ^{19}F NMR (DMSO-d_6): δ −116.6 (dd, J = 11.5 and 7.8 Hz). HRMS (ESI): m/z calcd for $C_{22}H_{28}FN_5O_3$ + H^+ [M + H^+], 430.225; found, 430.224. Elem anal. Calcd for $C_{22}H_{28}FN_5O_3$ (%): C, 61.52; H, 6.57; N, 16.31. Found: C, 61.18; H, 6.31; N, 16.22.

*N-(4-Amino-3-fluorobenzoyloxycarbonyl)piperidine (**12**).* To a cold (0 °C) solution of 4-amino-3-fluorobenzoic acid (0.50 g, 3.22 mmol) in dry pyridine (2.5 mL) was added dropwise 1-piperidinecarbonyl chloride (0.55 mL, 4.32 mmol). After stirring at rt overnight, the mixture was concentrated and partitioned between EtOAc (10 mL) containing a few drops of MeOH and 0.01 M HCl (5 mL). The EtOAc phase was dried over $MgSO_4$ and concentrated, affording an amber colored syrup that solidified upon standing overnight (0.768 g, 91.2%). R_f 0.78 (EtOAc). ^{1}H NMR ($CDCl_3$): δ 7.77−7.61 (m, 2H), 6.76 (t, J = 8.6 Hz, 1H), 4.30 (s, 2H), 3.59 (m, 2H), 3.51−3.35 (m, 2H), 1.63 (m, 6H). Upon heating **12** in MeCN in the presence of DMAP (cat), *N*-(4-amino-3-fluorobenzoyl)piperidine (**13**) was formed quantitatively.

*N-(4-Amino-3-fluorobenzoyl)piperidine (**13**).* To a cold (0 °C) solution of 4-amino-3-fluorobenzoic acid (0.50 g, 3.22 mmol) in dry pyridine (2.5 mL) was added dropwise 1-piperidinecarbonyl chloride (0.55 mL, 4.32 mmol). After stirring at rt overnight, water (5 mL) was added, and then the solution was acidified to pH 2 using concd HCl. The oily precipitate was extracted with Et_2O (2 × 10 mL), filtered through a bed of silica gel/Na_2SO_4, and concentrated, affording an orange colored oil that solidified upon standing overnight (0.755 g, 88.0%). R_f 0.64 (EtOAc); mp 265−268 °C. ^{1}H NMR ($CDCl_3$): δ 7.18−6.96 (m, 2H), 6.74 (t, J = 9 Hz, 1H), 3.92 (br s, 2H), 3.53 (m, 4H), 1.70−1.50 (m, 6H). ^{13}C NMR ($CDCl_3$): δ 169.7, 150.7 (d, J = 240 Hz), 136.4 (d, J = 13 Hz), 126.0 (d, J = 6 Hz), 124.1 (d, J = 3 Hz), 116.1 (d, J = 4 Hz), 114.9 (d, J = 20 Hz), 26.2, 24.7.

In Vitro Inhibition Assay. Human Cat K, L, and S were expressed in *Pichia pastoris*, as described elsewhere.[50,51] Active enzyme concentration was determined by titration with E-64 (Peptide Institute, Japan). Cat activities were monitored using synthetic fluorogenic peptide substrates Z-Gly-Pro-Arg-AMC (Cat K), Z-Phe-Arg-AMC (Cat L), and Z-Phe-Val-Arg-AMC (Cat S; Bachem, Switzerland) at an excitation wavelength of 370 nm and an emission wavelength of 460 nm, as previously described.[52] All reactions were performed in 100 mM sodium phosphate buffer pH 6.0 in the presence of 1 mM DTT. For the inhibition assays, 96-well black flat bottom microplates were used (Greiner, Germany); all measurements were performed using a Tecan M1000 plate reader (Tecan,

DOI: 10.1021/acs.jmedchem.5b00746
J. Med. Chem. 2015, 58, 6928−6937

Switzerland). Different concentrations of inhibitor (10 nM to 10 μM) were mixed with Cat K, L, or S (5−50 nM) and incubated for 30 min at 37 °C to allow the formation of enzyme−inhibitor complex in a total volume of 100 μL. After the incubation, substrate was added in a negligible volume to a final concentration of 50 μM (10 μM for Cat L) and fluorescence was continuously monitored over 10 min. The initial velocities in the presence (v_i) and absence of the inhibitor (v_o) were then calculated from the initial linear portion of the curves by linear regression analysis. The determination of K_i values was then performed, as previously described.[53] Briefly, by plotting $v_o/v_i - 1 = [I]/K_{i,app}$, the $K_{i,app}$ values were obtained,[37,41,52] and K_i values were then calculated from the $K_{i,app}$ values using the equation $K_i = K_{i,app}/(1 + [S_0]/K_M)$, where the following K_M values were used to correct for substrate competition: 48 μM[54] for Cat K and Z-Gly-Pro-Arg-AMC as the substrate, 2 μM for Cat L and Z-Phe-Arg-AMC as the substrate, and 8 μM for Cat S and Z-Phe-Val-Arg-AMC as the substrate.[55]

Kinetics of Cat K Inhibition by Compounds 5, 6, and 9. Initially, reversibility of binding of compounds **5**, **6**, and **9** to Cat K was first evaluated as described elsewhere.[41] Briefly, the compounds (5 μM) and Cat K (1 μM) were incubated for 30 min, which is sufficient for completion of the reaction, before rapidly diluted 100-fold in the reaction buffer containing the substrate. Progress of the reaction was then monitored continuously as described above. Next, kinetics of binding of the three compounds to Cat K was studied to better understand the mechanistic aspects of the inhibition. All measurements were performed under the same conditions as for the inhibition assay. A total of 10 nM preactivated Cat K was added to a reaction buffer containing the substrate and varying concentrations of inhibitor (100 nM to 10 μM). The inhibitor concentration was always at least 10-fold higher as the enzyme concentration to ensure the pseudo-first-order reaction conditions. The progress curves were continuously recorded for 1.5 h and could be best fitted to the following equation by nonlinear regression analysis and the GraphPad Prism software (GraphPad Software, U.S.A.):

$$[P] = v_s t + (v_o - v_s)(1 - e^{-kt})/k$$

where $[P]$ is the product concentration, v_{oi} and v_s are the initial and steady-state reaction velocities, respectively, whereas k stands for the observed pseudo-first order rate constant describing the presteady state of the reaction.[56] On this basis it is possible to calculate association and dissociation rate constants, k_{on} and k_{off}, using the following relationships: $k = k_{on}[I_0]/(1 + [S_0]/K_M) + k_{off}$ that enables k_{on} and k_{off} using linear regression analysis. However, due to unfavorable reaction condition, estimates of k_{off} and k_{on} values were calculated using the following relationships:

$$k_{off} = k \times (v_s/v_o);$$

$$k_{on} = (k \times (1 - v_s/v_o))/([I_0]/(1 + [S_0]/K_M)),\ K_i = k_{off}/k_{on}$$

The equations were adapted from standard protocols published previously.[56,57]

X-ray Diffraction. The cDNA of human procathepsin K was obtained from Deutsches Ressourcenzentrum für Genomforschung. Expression of human Cat K: cDNA for human procathepsin K was amplified by PCR. PCR products were cloned into vector pPIC9 (Invitrogen). Recombinant plasmids were electroporated in *P. pastoris* strain GS115 (Invitrogen). Recombinant clones were selected according to Pichia Expression kit manual (Invitrogen). The clone with the highest protein expression was chosen and expressed according to the Invitrogen protocol.[58] Afterward, the expression crude medium was dialyzed to buffer_1 (50 mM sodium acetate, pH 5.5, 2.5 mM EDTA, 2.5 mM DTT). Dialyzed medium containing Cat K was concentrated to 350 mL. Procathepsin K was activated by adding pepsin to final concentration of 0.03 mg/mL and reducing buffer pH to 4.0. The activation took 4 h at 37 °C. Afterward, the protein was dialyzed to buffer_2 (50 mM sodium acetate, pH 5.5) and blocked reversibly with 10-fold molar excess of *S*-methylmethanethiosulfonate. The prepared protein solution was applied to 25 mL of conditioned SP sepharose. Cat K was eluted using NaCl gradient from 0 to 2 M NaCl in buffer_2. Fractions containing pure Cat K were collected and dialyzed again against buffer_2. Afterward, pure cat K was concentrated to a final concentration of 36 mg/mL.

Protein crystallization and crystal harvesting: For crystallization trials, the protein was activated by adding DTT to the 2 mM final concentration and incubated for 10 min. Afterward, the inhibitor was added to 2 mM final concentration and incubated for 30 min at room temperature. Final complexes were concentrated to 10−15 mg/mL and centrifuged at 20000g for 30 min. Sitting drops of 0.1 μL protein solution and 0.1 μL reservoir solution against 70 μL of reservoir solution were set at 22 °C using commercial screens from QIAGEN and Hampton Research. Crystals grew in 1 week. Crystals of Cat K with compound **5** grew in 0.2 M KCl, 0.05 M HEPES (pH 7.5), 35% pentaerythritol propoxylate (5/4 PO/OH) solution. Crystals of Cat K with compound **6** grew in 0.2 M Li_2SO_4, 0.1 M NaOAc (pH 4.5), 30% PEG 8000 solution. Crystals of Cat K with compound **9** grew in 2.4 M $(NH_4)_2SO_4$/0.1 M NaOAc (pH 5.0) medium. All crystals were transferred for 5 s into cryobuffer, reservoir solution supplemented with 20% glycerol, and flash-cooled in liquid nitrogen.

Data Processing. Crystals of Cat K with compounds **5** and **6** were diffracting to 1.6 Å at home source using Bruker Microstar rotation anode, whereas the crystals of Cat K with compound **9** diffracted to 1.0 Å at BESSY BEAMLINE 14.1 in Berlin on PILATUS 6 M detector using 1.0 wavelength.[59] Data sets of the first two crystals were processed with PROTEUM (Proteum Software Pty Ltd.), whereas the third one was processed with XDSAPP.[60]

The diffraction pattern of all crystals revealed the presence of a single crystal lattice. Data sets were scaled in P21 for Cat K with compound **5**, in P212121 for Cat K with compound **6**, and Cat K with compound **9**. Data statistics are presented in Table S4.

Molecular Replacement. Matthews coefficient analysis ambiguously defined one molecule in the asymmetric unit for all three Cat K structures. The structures were solved using *PHASER*[61] with the PDB model 4DMX.[36] Refinement was performed using *REFMACS*[62] and *MAIN*.[63] Model building was done using *Coot*[64] and *MAIN*.[63] Complete chain of Cat K from 1st to 215th in all three crystals is well resolved by electron density maps. Several residues were built with alternative conformations. Additionally, one potassium and one chloride ion was built in the compound **5** complex, five SO_4^{2-} anions in the crystal complex of Cat K with compound **6** and six SO_4^{2-}, two NH_4^+, and one AcO^- anions in the compound **9** complex. Determined coordinates of Cat K with compounds **5**, **6**, and **9** were deposited in PDB as 4X6J, 4X6I, and 4X6H, respectively.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jmedchem.5b00746.

Additional details of combinatorial data set and consensus modeling, tables of crystallographic data, schematic representation of virtual synthetic pathways, reversibility test, NMR spectra (PDF)

CSV file containing molecular formula strings (CSV)

Additional experimental and combinatorial data and consensus results (XLSX)

Accession Codes

PDB ID codes: 4X6J, 4X6I, 4X6H.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: marjana.novic@ki.si.

*E-mail: boris.turk@ijs.si.

*E-mail: dusan.turk@ijs.si.

Notes

The authors declare no competing financial interest.

DOI: 10.1021/acs.jmedchem.5b00746
J. Med. Chem. 2015, 58, 6928−6937

■ ACKNOWLEDGMENTS

This work was supported by the Slovene Research Agency (Grants No. P1-0017, P1-0140 and P1-0048).

■ REFERENCES

(1) Fonovic, M.; Turk, B. Cysteine cathepsins and extracellular matrix degradation. *Biochim. Biophys. Acta, Gen. Subj.* **2014**, *1840*, 2560−2570.
(2) Novinec, M.; Lenarcic, B. Cathepsin K: a unique collagenolytic cysteine peptidase. *Biol. Chem.* **2013**, *394*, 1163−1179.
(3) Deaton, D. M.; Kumar, S. 6. Cathepsin K inhibitors: their potential as anti-osteoporosis agents. *Prog. Med. Chem.* **2004**, *42*, 245−375.
(4) Katunuma, N. Structure-based development of specific inhibitors for individual cathepsins and their medical applications. *Proc. Jpn. Acad., Ser. B* **2011**, *87*, 29−39.
(5) Nomura, T.; Katunuma, N. Involvement of cathepsins in the invasion, metastasis and proliferation of cancer cells. *J. Med. Invest.* **2005**, *52*, 1−9.
(6) Turk, V.; Turk, B.; Turk, D. Lysosomal cysteine proteases: facts and opportunities. *EMBO J.* **2001**, *20*, 4629−4633.
(7) Turk, V.; Stoka, V.; Vasiljeva, O.; Renko, M.; Sun, T.; Turk, B.; Turk, D. Cysteine cathepsins: from structure, function and regulation to new frontiers. *Biochim. Biophys. Acta, Proteins Proteomics* **2012**, *1824*, 68−88.
(8) Aguda, A. H.; Panwar, P.; Du, X.; Nguyen, N. T.; Brayer, G. D.; Brömme, D. Structural basis of collagen fiber degradation by cathepsin K. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 17474−17479.
(9) Hua, Y.; Nair, S. Proteases in cardiometabolic diseases: Pathophysiology, molecular mechanisms and clinical applications. *Biochim. Biophys. Acta, Mol. Basis Dis.* **2015**, *1852*, 195−208.
(10) Nagase, S.; Ohyama, M.; Hashimoto, Y.; Small, M.; Sharpe, J.; Manako, J.; Kuwayama, T.; Deacon, S. Bone turnover markers and pharmacokinetics of a new sustained-release formulation of the cathepsin K inhibitor, ONO-5334, in healthy post-menopausal women. *J. Bone Miner. Metab.* **2015**, *33*, 93−100.
(11) Cheng, X. W.; Shi, G. P.; Kuzuya, M.; Sasaki, T.; Okumura, K.; Murohara, T. Role for cysteine protease cathepsins in heart disease: focus on biology and mechanisms with clinical implication. *Circulation* **2012**, *125*, 1551−1562.
(12) Brömme, D.; Lecaille, F. Cathepsin K inhibitors for osteoporosis and potential off-target effects. *Expert Opin. Invest. Drugs* **2009**, *18*, 585−600.
(13) Vassalle, C.; Iervasi, G. Cathepsin K - A classical bone biomarker in cardiovascular disease: The heart is not alone anymore. *Atherosclerosis* **2013**, *228*, 36−37.
(14) Hua, Y.; Zhang, Y.; Dolence, J.; Shi, G. P.; Ren, J.; Nair, S. Cathepsin K knockout mitigates high-fat diet-induced cardiac hypertrophy and contractile dysfunction. *Diabetes* **2013**, *62*, 498−509.
(15) Frizler, M.; Stirnberg, M.; Sisay, M. T.; Gütschow, M. Development of Nitrile-Based Peptidic Inhibitors of Cysteine Cathepsins. *Curr. Top. Med. Chem.* **2010**, *10*, 294−322.
(16) Wijkmans, J.; Gossen, J. Inhibitors of cathepsin K: a patent review (2004 - 2010). *Expert Opin. Ther. Pat.* **2011**, *21*, 1611−1629.
(17) Chapurlat, R. D. Treatment of postmenopausal osteoporosis with odanacatib. *Expert Opin. Pharmacother.* **2014**, *15*, 559−564.
(18) Engelke, K.; Nagase, S.; Fuerst, T.; Small, M.; Kuwayama, T.; Deacon, S.; Eastell, R.; Genant, H. K. The Effect of the Cathepsin K Inhibitor ONO-5334 on Trabecular and Cortical Bone in Post-menopausal Osteoporosis: The OCEAN Study. *J. Bone Miner. Res.* **2014**, *29*, 629−638.
(19) Duong, L. T.; Wesolowski, G. A.; Leung, P.; Oballa, R.; Pickarski, M. Efficacy of a Cathepsin K Inhibitor in a Preclinical Model for Prevention and Treatment of Breast Cancer Bone Metastasis. *Mol. Cancer Ther.* **2014**, *13*, 2898−2909.
(20) Nakamura, T.; Shiraki, M.; Fukunaga, M.; Tomomitsu, T.; Santora, A. C.; Tsai, R.; Fujimoto, G.; Nakagomi, M.; Tsubouchi, H.; Rosenberg, E.; Uchida, S. Effect of the cathepsin K inhibitor odanacatib administered once weekly on bone mineral density in Japanese patients with osteoporosis-a double-blind, randomized, dose-finding study. *Osteoporosis Int.* **2014**, *25*, 367−376.
(21) Brixen, K.; Chapurlat, R.; Cheung, A. M.; Keaveny, T. M.; Fuerst, T.; Engelke, K.; Recker, R.; Dardzinski, B.; Verbruggen, N.; Ather, S.; Rosenberg, E.; de Papp, A. E. Bone density, turnover, and estimated strength in postmenopausal women treated with odanacatib: a randomized trial. *J. Clin. Endocrinol. Metab.* **2013**, *98*, 571−580.
(22) Stoch, S. A.; Zajic, S.; Stone, J. A.; Miller, D. L.; van Bortel, L.; Lasseter, K. C.; Pramanik, B.; Cilissen, C.; Liu, Q.; Liu, L.; Scott, B. B.; Panebianco, D.; Ding, Y.; Gottesdiener, K.; Wagner, J. A. Odanacatib, a Selective Cathepsin K inhibitor to Treat Osteoporosis: Safety, Tolerability, Pharmacokinetics and Pharmacodynamics - Results from Single Oral Dose Studies in Healthy Volunteers. *Br. J. Clin. Pharmacol.* **2013**, *75*, 1240−1254.
(23) Nagase, S.; Hashimoto, Y.; Small, M.; Ohyama, M.; Kuwayama, T.; Deacon, S. Serum and urine bone resorption markers and pharmacokinetics of the cathepsin K inhibitor ONO-5334 after ascending single doses in post menopausal women. *Br. J. Clin. Pharmacol.* **2012**, *74*, 959−970.
(24) Ng, K. W. Potential role of odanacatib in the treatment of osteoporosis. *Clin. Interventions Aging* **2012**, *7*, 235−247.
(25) Ochi, Y.; Yamada, H.; Mori, H.; Nakanishi, Y.; Nishikawa, S.; Kayasuga, R.; Kawada, N.; Kunishige, A.; Hashimoto, Y.; Tanaka, M.; Sugitani, M.; Kawabata, K. Effects of ONO-5334, a novel orally-active inhibitor of cathepsin K, on bone metabolism. *Bone* **2011**, *49*, 1351−1356.
(26) Rünger, T. M.; Adami, S.; Benhamou, C. L.; Czerwiński, E.; Farrerons, J.; Kendler, D. L.; Mindeholm, L.; Realdi, G.; Roux, C.; Smith, V. Morphea-like skin reactions in patients treated with the cathepsin K inhibitor balicatib. *J. Am. Acad. Dermatol.* **2012**, *66*, e89−e96.
(27) Falgueyret, J. P.; Desmarais, S.; Oballa, R.; Black, W. C.; Cromlish, W.; Khougaz, K.; Lamontagne, S.; Massé, F.; Riendeau, D.; Toulmond, S.; Percival, M. D. Lysosomotropism of basic cathepsin K inhibitors contributes to increased cellular potencies against off-target cathepsins and reduced functional selectivity. *J. Med. Chem.* **2005**, *48*, 7535−7543.
(28) McGrath, M. E.; Klaus, J. L.; Barnes, M. G.; Brömme, D. Crystal structure of human cathepsin K complexed with a potent inhibitor. *Nat. Struct. Biol.* **1997**, *4*, 105−109.
(29) Turk, D.; Guncar, G.; Podobnik, M.; Turk, B. Revised definition of substrate binding sites of papain-like cysteine proteases. *Biol. Chem.* **1998**, *379*, 137−147.
(30) Minovski, N.; Perdih, A.; Solmajer, T. Combinatorially-generated library of 6-fluoroquinolone analogs as potential novel antitubercular agents: a chemometric and molecular modeling assessment. *J. Mol. Model.* **2012**, *18*, 1735−1753.
(31) Minovski, N.; Perdih, A.; Novic, M.; Solmajer, T. Cluster-based molecular docking study for in silico identification of novel 6-fluoroquinolones as potential inhibitors against mycobacterium tuberculosis. *J. Comput. Chem.* **2013**, *34*, 790−801.
(32) Borišek, J.; Drgan, V.; Minovski, N.; Novič, M. Mechanistic interpretation of artificial neural network-based QSAR model for prediction of cathepsin K inhibition potency. *J. Chemom.* **2014**, *28*, 272−281.
(33) Enamine fragment database. URL: http://www.enamine.net/index.php?option=com_content&task=view&id=117 (accessed Feb., 15, 2013).
(34) Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A rule of three for fragment-based lead discovery? *Drug Discovery Today* **2003**, *8*, 876−877.
(35) Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **2002**, *45*, 2615−2623.
(36) Dossetter, A. G.; Beeley, H.; Bowyer, J.; Cook, C. R.; Crawford, J. J.; Finlayson, J. E.; Heron, N. M.; Heyes, C.; Highton, A. J.; Hudson, J. A.; Jestel, A.; Kenny, P. W.; Krapp, S.; Martin, S.; MacFaul, P. A.; McGuire, T. M.; Gutierrez, P. M.; Morley, A. D.; Morris, J. J.; Page, K.

DOI: 10.1021/acs.jmedchem.5b00746
J. Med. Chem. 2015, 58, 6928−6937

M.; Ribeiro, L. R.; Sawney, H.; Steinbacher, S.; Smith, C.; Vickers, M. (1R,2R)-N-(1-cyanocyclopropyl)-2-(6-methoxy-1,3,4,5-tetrahydropyrido[4,3-b]indole −2-carbonyl)cyclohexanecarboxamide (AZD4996): a potent and highly selective cathepsin K inhibitor for the treatment of osteoarthritis. *J. Med. Chem.* **2012**, *55*, 6363−6374.
(37) Palmer, J. T.; Bryant, C.; Wang, D. X.; Davis, D. E.; Setti, E. L.; Rydzewski, R. M.; Venkatraman, S.; Tian, Z. Q.; Burrill, L. C.; Mendonca, R. V.; Springman, E.; McCarter, J.; Chung, T.; Cheung, H.; Janc, J. W.; McGrath, M.; Somoza, J. R.; Enriquez, P.; Yu, Z. W.; Strickley, R. M.; Liu, L.; Venuti, M. C.; Percival, M. D.; Falgueyret, J. P.; Prasit, P.; Oballa, R.; Riendeau, D.; Young, R. N.; Wesolowski, G.; Rodan, S. B.; Johnson, C.; Kimmel, D. B.; Rodan, G. Design and synthesis of tri-ring P3 benzamide-containing aminonitriles as potent, selective, orally effective inhibitors of cathepsin K. *J. Med. Chem.* **2005**, *48*, 7520−7534.
(38) Frizler, M.; Lohr, F.; Furtmann, N.; Kläs, J.; Gütschow, M. Structural Optimization of Azadipeptide Nitriles Strongly Increases Association Rates and Allows the Development of Selective Cathepsin Inhibitors. *J. Med. Chem.* **2011**, *54*, 396−400.
(39) Frizler, M.; Lohr, F.; Lülsdorff, M.; Gütschow, M. Facing the gem-Dialkyl Effect in Enzyme Inhibitor Design: Preparation of Homocycloleucine-Based Azadipeptide Nitriles. *Chem. - Eur. J.* **2011**, *17*, 11419−11423.
(40) Falgueyret, J. P.; Desmarais, S.; Oballa, R.; Black, W. C.; Cromlish, W.; Khougaz, K.; Lamontagne, S.; Massé, F.; Riendeau, D.; Toulmond, S.; Percival, M. D. Lysosomotropism of Basic Cathepsin K Inhibitors Contributes to Increased Cellular Potencies against Off-Target Cathepsins and Reduced Functional Selectivity. *J. Med. Chem.* **2005**, *48*, 7535−43.
(41) Schröder, J.; Klinger, A.; Oellien, F.; Marhöfer, R. J.; Duszenko, M.; Selzer, P. M. Docking-based virtual screening of covalently binding ligands: an orthogonal lead discovery approach. *J. Med. Chem.* **2013**, *56*, 1478−1490.
(42) Barrett, A. J.; Kembhavi, A. A.; Brown, M. A.; Kirschke, H.; Knight, C. G.; Tamai, M.; Hanada, K. L-*trans*-Epoxysuccinyl-leucylamido(4-guanidino)butane (E-64) and its analogs as inhibitors of cysteine proteinases including cathepsins B, H and L. *Biochem. J.* **1982**, *201*, 189−198.
(43) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(44) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins: Struct., Funct., Genet.* **2003**, *52*, 609−623.
(45) Nissink, J. W.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. A new test set for validating predictions of protein-ligand interaction. *Proteins: Struct., Funct., Genet.* **2002**, *49*, 457−471.
(46) Kirchmair, J.; Markt, P.; Distinto, S.; Wolber, G.; Langer, T. Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection - What can we learn from earlier mistakes? *J. Comput.-Aided Mol. Des.* **2008**, *22*, 213−228.
(47) Carter, M. C.; Alber, D. G.; Baxter, R. C.; Bithell, S. K.; Budworth, J.; Chubb, A.; Cockerill, G. S.; Dowdell, V. C. L.; Henderson, E. A.; Keegan, S. J.; Kelsey, R. D.; Lockyer, M. J.; Stables, J. N.; Wilson, L. J.; Powell, K. L. 1,4-Benzodiazepines as inhibitors of respiratory syncytial virus. *J. Med. Chem.* **2006**, *49*, 2311−2319.
(48) Bernhart, C.; Bouaboula, M.; Casellas, P.; Duclos, O.; Herbert, J. M.; Jegham, S.; Mccort, G. Antineoplastic derivatives, preparation thereof, and therapeutic use thereof. Patent WO2010004198 A2, 2010.
(49) Olson, D. R.; Wheeler, W. J.; Wells, J. N. Rigid Analogs of Indomethacin. *J. Med. Chem.* **1974**, *17*, 167−171.
(50) Mihelic, M.; Dobersek, A.; Guncar, G.; Turk, D. Inhibitory fragment from the p41 form of invariant chain can regulate activity of cysteine cathepsins in antigen presentation. *J. Biol. Chem.* **2008**, *283*, 14453−60.
(51) Brömme, D.; Nallaseth, F. S.; Turk, B. Production and activation of recombinant papain-like cysteine proteases. *Methods* **2004**, *32*, 199−206.
(52) Rozman-Pungercar, J.; Kopitar-Jerala, N.; Bogyo, M.; Turk, D.; Vasiljeva, O.; Stefe, I.; Vandenabeele, P.; Brömme, D.; Puizdar, V.; Fonović, M.; Trstenjak-Prebanda, M.; Dolenc, I.; Turk, V.; Turk, B. Inhibition of papain-like cysteine proteases and legumain by caspase-specific inhibitors: when reaction mechanism is more important than specificity. *Cell Death Differ.* **2003**, *10*, 881−8.
(53) Greco, W. R.; Hakala, M. T. Evaluation of methods for estimating the dissociation constant of tight binding enzyme inhibitors. *J. Biol. Chem.* **1979**, *254*, 12104−9.
(54) Cherney, M. M.; Lecaille, F.; Kienitz, M.; Nallaseth, F. S.; Li, Z.; James, M. N.; Brömme, D. Structure-activity analysis of cathepsin K/chondroitin 4-sulfate interactions. *J. Biol. Chem.* **2011**, *286*, 8988−98.
(55) Kirschke, H.; Barrett, A. J.; Rawlings, N. D. Proteinases 1: lysosomal cysteine proteinases. *Protein Profile* **1995**, *2*, 1581−643.
(56) Turk, B.; Colić, A.; Stoka, V.; Turk, V. Kinetics of inhibition of bovine cathepsin S by bovine stefin B. *FEBS Lett.* **1994**, *339*, 155−159.
(57) Morrison, J. F. The slow-binding and slow, tight-binding inhibition of enzyme-catalysed reactions. *Trends Biochem. Sci.* **1982**, *7*, 102−105.
(58) Invitrogen. Pichia Expression Kit. http://tools.thermofisher.com/content/sfs/manuals/pich_man.pdf.
(59) Mueller, U.; Darowski, N.; Fuchs, M. R.; Förster, R.; Hellmig, M.; Paithankar, K. S.; Pühringer, S.; Steffien, M.; Zocher, G.; Weiss, M. S. Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin. *J. Synchrotron Radiat.* **2012**, *19*, 442−449.
(60) Krug, M.; Weiss, M. S.; Heinemann, U.; Mueller, U. XDSAPP: a graphical user interface for the convenient processing of diffraction data using XDS. *J. Appl. Crystallogr.* **2012**, *45*, 568−572.
(61) Mccoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658−674.
(62) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 355−367.
(63) Turk, D. MAIN software for density averaging, model building, structure refinement and validation. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013**, *69*, 1342−1357.
(64) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486−501.

DOI: 10.1021/acs.jmedchem.5b00746
J. Med. Chem. 2015, 58, 6928−6937