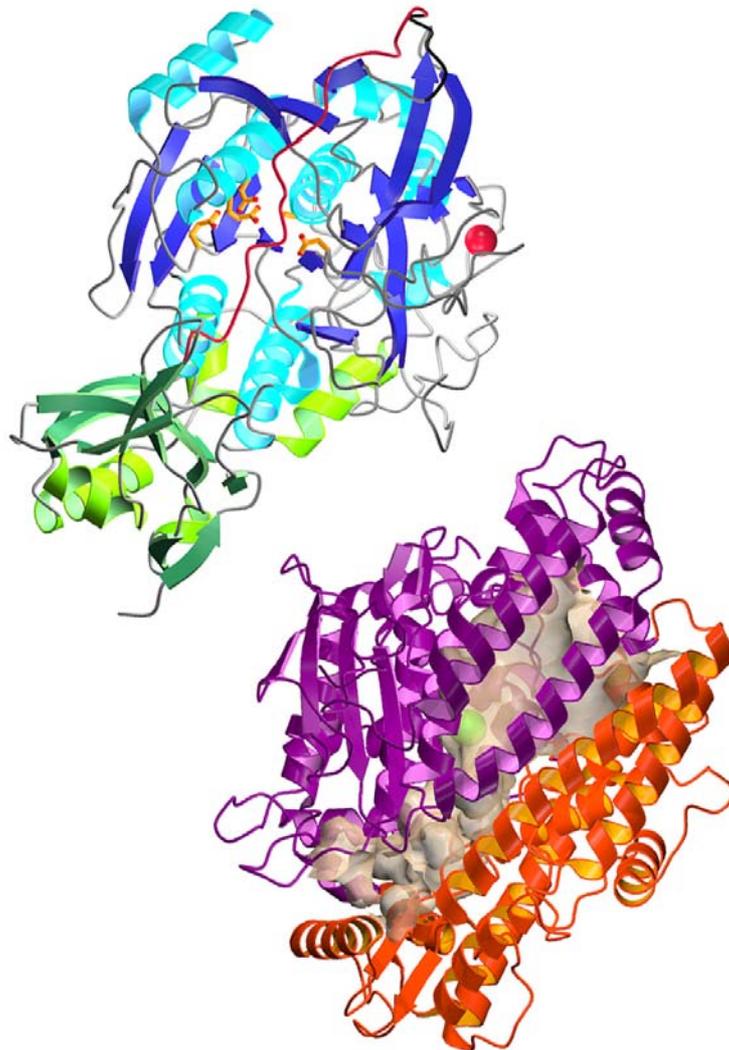


**Structural Studies of the
Serine-Carboxyl Proteinase Kumamolisin and
the Metallopeptidase Peptidyl-Dipeptidase Dcp**



Mireia Comellas-Bigler

June, 2004
Max-Planck-Institut für Biochemie
Abteilung Strukturforschung
D-82152, Martinsried, Germany

Max-Planck-Institut für Biochemie
Abteilung Strukturforschung

**Structural Studies of the
Serine-Carboxyl Proteinase Kumamolisin and
the Metallopeptidase Peptidyl-Dipeptidase Dcp**

Mireia Comellas Bigler

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität
München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Steffen J. Glaser
Prüfer der Dissertation: 1. apl. Prof. Dr. Dr. h.c. Robert Huber
2. Univ.-Prof. Dr. Wolfgang Hiller

Die Dissertation wurde am 01.07.2004 bei der Technischen Universität München eingereicht
und durch die Fakultät für Chemie am 09.09.2004 angenommen.

Pels meus pares
u für mine iai

Table of Contents

1 Summary	1
1.1 Kumamolisin a Serine-Carboxyl Proteinase	1
1.2 Peptidyl-Dipeptidase Dcp	3
2 Introduction	5
2.1 Proteolytic enzymes	5
2.2 Classification	5
2.3 The Peptidase Zymogen Proregions	6
2.3.1 Proregion Location in the Proenzyme	7
2.3.2 Proregions Exhibit Diverse Functions other than Inhibition/Regulation	8
2.4 Serine Proteinases	9
2.4.1 Zymogens of Serine Proteinases	10
2.4.2 The SB Clan of Subtilases	11
2.4.3 Catalytic Mechanism of Subtilases: Differences and Similarities between S6 and S53 Families	12
2.4.4 The Substrate Recognition Sites	13
2.4.4.1 The Unspecific Main Chain Substrate Binding	14
2.4.4.2 The Specificity Pockets	15
2.4.5 Kumamolisin: a Member of the S53 Serine Proteinases	15
2.5 Metallopeptidases	18
2.5.1 The MA(E) Clan of Gluzincins	20
2.5.2 Peptidyl-Dipeptidase Dcp and the Related Enzymes: Angiotensin-I- Converting Enzyme (ACE) and Neurolysin	21
2.5.2.1 Angiotensin-I-Converting Enzyme	21
2.5.2.2 Peptidyl-Dipeptidase Dcp	22
2.5.2.3 Neurolysin	23
2.6 Protein Crystallography	24

3 Experimental Procedures	26
3.1 Materials	26
3.1.1 Chemicals	26
3.1.2 Bacterias	26
3.1.3 Plasmid	27
3.1.4 Restriction Enzymes	27
3.1.5 Media	27
3.1.6 Devices.....	28
3.2 Molecular Biology Methods	28
3.2.1 Plasmid Preparation	28
3.2.2 Restriction Analysis	28
3.2.3 Agarose Gel Electrophoresis	29
3.2.4 <i>E. coli</i> Transformation by Electroporation	29
3.2.5 Cell Cultures	30
3.2.5.1 Plate Cultures	30
3.2.5.2 Liquid Cultures	30
3.2.5.3 Glycerol Cultures	30
3.3 Protein Biochemistry	30
3.3.1 Protein Concentration Determination	30
3.3.2 Apparent Molecular Wight: SDS Polyacrylamide Gel Electrophoresis	31
3.3.3 Protein Transfer and Western Blots	32
3.3.4 N-Terminal Sequencing by Edman Degradation	33
3.3.5 Mass Spectroscopy	34
3.3.6 Protein Concentration	35
3.4 Protein Production	35
3.4.1 Protein Over-Expression	35
3.4.1.1 Peptidyl-Dipeptidase Dcp Over-Expression	35
3.4.1.2 SeMet-Peptidyl-Dipeptidase Dcp Over-Expression	35
3.4.2 Cell Disruption	36
3.4.3 Protein Preparation	36
3.4.3.1 Purification of Native and SeMet-Peptidyl-Dipeptidase Dcp	36
3.4.3.2 Peptidyl-Dipeptidase Dcp/Inhibitor Complex Purification	37
3.5 Crystallography Methods	37
3.5.1 Crystallization	37
3.5.2 Inhibitor Soaking and Co-Crystallization	38

3.5.2.1 Kumamolisin-Inhibitor Complexes Co-Crystallization	38
3.5.2.2 Peptidyl-Dipeptidase Dcp/Inhibitor Complex Co-Crystallization	40
3.5.3 Crystal Seeding	40
3.5.1 Peptidyl-Dipeptidase Dcp Microseeding	41
3.5.4 Crystal Mounting	41
3.5.5 Data Collection	42
3.5.6 Phasing	42
3.5.6.1 The Patterson Function	43
3.5.6.2 Kumamolisin was solved by Molecular Replacement	43
3.5.6.3 Peptidyl-Dipeptidase Dcp was solved by MAD	44
3.5.7 Modeling and Refinement	45
3.5.8 Structural Analysis and Graphical Representation	45
4 Results	47
4.1 Kumamolisin a Serine-Carboxyl Proteinase from <i>Bacillus</i>	
novosp. MN-32	47
4.1.1 Protein Crystallization	47
4.1.2 Data Collection and Refinement	48
4.1.3 Description of the Structure	53
4.1.3.1 Overall Structure of Kumamolisin.....	53
4.1.3.2 Structural Comparison with Sedolisin and Subtilisin	56
4.1.3.3 Active Site and Substrate Binding Site	60
4.1.3.4. Site-Directed Mutagenesis Studies	63
4.2 Pro-Kumamolisin and Mutants E32A and W129A	64
4.2.1 Protein Crystallization	64
4.2.2 Data Collection and Refinement	65
4.2.3 Description of the Structures	69
4.2.3.1 Structure of the E32A and W129 Mutants of Kumamolisin	69
4.2.3.2 Overall Structure of Pro-Kumamolisin	72
4.2.3.3 Prodomain of Pro-Kumamolisin	74
4.2.3.4 Activation Peptide or Linker	76
4.3 Peptidyl-Dipeptidase Dcp from <i>E. coli</i>	79
4.3.1 Protein Over-Expression and Preparation	79
4.3.2 Protein Crystallization	80
4.3.3 Native and MAD Data Collection	81
4.3.4 Phasing and Refinement.....	83
4.3.5 Description of the Structure	87
4.3.5.1 Overall Structure of Peptidyl-Dipeptidase Dcp	87
4.3.5.2 Structural Comparison with other Metallopeptidases	89
4.3.5.3 Active Site and Substrate Binding	94

5 Discussion	99
5.1 Kumamolisin a Serine-Carboxyl Proteinase from <i>Bacillus</i>	
novosp. MN-32	99
5.2 Kumamolisin Mutants E32A and W129A	107
5.3 Pro-Kumamolisin	108
5.4 Peptidyl-Dipeptidase Dcp from <i>E. coli</i>	111
6 Abbreviations	119
7 References	121
8 Acknowledgments	135

1 Summary

1.1 Kumamolisin a Serine-Carboxyl Proteinase

In the early 70s, three proteinases had been isolated from *Scytalidium lignicolum* that exhibited optimal catalytic activity at acidic pH, but lacked primary structure similarity to any other known carboxyl proteinase. Several studies reported the insensitivity of these proteinases towards most common aspartyl proteinase inhibitors such as pepstatin and the presence of at least two catalytically important carboxyl groups in the enzymes of this subfamily. These experimental evidences lead to the proposal of a new class of carboxyl proteinases. More recently, several other enzymes have been classified as members of this family based on sequence similarities, among them the prokaryotic enzymes from *Pseudomonas* sp. (sedolisin), from *Bacillus coagulans* (J-4), and the thermostable enzyme from *Bacillus* novosp. MN-32 (kumamolisin), which share 30 to 35% sequence identity. This proteinase family also includes eukaryotic homologues, such as the human lysosomal tripeptidyl-peptidase I (CLN2), which has been recently assigned to the family. Mutations in the encoding *CLN2* gene are directly associated with a fatal neurodegenerative disease called classical late-infantile neuronal ceroid lipofuscinosis.

The crystal structures of native kumamolisin and in complex with two tripeptidyl aldehyde inhibitors reveal a new Ser-Glu-Asp catalytic triad connected to a Glu-Trp diad in a subtilisin-like scaffold. The catalytic domain of Kumamolisin consists essentially of a central eight-stranded parallel β -sheet, flanked by eight helices and a few short strand pairs arranged on both sides of the major sheet. Kumamolisin exhibits a Ser-Glu-Asp triad, in which a glutamate substitutes the histidine of the classical Ser-His-Asp catalytic triad of subtilases. The catalytic machinery of kumamolisin, characterized by particularly short inter-connecting hydrogen bonds, enables the reactive serine to attack substrate peptide bonds at quite acidic pH. These findings agree with the previously solved structure of the related enzyme sedolisin from *Pseudomonas* sp. and with mutagenesis studies on the human homologue, CLN2, which indicated that one serine and two aspartic acid residues were essential for catalytic activity. In accordance with these evidences, the enzyme family was renamed serine-carboxyl proteinase (SCP) family or, more recently,

sedolisin family. Hence, the sedolisin family, S53, was included in the SB clan of subtilases, together with the subtilisin family, S8.

An additional Glu/Trp pair, unique to kumamolisin, extends the active site hydrogen bond network in the thermostable enzyme. To evaluate the importance of the glutamic acid and the tryptophan residues in kumamolisin catalysis, mutants of each residue were prepared and their proteolytic activities were characterized. The determination of the three-dimensional structures for the Glu32Ala and Trp129Ala mutants (at atomic or near-atomic resolution) shows that the decrease in proteolytic activity does not result from the replacement of these two side chains alone, but is also a consequence of local structural changes. The absence of the glutamate or the tryptophan prevents the formation of the hydrogen bond network in the active site favoring double conformations in the other catalytic residues.

Most proteins identified as serine-carboxyl proteinases are synthesized as preproforms, with propeptides ranging 170 to 215 residues that have to be cleaved off for activation of the enzymes. To study the structural and functional features of the SCPs prodomain, the stable inactive full-length mutant of pro-kumamolisin enzyme, Ser278Ala, was isolated and crystallized. The prodomain exhibits a half- β -sandwich core docking to the catalytic domain similarly as the equivalent subtilisin prodomains in their catalytic-domain complexes. Nevertheless, the prodomain of pro-kumamolisin is considerably larger than its structurally homologous prodomains, giving rise to two additional sub-domains: a three-stranded multiple-turn appendix and a sub-domain comprising two helices, which are connected by an exposed large loop. This pro-kumamolisin structure displays, for the first time, the uncleaved linker segment running across the properly arranged active-site. The peptide linker is bound to the enzyme in a substrate-like manner and connects the prodomain with the catalytic domain. As presumed for pro-subtilisin and pro-furin, the structure of kumamolisin strongly points to an initial intra-molecular activation mechanism in subtilases, in which the first step corresponds to the cleavage of the peptide linker.

1.2 Peptidyl-Dipeptidase Dcp

Escherichia coli possesses a number of peptidolytic activities to meet the requirement for intracellular protein breakdown. Peptidases are involved not only in the utilization of peptides supplied in the medium, but also in the turnover of cytoplasmic proteins during cell growth and starvation as well as degradation of signal peptides resulting from protein export. About 20% of the proteins of gram-negative bacteria are exported to the periplasmic or secreted into the extracellular space.

Peptidyl-dipeptidase Dcp from *E. coli* is a 70 kDa exopeptidase, which belongs to the M3(A) family of metalloproteases also known as the thimet oligopeptidase family. All enzymes of this family possess one catalytic zinc in the active site and the conserved HEXXH metal ion binding motive. However, Dcp differs markedly in its specificity and susceptibility towards inhibitors from its related enzymes such as neurolysin and oligopeptidase A. To our knowledge, Dcp is the only C-terminal exopeptidases known to be present in *E. coli*. Dcp requires a free C-terminus and cannot hydrolyze Xaa-Pro peptide bonds or Gly-Gly peptide bonds. Interestingly, Dcp, in its catalytic activity and the susceptibility towards inhibitors, closely resembles mammalian angiotensin-I-converting enzyme (ACE), which is involved in blood pressure regulation.

The crystal structure of Dcp in complex with a peptidic inhibitor (N-1450) displays an overall prolate ellipsoid shape, with a large internal cavity formed by two domains. The enzyme groove encompassing the entire protein molecule buries the inhibitor and the catalytic zinc-binding site, which is conserved among the family members. The three-dimensional structure of Dcp in complex with the inhibitor N-1450 allowed us a detailed description of the active site cleft and an explanation for the peptidyl-dipeptidase activity of the enzyme. In addition, comparison with the native structure of the related neurolysin reveals a highly conserved scaffold of the two domains separately, but with a significant shift of one domain position with respect to the other, between both enzymes. The relative movement of the two domains in the structures of Dcp and neurolysin points to a ligand-dependant hinge-bending movement as the catalytic mechanism for members of family M3(A).

The overall fold of Dcp also resembles that of human ACE, despite the lack of sequence similarity between both enzymes. This indicates a case of divergent evolution

Summary

from a common ancestor between the families of peptidyl-dipeptidase Dcp (M3(A)) and angiotensin-converting enzyme (M2). The strong structural similarity between Dcp and ACE, together with a hinge-bending motion recently observed between native and inhibitor-bound structures of an ACE-related enzyme (ACE2), suggest a similar ligand-dependant movement for catalytic activity of angiotensin-I-converting enzyme.

2 Introduction

2.1 Proteolytic enzymes

Peptidases or proteases designate the subset of hydrolases that carry out proteolytic reactions on peptide bonds, elementary to numerous biological processes. Proteolytic enzymes are involved in a great variety of physiological settings ranging from the highly specific and precisely regulated systems of blood clotting, fibrinolysis, cellular and humoral immunity, fertilization, embryonic development, cell signaling/migration, wound healing or cell death through apoptosis, to the rather unspecific cleavages of digestive and degradative processes.

2.2 Classification

Within the proteins, peptidases in particular are assigned to families and clans, as proposed by Rawlings and Barrett (1993 and 1999a), based on their primary through tertiary structures, their catalytic mechanisms, and the nature and distribution of the residues in their active sites. Six major families of peptidases have been identified so far (MEROPS Peptidase Database, <http://merops.sanger.ac.uk>): aspartic peptidases, cysteine peptidases, glutamic peptidases, metallopeptidases, serine peptidases and threonine peptidases (Figure 2.1). The cysteine, serine and threonine peptidases differ from the aspartic, glutamic and metallopeptidases in that the catalytic nucleophile in the active site is an amino acid atom, whereas in the latter families the nucleophilic attack is performed by an activated water molecule. Thus, acyl intermediates covalently bound to the enzyme are only formed during the hydrolytic reactions of the Ser/Thr/Cys peptidases.

Peptidases can also be subdivided according to their cleavage sites: the endopeptidases, also called proteinases, which cleave peptide bonds at positions within the main body of the protein, and the exopeptidases, which remove amino acid residues from either N or C-terminus, and therefore are named aminopeptidases or carboxypeptidases, respectively (Figure 2.1).

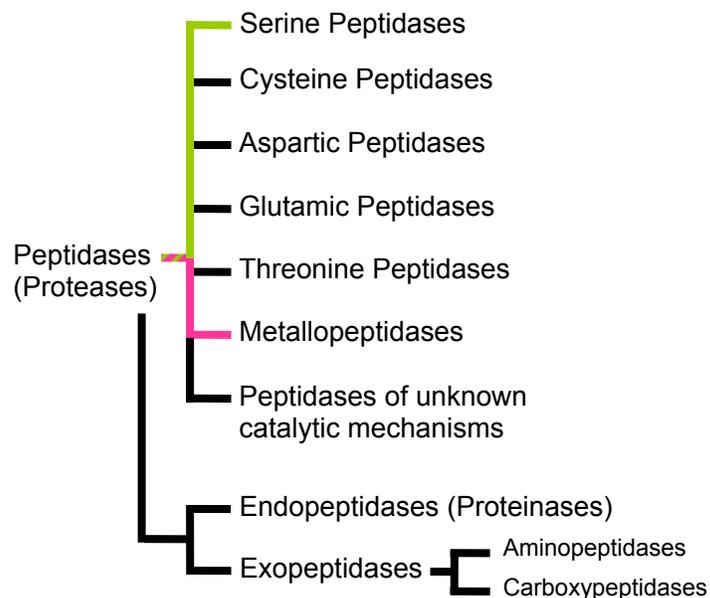


Figure 2.1: Scheme of the current peptidases classification according to the MEROPS data base. Green and pink lines indicate the two families discussed in more detail in the present work.

2.3 The Peptidase Zymogen Proregions

The synthesis of peptidases as larger precursors is one of several ways in which hydrolytic activity can be regulated, to prevent unwanted protein degradation and collateral damage from its actions in undesired locations or on unintended substrates. A vast majority of cell characterized peptidases, a part from numerous proteins, are synthesized as inactive zymogens that must be activated prior to being able to express their full biological activities. Noteworthy exceptions are signal peptidases (Dalbey and Vonheijne, 1992; Dalbey et al., 1997), or viral peptidases (Babe and Craik, 1997), which are translated within a large polyprotein precursor, part of which could play a role of a classic inhibitory proregion.

The most frequent mechanism of activation, as it leads to irreversible activation, is post-translational cleavage of peptide bond(s) within the proenzyme molecule. These cleavages can be intramolecular or intermolecular, being catalyzed either by peptidases acting on themselves or by activating partners (Lazure, 2002), respectively. Stepwise proenzyme activation via limited proteolysis plays a critical role in numerous biological

processes such as blood coagulation (reviewed by Davie et al., 1991) and the activation of the immune complement system (Mullereberhard, 1988). In both concerted activation systems, the term cascade has been employed to describe the complex set of reactions and regulation mechanisms.

More than any other experimental technique X-ray crystallography has contributed to our current understanding of zymogen activation. Most families of proteinases have currently at least one representative proenzyme structure in the protein database (PDB; <http://www.rcsb.org/pdb>), e.g. pepsinogen of the aspartyl peptidases (James and Sielecki, 1986), procathepsin B of the cysteine peptidases (Cygler et al. 1996; Turk et al., 1996), prostromelysin of the metallopeptidases family (Becker et al., 1995) and trypsinogen (Bode et al., 1976) and prosubtilisin (Gallagher et al., 1995), which are trypsin-like and subtilisin-like serine peptidases, respectively. Nevertheless, among the several crystal structures of subtilisin-like proteinases so far reported, none of them show the uncleaved zymogen, but the complex between the propeptide and its corresponding catalytic enzyme. Interestingly, although the primary structure of proregions even within endopeptidase families appears poorly conserved, the crystallographic studies showed that the overall folds are preserved (Lazure, 2002).

2.3.1 Proregion Location in the Proenzyme

It is conceivable that synthesis of the propeptide, which can range in length from few residues to multidomain structures, precedes that of the catalytic unit avoiding the risk of developing peptidase activity before complete synthesis of the polypeptide. Indeed, most propeptides are located at the N-terminus of the zymogen often immediately following the signal peptide in the case of secreted peptidases. However, activation segments have also been found C-terminally of the catalytic domain, where they play an important role in protein secretion for example in aqualysin from *Thermus aquaticus* (Lee et al., 1994). Finally, activation segments can be found as integral part of the catalytic domain, as in the case of human protective protein, where removal of an internal peptide is necessary to confer activity through postulated conformational changes (Rudenko et al., 1995).

2.3.2 Proregions Exhibit Diverse Functions other than Inhibition/Regulation

Unification of a general hypothesis concerning the nature and role of proregions becomes rather difficult as they exhibit diverse structures and properties among and within the different families. Besides the above mentioned role of C-terminally located proregions in secretion into the yeast vacuolar space (Lee et al., 1994), propeptides are also essential for the assembly of the various subunits of the eukaryotic 20S proteasome into the mature particle (Chen and Hochstrasser, 1996; Voges et al., 1999). However, the presence of prosegments is not required for oligomerization and full activity of the 20S proteasome in *Thermoplasma acidophilum* (Zwickl et al., 1994). Proregions of certain caspases (cysteine-dependant aspartate specific proteases) play an important role in their activation and regulation, as they contain domains that direct the zymogens to specific activating platforms via homotypic protein-protein interactions (Boldin et al., 1996; Li et al., 1997).

The role of propeptides as intramolecular chaperones, assisting peptidases to gain their final active three-dimensional structure, has been widely documented and experimentally demonstrated with studies of refolding *in vitro* of denatured subtilisin upon addition of its prosegment as a separate entity (Zhu et al., 1989; Eder et al., 1993). Intramolecular chaperones can act directly to catalyze folding of their cognate peptidases from a molten globule state to a completely folded state, either by lowering the energy barrier between the two states, and/or by stabilizing the rate-limiting folding intermediate. The chaperonine activity of the proregions implicates a close complementarity between the prodomain and the catalytic domain, which involve numerous non-covalent interactions located at their interface (Shinde et al., 1997). It is therefore not surprising that the prosegments of several proteinases remain tightly associated with the enzyme after cleavage, forming 1:1 inhibitor-enzyme complexes as is the case of the subtilisin-prosubtilisin complex (Gallagher et al., 1995).

2.4 Serine Proteinases

Almost one-third of all peptidases are classified as serine peptidases, based on the presence of a nucleophilic serine residue in their active site. This mechanistic class was originally distinguished by the presence of the Ser-His-Asp catalytic triad, which can be found with strictly conserved geometry in at least four radically different protein scaffolds. Thus, this catalytic machinery has evolved on at least four separate occasions, giving rise to the PA(S), SB, SC and SK clans of serine peptidases (Figure 2.2) typified by chymotrypsin-like (Lesk and Fordham, 1996), subtilisin-like (Siezen and Leunissen, 1997), α/β -hydrolase (Ollis et al., 1992) and Clp protease fold (Wang et al., 1997). More recently, serine peptidases with novel catalytic triads and dyads have been identified, including Ser-His-Glu, Ser-Lys/His, His-Ser-His, and N-terminal Ser (for a review see Hedstrom, 2002). These serine peptidases show structurally unrelated three-dimensional structures, increasing to twelve the last clan counter of this proteinase type (MEROPS, release 6.60, <http://merops.sanger.ac.uk>) (Figure 2.2).

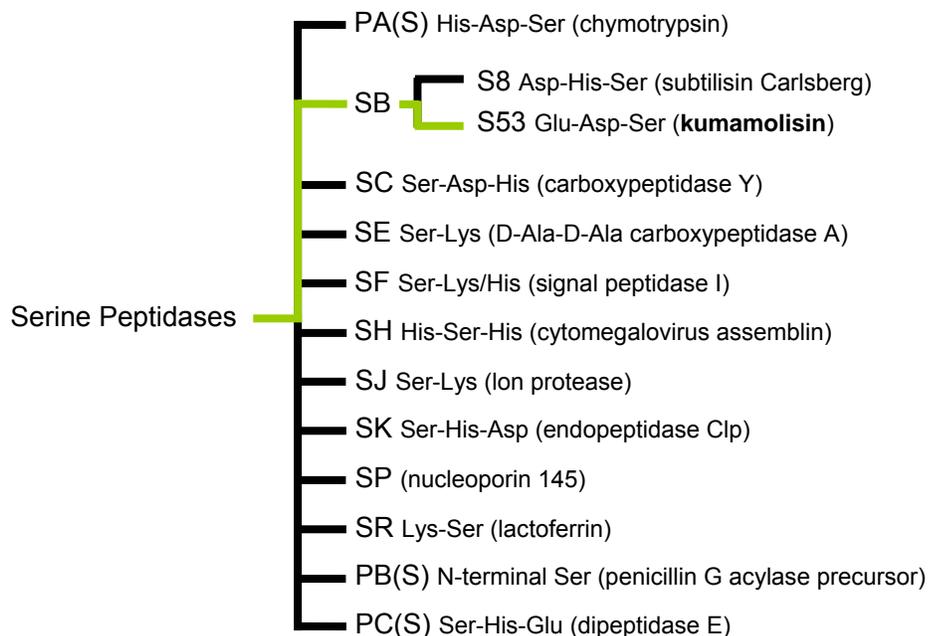


Figure 2.2: Schematic classification of serine peptidases. The catalytic residues are indicated beside the family name. A representative example of each family is given in brackets.

Serine proteinases must overcome three main obstacles to hydrolyze a peptide bond: (i) amide bonds are very stable due to electron delocalization between the amide nitrogen and the carbonyl oxygen. The peptide bond is usually activated via the interaction of the carbonyl oxygen with a general acid; (ii) water is a poor nucleophile, which must be activated by the peptidases, normally via a general base; and (iii) amines are chemically poor leaving groups. Prior to expulsion, proteinases always protonate the amine group. These tasks are carried out by serine proteinases with a high catalytic efficiency, yielding $\sim 10^{10}$ -fold greater rates of peptide hydrolysis than the uncatalyzed reactions. Besides the hydrolysis of peptide bonds, serine peptidases are able to hydrolyze other acyl compounds such as amides, anilides, esters and thioesters.

2.4.1 Zymogens of Serine Proteinases

As reflection of the large variety of serine peptidase folds, different activation mechanisms more or less related are known so far for these enzymes. The first one, which was one of the earliest discovered, involves a significant conformational change from the zymogen into the active enzyme upon cleavage of a single peptide bond. The structural rearrangements that follow insertion of the new N-terminal segment into the main body of the enzyme, result in the proper positioning of the catalytic machinery including the active site residues and the oxyanion hole. This mode of activation is characteristic of chymotrypsin-like enzymes and thus, the inactive state of their zymogens can be explained by the fact that catalytic machinery and cleft are not properly formed (Bode et al., 1976; Khan and James, 1998). A variant of this mechanism, highlighted by the plasminogen structure (Peisach et al., 1999), is related to the immature substrate binding cleft model observed in chymotrypsinogen. However, in plasminogen, the misplacement of the catalytic residues and the oxyanion hole is further reinforced by the presence of a main chain residue occupying the S1 subsite and hence totally preventing of possible substrate binding (Peisach et al., 1999).

A different activation mechanism has been described for the SB clan of subtilases. In this case, the residues of the C-terminal proregion occupy the substrate binding cleft in a product-like manner, pointing to an autoactivation model. Thus, as a general rule, the nature of the residues occupying the C-terminus of the prosegment corresponds closely to

the specificity of the cognate enzyme. The tight enzyme-substrate interaction between catalytic domains and prodomains have been revealed in several structures of complexes of subtilisin-like proteinases with their proregions, e.g. α -lytic protease (Sauter et al., 1998), prohormone convertase furin (Anderson et al., 2002) and subtilisin BPN' (Gallagher et al., 1995). In addition, stable complexes between the prosegments and their corresponding catalytic domains of several subtilases have been isolated. These facts confirm two further properties of these proregions, namely that (i) they can behave as transient potent inhibitors and that (ii) at least a secondary cleavage at an accessible site is necessary in order to degrade the proregion and prevent further inhibition (Lazure, 2002) to fully activate the enzyme (Kojima et al., 1997).

2.4.2 The SB Clan of Subtilases

The SB clan of serine proteinases, which is characterized by an α/β catalytic domain scaffold, consist of family S6 of subtilisin-like peptidases and family S53 of sedolisins (Wlodawer et al., 2003). Almost all subtilases possess at least one of the five calcium ion binding sites known for this superfamily. These calcium ion sites, which are essential for catalysis and stability of the three-dimensional structures, vary in their locations and in their binding strength. Among the subtilisin family members, Ca1 (strong) and Ca3 (weak) are the most frequently observed ion binding sites, whereas the medium-strength Ca2 site is less common (Siezen et al., 1991) and the weak Ca4 site is only found in proteinase K (Betzel et al., 1988a, 1988b). Sedolisins show a single Ca^{2+} site, which is topologically distinct from sites Ca1-Ca4. Over the last years, more than 100 new subtilases have been discovered with extremely widespread occurrence, including bacteria, archaea, fungi, yeasts, viruses and higher eukaryotes such as slime molds, plants, insects, amphibian, fish and mammals (Siezen and Leunissen, 1997).

2.4.3 Catalytic Mechanism of the Subtilases: Differences and Similarities between S6 and S53 Families.

The standard catalytic mechanism employed by serine proteinases belonging to the SB clan relies on a catalytic triad and has been divided in two steps, namely: the acylation phase and the deacylation phase. The subtilisin catalytic triad consists of a histidine general base, which in the acylation step abstracts the proton from the neighboring serine hydroxyl side chain, allowing the latter residue to act as a nucleophile and attack the carbonyl group of the substrate amide bond (Figure 2.3.A). A covalent bond forms between the serine O δ and the substrate to yield a complex known as the tetrahedral intermediate, because of its tetrahedral geometry in contrast to the

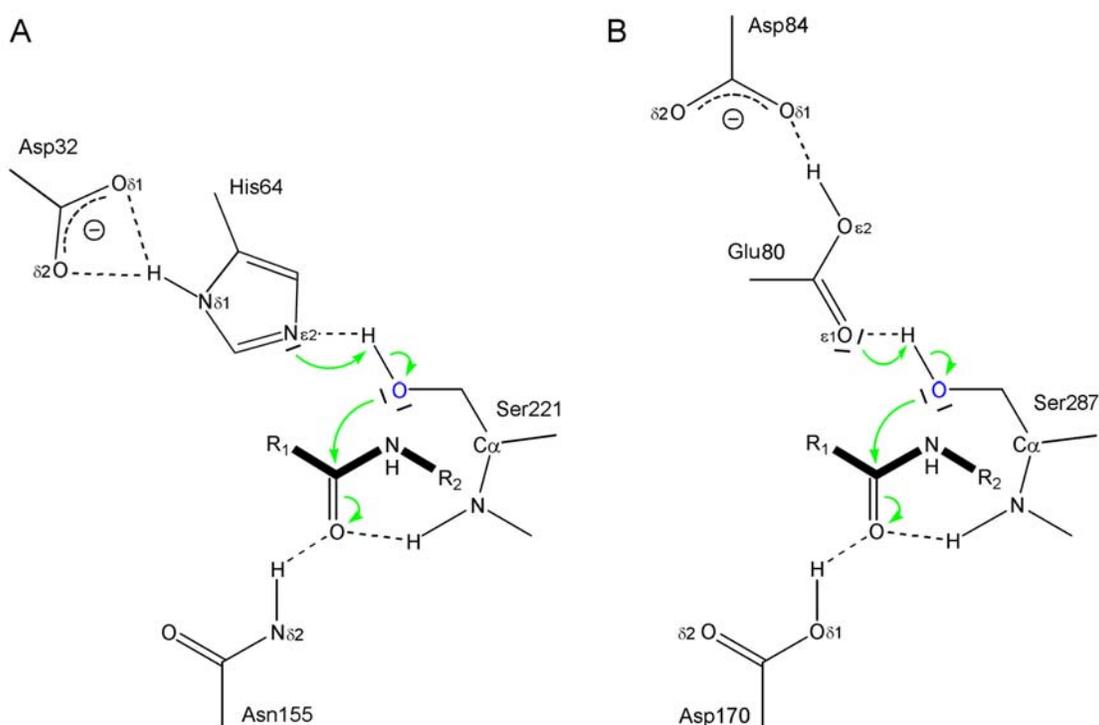


Figure 2.3: Schematic representation of the nucleophilic attack in the subtilase super-family: **A)** Subtilisin Carlsberg (Bode et al., 1987); and **B)** Sedolisin (Wlodawer et al., 2001). The peptidase residues are drawn as thin black lines while the substrate is shown in bold lines. The nucleophilic oxygen of the catalytic serine is highlighted in blue and the first electron movements are indicated with green arrows.

planar triangular geometry of the amide group. In the sedolisin family the catalytic machinery is optimized for hydrolytic activity at acidic pH (Figure 2.3.B). While the nucleophilic attack is also performed by the side chain oxygen of a serine residue, the proton acceptor is a glutamic acid residue instead of a histidine, which at acidic pH would be completely protonated and thus unable to act as a general base. The third component of the triad is an aspartic acid in both families. This aspartate orients the histidine residue and neutralizes the charged histidine intermediate in subtilisins, whereas in the sedolisin family it probably plays the role of a second proton acceptor.

The tetrahedral intermediate formed as a result of the nucleophilic attack possesses a negatively charged oxygen stabilized by two hydrogen bonds donated by the enzyme. This stabilizing region is known as the "oxyanion hole" because it is occupied by the oxyanion group of the intermediate. A hydrogen bond donated by the main chain nitrogen of the catalytic serine is conserved in the two subtilase families. In sedolisins the second component of the oxyanion hole has evolved to an aspartic acid residue, which constitutes a much better hydrogen bond donor in acidic media than the equivalent asparagine residue of subtilisin (Figures 2.3.A and B).

The following final steps of hydrolysis are common for all subtilases. The tetrahedral intermediate quickly decomposes back to a planar carbonyl group upon disruption of the C-N bond, releasing the C-terminal portion of the substrate. The remaining complex with the peptidase serine residue is covalently bound to the N-terminal segment of the substrate through an ester linkage and is called the acyl-enzyme intermediate. Deacylation proceeds as a hydroxyl group derived from an activated water molecule reacts with the acyl-enzyme forming a second tetrahedral intermediate. When the carbonyl bond reforms, the bond to the serine is broken, freeing the N-terminal half of the peptide chain and allowing the reestablishment of the hydrogen bond system between the catalytic residues.

2.4.4 The Substrate Recognition Sites

The active site of serine proteinases is shaped as a cleft capable of accommodating at least six amino acid residues of the polypeptide substrate or inhibitor (pseudo-substrate). The substrate recognition sites include the polypeptide binding site and the

binding pockets for the side chains of the peptide substrate. Schechter and Berger established in 1967 a systematic nomenclature for substrates and proteinases (Figure 2.4), where the polypeptide substrate amino acid residues are labeled from N- to C- terminus as P_i, ..., P₃, P₂, P₁, P₁', P₂', P₃', ..., P_j; and their respective binding subsites (S_i, ..., S₃, S₂, S₁, S₁', S₂', S₃', ..., S_j) . In this nomenclature, the scissile peptide bond is located between residues P₁ and P₁'.

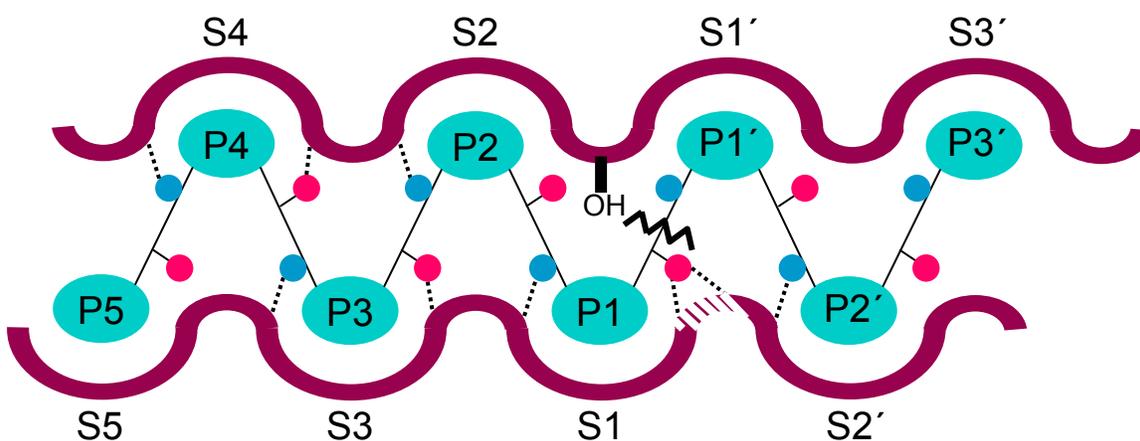


Figure 2.4: Schematic representation of substrate/inhibitor binding to subtilisin-like serine proteinases (subtilases) (Siezen and Leunissen, 1997). Nomenclature P₅-P₃' and S₅-S₃' according to Schechter and Berger (1967). Side chains of the P₅-P₃' residues are shown as cyan ellipses interconnected by peptide bonds (thin black lines) and red and blue spheres indicate their main chain oxygens and nitrogens, respectively. The corresponding binding sites of the enzyme are symbolized by smooth maroon surfaces and a dashed area in the maroon surface indicates the location of the oxyanion hole. Dotted black lines and a jagged line represent the hydrogen bonds between enzyme and substrate/inhibitor and the scissile peptide bond, respectively. The nucleophilic hydroxyl group of the catalytic serine is indicated near the hydrolyzed bond.

2.4.4.1 The Unspecific Main Chain Substrate Binding

Most serine proteinases have no absolute substrate specificity, i.e. they are able to cleave peptide bonds with a variety of side chains adjacent to the scissile peptide bond. Main chain to main chain hydrogen bond interactions between enzyme and substrate are primarily responsible for the unspecific binding of polypeptides. In the subtilase superfamily, the P₄-P₁ backbone is richly hydrogen bonded to the enzyme backbone

(Figure 2.4) forming a short antiparallel β -sheet. In contrast, the C-terminal or leaving portion P1'-P3' of the substrate appears to be held less tightly with only one hydrogen bond observed in the known complex structures.

2.4.4.2 The Specificity Pockets

Even though serine peptidases exhibit no absolute specificity, many of them show a preference for particular side chain located in certain positions of the polypeptide substrate. In general, the specificity of subtilases appears to be largely determined by interactions of the P4-P1 residue side chains with the S4-S1 binding sites of the enzyme (Siezen and Leunissen, 1997). The different shape and hydrophobicity or hydrophilicity of the binding subsites characterize the enzyme pockets and confer specificity for determinate residues in the corresponding positions of the substrate. Though, a general description for the binding sites specificities of subtilases is difficult to establish, since small mutations in the residues that constitute the active-site cleft cause substantial changes in its shape.

2.4.5 Kumamolisin: a Member of the S53 Serine Proteinases

In the early 70s, three proteinases were isolated from *Scytalidium lignicolum* that exhibited optimal catalytic activity at acidic pH but lacked primary structure similarity to any other known carboxyl proteinase (Murao et al., 1972). Later, the finding of at least two catalytically important carboxyl groups and the insensitivity of these proteinases towards most common aspartyl proteinase inhibitors such as pepstatin led to the proposal of the new family of pepstatin-insensitive carboxyl proteinases (Murao et al., 1993; Oda et al., 1994; Oda et al., 1996). More recently, several other enzymes have been classified as members of this family based on sequence similarities, among them the prokaryotic enzymes from *Bacillus* novosp. MN-32 (kumamolisin) (Murao et al., 1993), from *Pseudomonas* sp. 101 (Oda et al., 1994), from *Xanthomonas* sp. T-22 (Oda et al., 1996), and from *Bacillus coagulans* (J-4) (Shibata et al., 1998). This proteinase family is widely distributed and includes eukaryotic homologues such as LYS60 and LYS45, markers for late lysosomes in *Amoeba proteus* (Kwon et al., 1999) and the human lysosomal CLN2, which is particularly interesting due to its involvement in a human disease (Sleat et al.,

1997). Mutations in the encoding *CLN2* gene leading to the lack of active enzyme in the cells are directly associated with a fatal neurodegenerative disorder called classical late-infantile neuronal ceroid lipofuscinosis (LINCL) (Sleat et al., 1997).

A recent report by Rawlings and Barrett (1999b), based on the matching between the known amino acid sequence of CLN2 and fragmentary sequence data of the rat lysosomal enzyme tripeptidyl-peptidase I (TPI), pointed out that the CLN2 protein deficient in LINCL patients was tripeptidyl-peptidase I. This observation and the evidence that TPI was a serine peptidase (Doebber et al., 1978; McDonald et al., 1985) led to questioning of the carboxyl proteinase assignment (Rawlings and Barrett, 1999b). In addition, subsequent mutagenesis work on CLN2 identified two aspartic acid and a serine residue but no histidine as essential for catalytic activity and inhibition studies confirmed inactivation of the enzyme by DCI and DFP, which are typical serine peptidases inhibitors (Lin et al., 2001). However, the unambiguous assignment of all these related enzymes to the serine proteinase SB clan, which previously consisted exclusively of the subtilisin family (S8), became possible only after crystal analysis of *Pseudomonas* sp. proteinase (Wlodawer et al., 2001). The crystal structure revealed a subtilisin-like fold with a novel catalytic triad composed of a nucleophilic serine residue, a glutamic acid and an aspartic acid, which explains their hydrolytic activity at acidic pH (Wlodawer et al., 2001). This family was thus renamed as the serine-carboxyl proteinase (SCP) family (S53) and is now assigned in MEROPS (<http://merops.sanger.ac.uk>) as the second member of the SB clan. Since the presence of the unique catalytic triad (Ser-Glu-Asp or SED in single-letter code) is the defining feature of the family, sedolisin was proposed as the systematic nomenclature for the serine-carboxyl proteinase family (Wlodawer et al., 2003). Accordingly, the *Pseudomonas* sp. 101 peptidase (also called PSCP, for *Pseudomonas* serine-carboxyl proteinase), as the prototype for the family, will be in the following named sedolisin.

Some members of the sedolisin family possess, in addition to the catalytic domains, of N-terminal propeptides, and/or quite large C-terminal domains of unknown function. For example, sedolisin-B, the isolated from *Xanthomonas* sp. T-22, contains a C-terminal propeptide that is removed upon activation of the enzyme (Oyama et al., 1999). Whereas the bacterial serine-carboxyl peptidases share a rather low sequence identity among them

and with CLN2 (30 to 35%), the mammalian homologues identified form a highly conserved subfamily of the sedolisins (Wlodawer et al., 2003). For the full-length enzymes of human, macaque, mouse, rat, cow and dog (563 residues, comprising the catalytic domain and the propeptide), sequence similarity is above 92%, while the identity yields 81% of matches with only a single-residue deletion in the mouse peptidase (Figure 2.5).

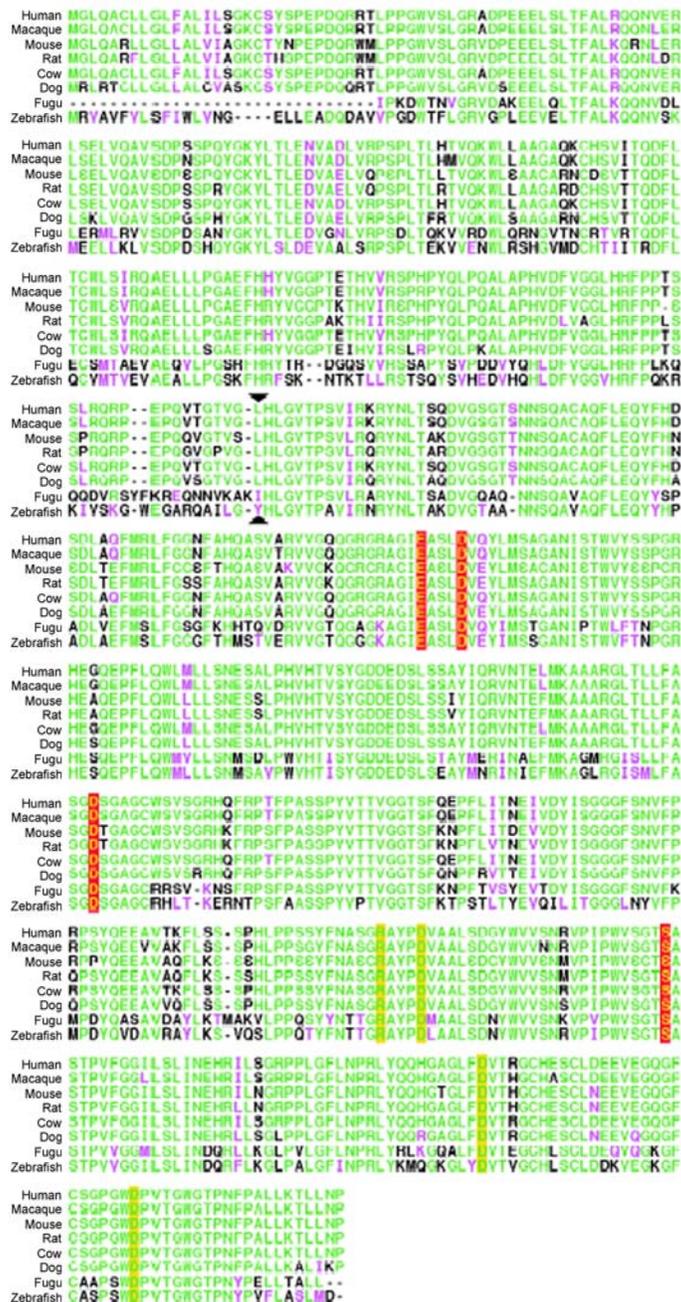


Figure 2.5: Sequence comparisons of the CLN2-like enzymes from mammals and fishes. Conserved residues in the majority of the sequences are colored in green, whereas residues similar in their characteristics are shown in magenta. Residues forming the active site are indicated in yellow on red background, other conserved residues identified as important for stability of the enzyme are marked with yellow background. N-terminus of the mature enzyme (experimentally determined for mammalian enzymes and predicted for fish) is marked with black triangles. Alignment extracted from Wlodawer et al., 2003.

Kumamolisin, which represents the first thermostable member of the sedolisin family, was isolated from the culture filtrate of *Bacillus novosp* MN-32 (Murao et al., 1988; Murao et al., 1993). This is a Gram-negative bacterium found in the acidic hot spring water at the Mt. Aso, Kumamoto prefecture in Japan. Kumamolisin (before kumamolysin), also called kumamolisin serine-carboxyl proteinase (KSCP), is expressed as a 64kDa precursor, which is believed to be secreted upon cleavage of the signal peptide. The 552-amino acid residue zymogen comprises a 188-residue prosegment (with its amino acid residues in the following designated by sequential numbers with a suffix “p” added) and a 364-amino acid mature kumamolisin (numbered starting at the first residue of the catalytic domain with the suffix “e”). Under the acidic conditions of the media, the proenzyme is autocatalytically converted to the 43kDa active peptidase (Oyama et al., unpublished). The proteolytic activity of kumamolisin is optimal at pH 3.0 and 70 °C, and is not detectable below pH 1.0 and above pH 6.0 (Murao et al., 1993). Experimental studies showed that kumamolisin preferentially cleaves the oxidized insulin B-chain between Leu15 and Tyr16, while hydrolysis of the Phe25-Tyr26 peptide bond occurs at a much slower rate.

2.5 Metallopeptidases

Metallopeptidases are characterized by binding of one or two metal ions, which participate directly in the hydrolytic activity. The great flexibility of zinc ion in both, number of ligands and geometry of coordination, is one of the main reasons for the prevalence in nature for this metal over other elements. In addition, zinc ion exhibits a filled outer d shell being thus very stable and unable to further oxidate, opposite to the neighboring iron or copper metals. Even though the majority of metallopeptidases naturally involve zinc in catalysis, these peptidases show altered activities with other substituted metals (Auld, 1995) being normally increased, for example, if cobalt ion is bound instead of zinc.

Metallopeptidases can be classified on the basis of structural similarities, the ion binding site motives and the number of ions implicated in catalysis. In many metallopeptidases, only one zinc is required, but in some families there are two metal ions that act cocatalytically. All metallopeptidases in which cobalt or manganese is

essential require two metal ions, but some families of zinc-dependent metallopeptidases have been identified where two zinc ions act cocatalytically. Although traditionally the HEXXH motive (being X any residue) has been used to detect zinc binding sites when new amino acid sequences are obtained, several other binding motives have been identified in the rapidly increasing number of known metallopeptidase structures (Figure 2.6).

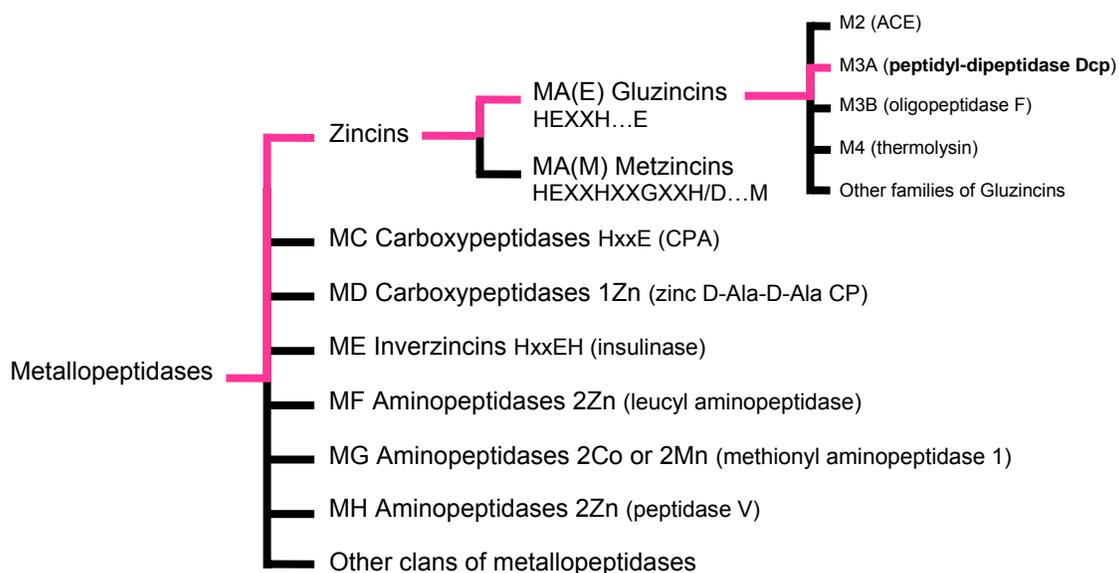


Figure 2.6: Scheme of the metallopeptidases classification. Conserved ion binding motives or metal ions involved in catalysis are indicated beside the family name.

The coordination geometry of the zinc ligands residues as well as their limited mobility play a fundamental role in the strong binding of the ion in the active site ($K_{\text{diss}} \sim 10^{-12}$) (Auld, 1997). Natural zinc binding residues are histidine, glutamate, aspartate and lysine, whereas cysteine residue is found to coordinate the metal ion exclusively in proenzymes, stabilizing the inactive form of the metallopeptidases. A common ion ligand throughout the metallopeptidases is a water molecule, which acts as the primary nucleophile in the hydrolysis of peptide bonds. The coordinating water molecule is activated by the polarization of the metal ion and a neighboring residue, which is normally a glutamate or an aspartate (Figure 2.7). The carbonylic group of the

scissile peptide bond, polarized at its turn by the metal ion, is nucleophilically attacked by the activated water at the carbonyl carbon. The resulting tetrahedral intermediate decomposes by breaking of the peptide bond, after two consecutive protonations of the amido nitrogen and deprotonation of the new incorporated oxygen.

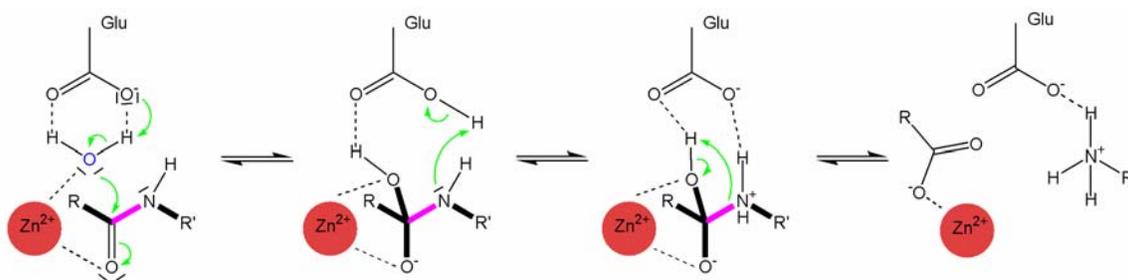


Figure 2.7: Scheme of the postulated catalytic mechanism for metalloproteases. The residues are shown as thin black lines while the substrate is drawn as bold black lines and the scissile peptide bond as a bold pink line. Zinc metal ion is represented by a red sphere. The nucleophilic oxygen of the activated water is highlighted in blue and the electron movement is indicated with green arrows.

2.5.1 The MA(E) Clan of Gluzincins

Up to sixteen different families have been classified within the MA(E) clan (gluzincins) clan based on the conserved HEXXH motif. Although outside of this motif there is no significant sequence similarity between the families of gluzincins, the three-dimensional structures solved (Brown et al., 2001; Natesh et al., 2003) showed that in different families a large portion of the active site region resembles that of thermolysin (Matthews et al., 1972). Both exopeptidases and endopeptidases are found between the metalloproteases belonging to the gluzincins, which are widely distributed among the kingdoms of organisms. Clan MA(E) includes zinc-dependent metalloproteases in which the metal ion ligands are the two histidines in the HEXXH motif and a glutamate several residues C-terminal to the binding motif (Rawlings and Barrett, 1995). The glutamic acid residue conserved in the consensus sequence, together with the zinc ion, are directly involved in the hydrolytic mechanism (Figure 2.7). The scaffold of thermolysin, which is the type example for clan MA(E), is characterized by a five-stranded β -sheet and two

different α -helices carrying the consensus sequence HEXXH and the glutamate, which serves as the third zinc ligand, respectively.

2.5.2 Peptidyl-Dipeptidase Dcp and the Related Enzymes: Angiotensin-I-Converting Enzyme (ACE) and Neurolysin

The name peptidyl-dipeptidase denotes enzymes which cleave C-terminal dipeptides from longer substrates. In 1970, only one activity of this type was known and was attributed to the mammalian angiotensin-I-converting enzyme (ACE) (Yang et al., 1970). The analysis of mutant strains of *Escherichia coli* and of the closely related *Salmonella typhimurium* led to the detection of a series of exopeptidases with partly overlapping substrate specificities including an enzyme with catalytic properties very similar to those of ACE, named peptidyl-dipeptidase Dcp (Deutch and Soffer, 1978). Although initially human and bacterial enzymes were classified into the same family due to their common peptidyl-dipeptidase activity, on the basis of sequencing studies (Hamilton and Miller, 1992), ACE and Dcp were separated out and transferred to the M2 and M3A families, respectively (Figure 2.6).

2.5.2.1 Angiotensin-I-Converting Enzyme

Angiotensin-I-converting enzyme (ACE, previously called peptidyl-dipeptidase A or kininase II) is shown to be a chloride-dependent metalloenzyme that cleaves a dipeptide from the carboxy terminus of the decapeptide angiotensin I to form the potent vasopressor (blood vessel constrictor) angiotensin II (Skeggs et al., 1956). In addition, it inactivates the vasodilator bradykinin by sequential removal of two carboxy-terminal dipeptides (Ehlers and Riordan, 1989; for a review see Riordan, 2003). Thus blood pressure is enhanced by ACE in two different ways: via production of angiotensin II and via inactivation of bradykinin (Waeber et al., 1990). Nevertheless ACE is a broad-specificity dipeptidyl carboxypeptidase and also may act on non-vasoactive peptides. For example, the human enzyme has recently been implicated in the *in vivo* hydrolysis of the tetrapeptide *N*-acetyl-seryl-aspartyl-lysyl-proline, which is involved in the control of hematopoietic stem cell differentiation and proliferation by preventing their recruitment

into S phase (Rieger et al., 1993). *In vitro* ACE hydrolyzes a wide range of substrates including [Met⁵]-enkephalin, β -neoendorphin, dynorphins (Skidgel and Erdos, 1987) and neurotensin, which is a 13-residue neuropeptide (Skidgel et al, 1984).

There are two isoforms of human ACE transcribed from the same gene in a tissue-specific manner (Hubert et al., 1991). In somatic tissues, it exists as a glycoprotein composed of a single large polypeptide chain of 140 kDa, whereas in sperm cells it is a lower-molecular-mass glycoform of 90 kDa. Both forms of angiotensin-I-converting enzyme show the same carboxy-terminal transmembrane and cytosolic sequences but differ in the extracellular part. Somatic ACE consists of two homologous extracellular domains each of which contain an active site, while the testicular form possesses only one of these domains with one active site (Riordan, 2003). The importance of angiotensin-I-converting enzyme in blood pressure regulation leads to intensive inactivation studies and design of potent competitive inhibitors.

2.5.2.2 Peptidyl-Dipeptidase Dcp

Peptidyl-dipeptidase Dcp closely resembles the mammalian ACE on the basis of its susceptibility to common inhibitors and of its specificity for cleavage of C-terminal dipeptides of several oligopeptides. Hence, peptidyl-dipeptidase Dcp has been used in the evaluation of drugs designed to ameliorate human hypertensive disease. However, structural dissimilarity in the overall folds of Dcp and ACE was expected according to the poor amino acid sequence homology and due to the lack of Dcp activity inhibition in the presence of rabbit ACE antibodies (Das and Soffer, 1976). The bacterial enzyme Dcp, which acts as a monomer of 77.5 kDa, is the only C-terminal exopeptidase known from *E. coli* and *S. typhimurium*, apart from the D-Ala-D-Ala carboxypeptidases (Conlin and Miller, 1995). Dcp shows hydrolytic activity on α -N-blocked tripeptides, free tetrapeptides and higher peptides, whereas free tripeptides are not cleaved. Peptides with a proline residue in the penultimate position, with a blocked C-terminus or with a D-amino acid in the C-terminal position are also not accepted as substrates. The enzyme is strongly inactivated in the presence of Cu²⁺, Ni²⁺ and Zn²⁺ (in concentrations higher than 0.1mM), whereas it shows slightly activation by addition of 1mM Mn²⁺, Mg²⁺, Ca²⁺ or

Co^{2+} (Yaron et al., 1972); but unlike ACE, chloride has no influence on the hydrolytic activity of Dcp (Skeggs et al., 1956).

2.5.2.3 Neurolysin

A human relative of peptidyl-dipeptidase Dcp in the M3A family is the neuropeptidase neurolysin, which is widely distributed in mammalian tissues (Brown et al., 2001). Although both enzymes share 24% sequence identity, neurolysin is a strict endopeptidase, opposite to the exopeptidase activity of Dcp and ACE. *In vitro*, the 78 kDa neurolysin cleaves a number of bioactive peptides at sequences that vary widely, and its longest known substrate is only 17 residues in length. The most established role *in vivo* of the human enzyme (along with thimet oligopeptidase) is the hydrolysis of neuropeptide neurotensin between residues 10 and 11 creating shorter fragments that are believed to be inactive (Vincent et al., 1995). Neurotensin peptide is found in a variety of peripheral and central tissues, where it is involved in a number of effects, including thermoregulation, intestinal mobility, and blood regulation (Goedert, 1984). Not appreciably activation of neurolysin was measured in the low concentrations of thiol compounds that enhance the activity of other members of the M3A family, such as thimet oligopeptidase. Neurolysin, is inhibited by EDTA and 1,10-phenanthroline, like other metallopeptidases, whereas specific inhibition with the dipeptide Pro-Ile serves to distinguish neurolysin from thimet oligopeptidase (Serizawa et al., 1995).

Different subcellular localizations have been reported for neurolysin. Experimental analyses show that in tissues of rat and pig the enzyme is predominantly mitochondrial and cytosolic, with the ratio between these two compartments varying with tissue and species (Serizawa et al., 1995). However, the indications that neurolysin contributes significantly to the physiological cleavage of neurotensin has led to emphasis on the fact that a fraction of the endopeptidase may be extracellular. This theory was confirmed by the discovery that an inhibitor of neurolysin protected exogenous neurotensin from degradation in rat brain (Vincent et al., 1997). The existence of two forms of neurolysin that differ only in the length of the N-terminus apparently targets the protein to cytosol and to mitochondria. The longer form contains a cleavable mitochondrial targeting sequence at the N-terminus (Serizawa et al., 1995) and it is directed to mitochondria,

whereas the shorter form, lacking the targeting sequence, remains in the cytosol. It has been proposed that the two forms of the protein arise from the gene using two alternative initiation codons, which is supported by the absence of multiple mRNAs that could undergo mRNA splicing (Serizawa et al., 1997).

2.6 Protein Crystallography

Few techniques are currently available for determining of three-dimensional macromolecular structures. The visualization methods of a protein structure at microscopic level are limited by the applied wavelength. At the scale of angstroms, the radiation needed to analyze atomic distances (e.g. 1.54 Å for a carbon-carbon σ -bond) lies within the spectral range of X-rays, used in protein crystallography, but, at certain speeds, neutrons and electrons can serve as radiation as well. The electron microscopy technique, for example, allows elucidation of macromolecular structures, by focusing the scattered electrons with magnetic fields. The poor signal-to-noise ratio and the fast damaging of the sample in electron microscopy limits the method to three-dimensional protein surfaces at low resolution and makes it more appropriate for very large assemblies.

The ability of X-ray diffraction analysis for solving protein structures at atomic resolution makes protein crystallography the most effective technique for macromolecular studies. This is confirmed by the exponentially growing number of deposited X-ray structures in the Protein Data Bank, since its beginning in the early seventies. Nevertheless, protein crystallography requires large three-dimensional crystals of the sample suitable for diffraction of X-ray, which often represents a limiting step. Another method that permits analysis at high resolution level is nuclear magnetic resonance (NMR). NMR measurements are based on a different principle than radiation scattering, namely: the particular response of atoms to applied magnetic fields depending on their local environment. As NMR allows structural elucidation of soluble proteins, it provides information of dynamic processes that cannot be obtained from the rather rigid surroundings of a crystal. However, NMR is limited to proteins of small molecular weight, up to 35 kDa. Although crystallography, when compared to NMR, gives a more

static description of the macromolecular structures, there are no limits in the size of the molecule to be analyzed. This makes X-ray crystallography the method of choice for studying macromolecules and their complexes with inhibitors or other proteins at an atomic level. A detailed discussion of protein crystallography is beyond the scope of this work and can be found in relevant textbooks (Blundell and Johnson, 1976; Drenth, 1994; Massa, 1994; *Methods in Enzymology*, vol. 276, Part A, 1997).

3 Experimental Procedures

3.1 Materials

3.1.1 Chemicals

Unless otherwise mentioned, the chemicals used had at least analytic purity and were provided by BioRad (Munich), Eppendorf (Hamburg), Fluka (Neu-Ulm), Hampton Research (Laguna Hills, USA), Merck (Darmstadt), Millipore (Eschborn), Quiagen (Hilden), Riedel de Hën (Seelze), Roth (Karlsruhe), Serva (Heidelberg), Sigma (Deisenhofen) and Stratagene (Heidelberg).

3.1.2 Bacterias

For cloning and expression of recombinant proteins, the following *E. coli* cells belonging to different strains were used:

For Cloning:

E. coli K12 strain DH5 α

Genotype: F⁻ *deoR*, *endA*, *gyrA96*, *hsdR17*(r_k⁻ m_k⁺), *recA1*, *relA1*, *supE44*, *thi-1*, Δ (*lacZYA-argF*, Φ 80*lacZ* Δ M15F⁻ λ ⁻).

For Expression of Wild Type Proteins:

E. coli K12 strain JM83 (Yanischperron, 1985).

Genotype: F⁻ *ara*, D (*lac-pro AB*), *rpsL* (= *strA*), F80, *lacZDM15*.

For Expression of SeMet-Substituted Proteins:

E. coli B strain B834(DE3)

Genotype: F⁻ *ompT*, *hsdS6*(r_B⁻ m_B⁻), *gal*, *dcm*, *met*, (DE3).

E. coli B strain B834(DE3)RIL

Genotype: F⁻ *ompT*, *hsdS6*(r_B⁻ m_B⁻), *gal*, *dcm*, *met*, (DE3).

3.1.6 Devices

The centrifuges used were from Beckmann, Eppendorf (Hamburg), Hettich (Kirchlengern) and Sorvall (Langenselbold). Pipettes were from the company Eppendorf (Hamburg) and gene pulser from BioRad (Munich). All chromatography steps were performed with FPLC or ÄKTA (Amersham Pharmacia Biothech. AP, Uppsala, Sweden) systems. Diffraction data was recorded in-house using an image-plate detector (MAR Research) coupled to a RU-200 rotating anode X-ray generator (Rigaku International Corp., Kyoto, Japan) or at the DESY BW6 beamline (Deutsches Elektronen Synchrotron, Hamburg) with a MAR Research CCD detector. Data processing and calculations were carried out in digital ALPHA-stations (OSF) or SGI (Silicon Graphics Inc., Mountain View, USA) work-stations and computer servers (IRIX).

3.2 Molecular Biology Methods

All instruments used were sterilized with 70% ethanol or autoclaved before and after working with bacteria. All buffers employed were sterilized by autoclaving or sterile filtration.

3.2.1 Plasmid Preparation

Glycerol cultures of transformed *E. coli* strain JM83 or DH5 α were grown overnight at 37° C in 150 ml LB medium with 100 mg/ml ampicillin. After centrifugation, plasmid DNA was extracted from the cells following the Qiagen Mini Prep Kit protocol and stored at -20 °C.

3.2.2 Restriction Analysis

Plasmid DNA was incubated for 3 h with 1-2 U of the required restriction enzyme per μ g DNA in the buffer recommended by the manufacturer.

3.2.3 Agarose Gel Electrophoresis

Agarose gel electrophoresis was used for separation and detection of DNA fragments of different lengths. To prepare an agarose gel a mixture of 0.7% agarose (w/v) (Biozym) with TAE-buffer was warmed in the microwave until the agarose was completely solubilized. Ethidium bromide was added to a final concentration of 0.5 µg/ml. The mixture was poured in an electrophoresis chamber, which had been previously sealed with custom-made dams. After solidification at room temperature, the gel was covered with TAE-buffer. The DNA samples were mixed with 6x gel loading dye buffer and pipetted in the corresponding gel wells. The gel was electrophoresed at a constant 50 V and DNA bands were visualized under UV light ($\lambda = 254 \text{ nm}$).

- TAE-buffer: 1 mM EDTA
 20 mM Acetic Acid
 40 mM Tris, pH 8.0
- 6x Gel Loading buffer: 0.2 g Bromophenol Blue
 0.2 g Xylene Cyanol
 10 ml 1x TAE-buffer
 80 ml Glycerin (87 %)

3.2.4 *E. coli* Transformation by Electroporation

Fifty µl electrocompetent cells were thawed on ice and mixed with 2 µl plasmid DNA. The cell/DNA mixture was pipetted into an electroporation cuvette at 4° C. The cells were electroporated at 1650 V in the gene pulser. One ml LB medium was added to the electroporated cells and the suspension transferred into a culture tube was incubated with shaking during 1 h at 37° C. The incubated cells were subsequently plated and/or further grown in liquid cultures.

3.2.5 Cell Cultures

3.2.5.1 Plate Cultures

Autoclaved LB medium with 15 g/l agar was cooled down to 50° C, mixed with 100 mg/ml ampicillin and poured into sterile plates (LB/amp plates). After solidification, plates were stored at 4° C. For growing colonies, 100 µl ampicillin-resistant bacteria were plated and incubated at 37° C overnight.

3.2.5.2 Liquid Cultures

Five ml autoclaved LB medium supplemented with 100 mg/ml ampicillin (LB/amp medium) were inoculated with a single colony picked from the agar plate. After overnight incubation at 37° C, the resulting incubated pre-culture was used either for plasmid preparation or for inoculation of larger cultures in ratios of 1:100 to 1:1000.

3.2.5.3 Glycerol Cultures

To freeze liquid cultures, 150 µl glycerin (87% v/v) was mixed in sterile conditions with 0.85 ml of an exponentially growing cell culture. The mixture was immediately shock-frozen in liquid nitrogen and stored at -80° C.

3.3 Protein Biochemistry

3.3.1 Protein Concentration Determination

Comparison of the absorbance value of a test solution with that of a standard solution provides a sensitive method of quantification as aromatic amino acid residues tyrosine and tryptophan exhibit an absorption maximum at 280 nm (Mach et al., 1992). Thus, the molar extinction coefficient, $\epsilon_{280\text{nm}}$, varies as a linear function of the aromatic and cysteine residue content (Gill and Vonhippel, 1989). Knowing the value of $\epsilon_{280\text{nm}}$, the protein concentration can be determined according to the Beer-Lambert law:

$$A = \epsilon_{280\text{nm}}cd$$

A = absorbance

c = protein concentration

d = length of the light path in solution (cm)

ϵ = molar extinction coefficient

3.3.2 Apparent Molecular Weight: SDS Polyacrylamide Gel Electrophoresis

Analytical protein separation, as result of different migration depending basically on the molecular weight, is routinely carried out using denaturing discontinuous SDS polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli et al., 1970). Proteins in loading dye buffer, strongly negatively charged by bound SDS molecules, migrate responding to an external electric field. The porous SDS-PAGE allows proteins separation by retarding migration of the enzymes inversely proportional to their molecular weights.

SDS-PAGE used consisted of a stacking gel (5% AA), which contains the sample wells, and a 12% AA separating gel. The separating gel mixture was poured into the gel apparatus (MPI, Martinsried) between the glass plates and straightened with a 2-mm isopropanol layer per gel. After complete polymerization of the separating gel and isopropanol removal, the stacking gel solution was pipetted and 10- to 12-wells combs were placed between the glass plates. The finished gels were packed in aluminium foil and stored in a water-containing box at 4° C.

- 1x Loading Dye Buffer: 0.1 M Tris-HCl, pH 6.8
0.1% Bromophenol Blue
4% SDS
20% Glycerol
0.25 M DTT (added fresh before use)
- Separating Gel (for 8 gels): 10.5 ml 1.5 M Tris-HCl, pH 8.8
16.9 ml 30% AA/0.8% Methylene bis-AA
13.9 ml H₂O
35 μ l TEMED (added at the end)

- Stacking Gel (for 8 gels):
 - 150 μ l 10% APS (added at the end)
 - 200 μ l 10% APS buffer (added at the end)
 - 1 ml 30% AA/0.8% Methylene bis-AA
 - 2.5 ml 0.5 M Tris-HCl, pH 6.8
 - 12.8 ml H₂O
 - 20 μ l TEMED (added at the end)
- APS Buffer:
 - 10% (w/v) Ammonium persulfate

For performing SDS-PAGE, protein samples were mixed with loading dye buffer and after incubation for 2 min at 96° C and centrifugation, loaded into the wells. In addition, a sample containing a protein mixture of known composition (molecular weight marker; Broad Range SDS-Marker, BioRad, Munich) was included in all experiments to determine molecular masses of proteins in the neighboring lanes. SDS-PAGE were run at 140 V in SDS running buffer and subsequently stained by heating and shaking (10 min) the gel in Coomassie Brilliant Blue R250 staining solution. Finally destaining buffer was used to wash repeatedly the gels until the background Coomassie was complete removed.

- Running Buffer:
 - 0.1% (w/v) SDS
 - 25 mM Tris-HCl, pH 8.3
 - 200 mM Glycine
- Staining Solution:
 - 0.1% Coomassie Brilliant Blue R-250
 - 12.5% (v/v) Acetic acid
 - 25% (v/v) Ethanol
- Destaining Solution:
 - 12.5% (v/v) Acetic acid
 - 25% (v/v) Ethanol

3.3.3 Protein Transfer and Western Blots

Proteins were transferred from a non stained polyacrylamide gel to a polyvinylidene fluoride (PVDF) membrane by electroblotting in a Trans-Blot transfer cell (Towbin et al., 1979). Before use, the hydrophobic PVDF membrane was submerged in

Automated N-terminal Edman degradation of a protein consists of repetitive cycles of Edman chemistry followed by PTH analysis on a HPLC column.

Purified proteins or solubilized crystals were fixed and stained in a PVDF membrane (as described in section 3.3.3). The Edman Degradation method was applied to the required protein band from which N-terminal sequence (5-7 amino acid residues) was determined. Micro-sequencing was kindly carried out by R. Mentele (Department for Protein Chemistry, Max-Planck-Institute for Biochemistry, Martinsried), using an Applied Biosystems Sequencer.

3.3.5 Mass Spectroscopy

Mass spectrometry (MS) allows high-accuracy determination of the ions mass in the gas phase based on the measurement of their mass-to-charge ratio. Analyte ions can be transferred to the gas phase from liquid or solid samples through techniques such as electrospray or MALDI (matrix assisted laser desorption ionization). The matrix assisted laser desorption ionization is a soft technique that allows the mass determination of large biological molecules such as proteins, since the molecule being ionized does not fall apart or break-up during the process (Karas and Hillenkamp, 1990). In MALDI, the analyte is first co-crystallized with a large molar excess of a matrix compound, usually a UV-absorbing weak organic acid. This analyte-matrix mixture is pulsed with a UV laser radiation resulting in the vaporization of the matrix which carries the analyte with it. The matrix plays a key role by strongly absorbing the laser light energy and acting as a proton donor and receptor, which serves to ionize the analyte in both positive and negative ionization modes, respectively. The resulting ionized analyte in vapor phase is submitted to a force produced by electric or magnetic fields. As this force causes an acceleration that depends directly on their mass and ionic charge, the spectrometer separates and quantifies the ions in time or space based on their mass-to-charge ratio.

The molecular masses of purified protein samples were determined by MALDI mass spectrometry, which was gently carried out by M. Zwaboda (Department for Protein Chemistry, Max-Planck-Institute for Biochemistry, Martinsried).

3.3.6 Protein Concentration

Purified proteins were concentrated using the centrifugation concentrators Centriprep for volumes of 15 ml or Centricon for volumes less than 2 ml (Millipore, Eschborn). The final concentrated protein samples were filtrated through an Ultrafree (Millipore, Eschborn) membrane (0.22 μm) in order to completely remove suspended particles.

3.4 Protein Production

3.4.1 Protein Over-Expression

3.4.1.1 Peptidyl-Dipeptidase Dcp Over-Expression

pDcp-coding plasmid was isolated from an overnight glycerol culture (*E. coli* strain JM83) and prepared in *E. coli* strain DH5 α as described above (section 3.2.1). Fresh *E. coli* strain JM83 cells were transformed with the purified pDcp plasmid and after 1 h incubation plated in LB/amp plates. A single colony was picked for inoculation of 150 ml LB/amp medium, which was incubated overnight with shaking at 37° C. Cells from the overnight pre-culture were used to inoculate 3 l of selective LB/amp medium (freshly prepared), and the cell culture was grown at 37° C to an optical density (OD_{600nm}) of 1.0 – 1.2. After cooling down to RT, protein expression was induced with 1 mM IPTG and cells grown for additional 15 h at RT. Cells were harvested by centrifugation at 4200 rpm, 4° C for 25 min.

3.4.1.2 SeMet-Peptidyl-Dipeptidase Dcp Over-Expression

After transformation of *E. coli* strain B834 with the plasmid pDcp, the cells were incubated for 1 h and plated in LB/amp plates. One hundred and fifty ml LB/amp medium inoculated with a single picked colony were grown overnight with shaking at 37° C. Five liters of NM medium complemented with 100 mg/ml ampicillin were inoculated with 35 ml overnight culture. Cells in NM medium were grown for 24 h to an OD_{600nm} of 1.0 – 1-2. After cooling down and inducing with 1mM IPTG, the culture was

grown with shaking for further 24 h at 37° C. Cells were centrifuged at 4200 rpm, 4° C for 25 min, and the supernatant discarded.

3.4.2 Cell Disruption

Harvested cells were resuspended in 10 ml of lysis buffer per liter culture, frozen and thawed, and subsequently sonicated on ice (Branson, microtip, level 9, *duty cycle* 60%, 4 times for 3 min). Disrupted cells were centrifuged (25000 rpm) for 30 min at 4° C and the resulting pellet was discarded.

- Lysis Buffer: 20 mM EDTA
100 mM Tris, pH 8.0

3.4.3 Protein Preparation

3.4.3.1 Purification of Native and SeMet Peptidyl-Dipeptidase Dcp

Throughout the purification steps, protein solutions were always handled at 4° C. The following purification procedure was applied to both native and SeMet-derivatized peptidyl-dipeptidase Dcp.

After centrifugation of the disrupted cells, the supernatant was dialyzed twice against Tris buffer (10 mM Tris-HCl, pH 8.0) in a ratio 1:100. Dialyzed protein was loaded on an anion exchange DE52 (Whatman, Maidstone, UK) 50-ml FPLC column (Amersham Pharmacia Biotech AP, Uppsala, Sweden) previously equilibrated in Tris buffer. The column was washed with three column volumes of the equilibrating buffer and the protein was eluted with a 0 – 400 mM NaCl gradient in the same Tris buffer. Collected fractions were analyzed by SDS-PAGE and UV/Visible spectrum.

The pooled fractions were mixed with 3 M AS solution to a final concentration of 0.8 M AS. The mixture was loaded on a Phenyl Sepharose High Performance (Amersham Pharmacia Biotech, AP, Uppsala, Sweden) 20-ml FPLC column (Amersham Pharmacia Biotech, AP, Uppsala, Sweden), which was previously equilibrated in Tris-HCl (pH 8.0), 800 mM AS. Following the column wash, the protein was eluted with a 800 to 0 mM AS gradient in Tris buffer. Protein-containing fractions were determined by SDS-PAGE analysis and dialyzed twice against buffer consisting of 20 mM Tris-HCl (pH 8.0), 50

mM NaCl (ratio 1:100). Dialyzed purified protein was filtered, aliquoted and stored at –20° C after concentration to 13 -14 mg/ml in the case of native Dcp and to 5 - 6 mg/ml for the SeMet-Dcp sample. Protein concentration was determined by absorption spectroscopy at 280 nm (see section 3.3.1) using a calculated extinction coefficient of $113630 \text{ M}^{-1} \text{ cm}^{-1}$ (Protparam tool from Expasy: <http://www.expasy.org/cgi-bin/protparam>).

Protein purity and correct folding were verified by SDS-PAGE, N-terminal peptide sequencing, MALDI-MS and CD spectroscopy. Complete Selenomethionine incorporation was confirmed by MALDI-MS.

3.4.3.2 Peptidyl-Dipeptidase Dcp/Inhibitor Complex Purification

Stored purified peptidyl-dipeptidase Dcp was thawed on ice and mixed with a 50 mM N-1450 (Bachem, Weil am Rhein) inhibitor (solubilized in bidistilled water) to a 0.25 mM final inhibitor concentration. Size exclusion chromatography of the protein-inhibitor complex was performed at 4° C in a ÄKTA (Amersham Pharmacia Biotech AP, Uppsala, Sweden) equipment with a Superdex 200 column previously equilibrated in 80 mM Tris-HCl (pH 8.0), 200 mM NaCl. Protein elution was carried out at a flow of 0.20 ml/min and followed online by UV absorbance spectrum at 280 nm. Fractions containing the protein-inhibitor complex were pooled and concentrated to 15 mg/ml and 6 mg/ml for native Dcp and SeMet-Dcp, respectively.

3.5 Crystallography Methods

3.5.1 Crystallization

The vapor-diffusion technique is the most commonly used method for the crystallization of macromolecules. Vapor diffusion refers to equilibration between two different concentrated solutions that are connected by vapor phase. In crystallization trials, a *sitting* or *hanging* drop, composed of a mixture of protein solution, crystallization reagent and eventually additives, is allowed to equilibrate against a much larger volume of a reservoir. Typically the drop contains a lower reagent concentration than the

reservoir because of mixing with the protein sample. To achieve equilibrium, volatile species (water and/or organic solvents) diffuse until vapor pressure in the droplet equals the one in the reservoir. As equilibration proceeds, the concentration of both protein and precipitant increase in the drop, which can lead to crystallization of the protein.

Crystallization trials were carried out in CrystalClear Strips on Nunc plates (Emerald BioStructures, Bainbridge Island, USA) and in CrysChem plates (Charles Supper Company, Natick, USA) for sitting-drop experiments and in Linbro tissue culture plates with siliconized cover slides (Hampton Research, Laguna Hills, USA) for hanging-drops. To cover a broad spectrum of buffers, precipitation agents and salts, crystal screenings were performed using the House Factorials (in-house made crystal screen) and several of the Crystallization Kits (Hampton Research, Laguna Hills, USA). Promising conditions, where needles or micro-crystals appear, were optimized independently varying several (e.g. pH, protein concentration, salt concentration, additives) as well as by trying different protein:solution drop ratios. Crystallization experiments were conducted at both 21° C and 4° C.

3.5.2 Inhibitor Soaking and Co-Crystallization

Both techniques, soaking and co-crystallization, are suitable to obtain crystals of a protein in complex with an inhibitor, substrate or substrate analog. Soaking can only be used, if crystals are available, and it is especially appropriate, if these crystals are stable and the soaked molecule is small. Otherwise, diffusion and binding of the soaked molecule into the crystal can cause its disordering (sometimes detectable as crystal crashing). Co-crystallization is adequate if crystals are damaged after soaking or to stabilize a protein and favor growth of bigger crystals.

3.5.2.1 Kumamolisin-Inhibitor Complexes Co-Crystallization

For co-crystallization experiments, (Figure 3.1) stock solutions of the inhibitors (Ac-IPF-CHO and Ac-IAF-CHO, kindly supplied by Dr. K. Oda, Department of Applied Biology, Faculty of Textile Science, Kyoto, Japan) were prepared by dissolving them in DMSO, followed by 10-fold dilution with crystallization buffer (to ~4 mg/ml inhibitor concentration). 0.3 µl of these solutions were added to two-µl drops containing a mixture

of native kumamolisin and crystallization buffer to obtain a final 10 mM inhibitor concentration.

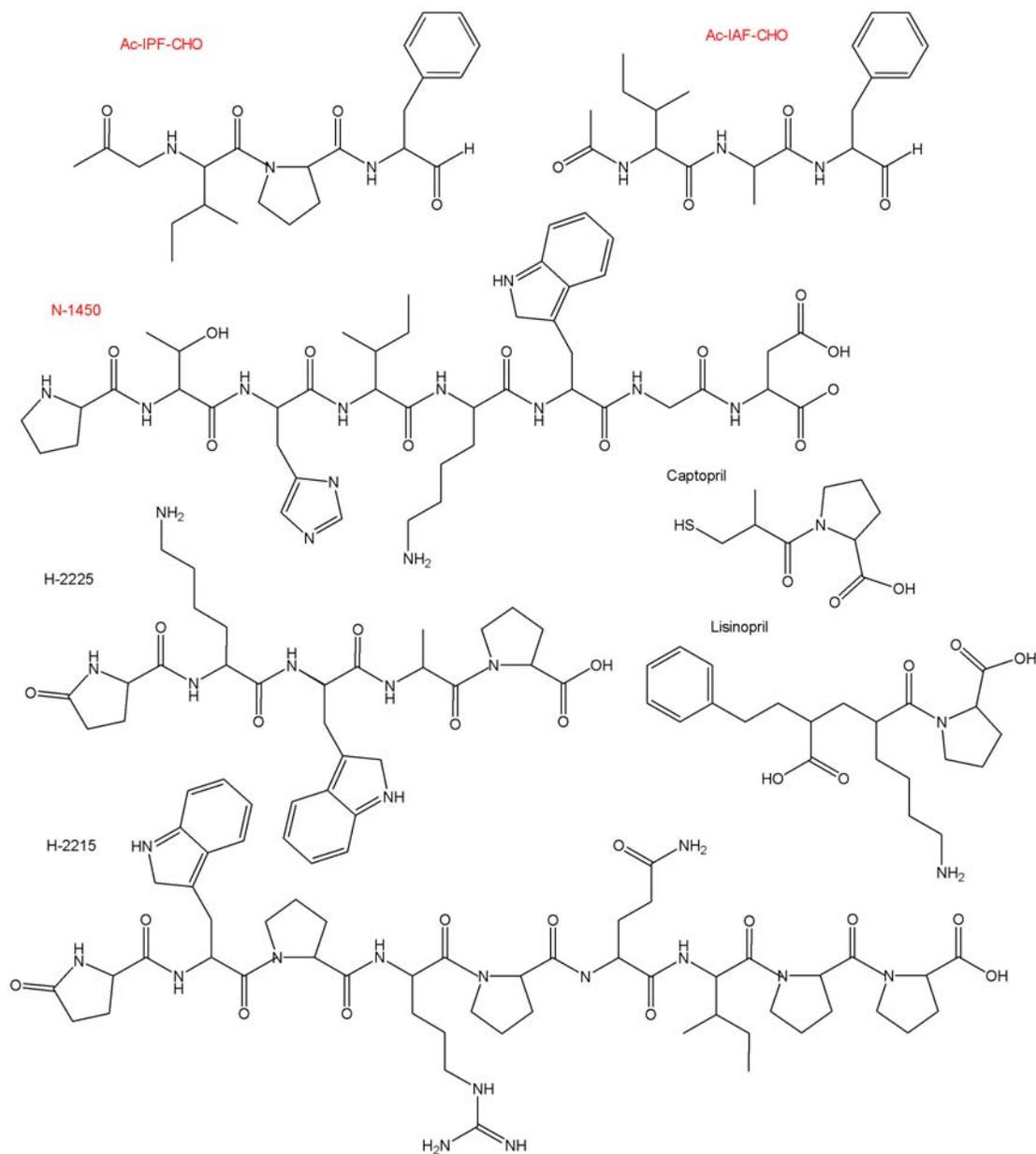


Figure 3.1: Structure formulas of kumamolisin and peptidyl-dipeptidase Dcp inhibitors used for co-crystallization experiments. Ac-IPF-CHO and Ac-IAF-CHO inhibitors of kumamolisin were synthesized in Faculty of Textile Science of Kyoto (Japan), whereas N-1450, H-2225, H2215, captopril and lisinopril were acquired at the company BACHEM (Weil am Rhein). Names written in red indicated the inhibitors successfully co-crystallized.

3.5.2.2 Peptidyl-Dipeptidase Dcp/Inhibitor Complex Co-Crystallization

To favor crystallization of Dcp co-crystallization trials were carried out with several inhibitors (Figure 3.1), e.g. N-1450, H-2215, H-2225, captopril, and lisinopril (Bachem, Weil am Rhein). Co-crystallization of Dcp, both native and Selenomethionine derivative, succeeded after size exclusion chromatography (see section 3.4.3.2) of the protein in complex with N-1450 inhibitor (H-PYHIKWGD-OH).

3.5.3 Crystal Seeding

If the concentration of crystallizable protein is plotted against the concentration of a precipitant, the resulting diagram divides the space into several areas depending on the physical state of the protein (Figure 3.2; extracted from www.douglas.co.uk/rep1.htm). If the protein concentration needed for nuclei formation (Figure 3.2.2. nucleation zone) is substantially higher than the one needed for crystal growth (Figure 3.2.3. metastable zone), it might be useful to add reagents like salts or detergents that increase the protein solubility and help to “push it” across the seed formation threshold. However, it is often found that crystals grown in the metastable zone are better ordered and thus diffract better than crystals grown at higher concentrations. To grow large crystals in the metastable zone, small crystals, which act as nuclei, can be transferred into crystallization drops with lower protein and/or precipitant concentration. Crystal seeding can be performed as *microseeding*, by introducing only fragments of crushed crystals, or as *macroseeding*, by transferring an entire crystal to a new drop.

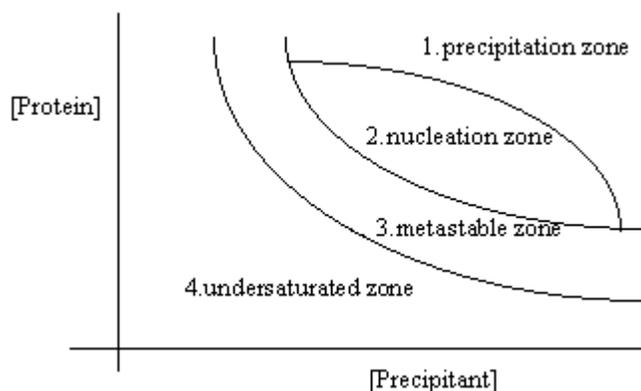


Figure 3.2: Schematic phase diagram of protein concentration versus concentration of precipitant.

1. At high concentrations of both protein and precipitant, the protein precipitates as an amorphous material.
2. At lower concentrations, crystal nuclei may form, which may grow to form diffracting crystals.
3. At still lower concentrations, nuclei will not form, so generally no crystals appear. However, if a nucleus or crystal is placed in such a solution, it will grow to form a large crystal. This area, where crystal growth but not nucleation takes place, is sometimes referred to as "the metastable zone".
4. At still lower concentrations, the protein is completely soluble.

3.5.3.1 Peptidyl-Dipeptidase Dcp Microseeding

Microseeding appear to be an essential step for obtaining crystals of SeMet-Dcp. Seed transfer of crushed native Dcp crystals to a variety of conditions (around the crystallization buffer of native Dcp) produced SeMet-Dcp diffracting crystals. Seeding was carried out by touching grown crystals of native Dcp with a cat whisker, previously washed in ethanol, and streaking the whisker across the centre of freshly pipetted drops, which contained SeMet-Dcp mixed with crystallization buffer in a concentration approximately 80% of that used for growing of native Dcp crystals. The whisker was washed and the microseeding procedure was repeated every 4 drops.

3.5.4 Crystal Mounting

First tests for inherent diffraction quality were carried out measuring the crystals mounted in capillaries at 21° C. The test crystal was picked up from the droplet and mounted in siliconized quartz capillary (1 mm Ø, Mark-Rörchen, Berlin) between two mother liquor layers, to prevent drying out of the crystal. After sealing it with wax (Harvard, Berlin), the capillary was fixed on a goniometer head.

To overcome crystal damage in the X-ray beam, crystals mounted in cryo-loops (Hampton Research, Laguna Hills, USA) were shock frozen at 100 K with a nitrogen stream cooling system (Oxford Cryosystems, Oxford, UK). Before freezing, crystals were shortly incubated (10 - 20 s) in a mixture of mother liquor buffer and cryoprotectant agent (20% to 30% (v/v) glycerol or 15% (v/v) MPD) to prevent strong powder diffraction rings caused by formation of ice crystals in the solvent. Datasets used for

solving all structures of kumamolisin and peptidyl-dipeptidase Dcp were collected from flash frozen crystals.

Crystals able to be measured either in capillary at 21° C or in cryo-loop at 100 K were centered in the X-ray area adjusting the goniometer head with help of a monitoring camera.

3.5.5 Data Collection

High-resolution datasets used for solving the structures of kumamolisin (native and mutants) and pro-kumamolisin as well as MAD and native data of peptidyl-dipeptidase Dcp were collected at wiggler beamline 6 (BW6) of DESY (Deutsches Elektronen Synchrotron, Hamburg) on a Mar Research CCD detector. Crystals of inhibited kumamolisin and of the native enzyme in different crystallization conditions were taken on our in-house Rigaku rotating-anode X-ray generator (Tokyo, Japan) at a wavelength of $\lambda_{\text{Cu}}=1.5418 \text{ \AA}$ combined with a MAR Research image-plate system (Hamburg).

All datasets were indexed and integrated using the programs DENZO (Otwinowski and Minor, 1993) or MOSFLM6.11 (Leslie, 1991). Programs supported by the Collaborative Computational Project N° 4 (CCP4, 1994) were used for further steps of merging, scaling and truncation.

3.5.6 Phasing

The goal of a crystallographic experiment is to calculate the distribution of electron density in the asymmetric unit of the crystal in order to be able to place an atomic model of the crystallized molecule therein. The electron density distribution ρ for every point (x,y,z) can be calculated as a Fourier summation over all structure factors $F(h,k,l)$:

$$\rho(x,y,z) = V^{-1} \sum_{hkl} |F(h,k,l)| e^{i\varphi(h,k,l)} e^{-2\pi i(hx+ky+lz)}$$

While the structure factor amplitude $|F_{hkl}|$ can be derived from the measured intensity, information about the phase angle $\varphi(h,k,l)$ is lost, which is commonly referred

to as the *phase problem* of crystallography. Experimentally the phase problem can be solved by a number of methods:

- Molecular Replacement (MR)
- Multiple Isomorphous Replacement (MIR)
- Multiple-wavelength Anomalous Dispersion (MAD)
- Direct Methods

3.5.6.1 The Patterson Function

An important part of the methods for solving the phase problem use the *Patterson function* to obtain information about the phases. While it is not possible to calculate an electron density map without phase information, the intensities of the reflections can be used to calculate a Patterson function $P(u,v,w)$, which gives information of the interatomic distances in the crystal:

$$P(u,v,w) = V^{-1} \sum_{hkl} |F(h,k,l)|^2 \cos 2\pi(hu+kv+lw)$$

In the Patterson function, u , v and w represent the axis of the Patterson cell unit, whose dimensions are identical to those of the crystal unit cell x , y and z . Distances between atoms in the real structure show up as vectors from the origin to maxima in the Patterson cell unit, forming a so called Patterson map.

3.5.6.2 Kumamolisin was solved by Molecular Replacement

Molecular Replacement is a method to obtain a starting set of phases and, hence, a first model of the measured protein using a homologous enzyme with known structure as a search model. To this propose, the search molecule must be oriented and translated in the unit cell of the target molecule in such a manner that maximum overlap is achieved between the calculated diffraction pattern of the models and the observed diffraction pattern of the protein to solve. Knowing the coordinates of the model positioned in the cell unit of the target molecule, it is possible to calculate the structure factor amplitudes and the phase angle for the model. A first electron density of the target enzyme can be estimate with its experimental structure factor amplitudes and the calculated model

phases. For positioning of the search model, the rotation and translation functions are calculated as correlation functions between observed and calculated Patterson maps.

The availability of a three-dimensional structure from the homologous protein sedolisin (Wlodawer et al., 2001) allowed to solve the crystal structure of native kumamolisin by molecular replacement using the program AMORE (Navaza, 1994). All other complexes, mutants and the full-length zymogen were solved as well by molecular replacement with the coordinates file of native kumamolisin and the same program.

3.5.6.3 Peptidyl-Dipeptidase Dcp was solved by MAD

MAD is a method for solving protein structures based on the anomalous diffraction properties of heavy atoms inside a protein crystal. Upon interaction with the atoms in a crystal, the oscillating electrical field of an X-ray photon induces an oscillation of equal frequency in the electron hull of the atom. The electrons act as oscillating dipoles emitting secondary radiation of the same frequency as the incident radiation, but with a phase difference of 180° . This elastic or coherent diffraction is no longer valid if the incident X-ray wavelengths have a maximum atomic absorption coefficient in the irradiated element. On these absorption edges, the photon energy of the X-rays is sufficient to eject an electron from the atom and so the absorbed energy will not contribute to the scattering of the atom at this wavelength. In addition, it is incorrect to consider electrons as free particles when measuring close to an absorption edge. As a result of the interaction of the electrons with the nucleus, which is more significant in heavy atoms, the shift of the diffracted beam phase differs from the regular 180° and is called *dispersion* effect. As a consequence of the dispersion and absorption effects, the structure factors of Friedel pairs, F^+_{PH} and F^-_{PH} (reflections with identical structure factor amplitudes and phase angles with reverse signs), are no longer equal. The difference of the intensities of the Friedel pairs can be used to search for the anomalous scatterers positions. An anomalous difference Patterson map calculated using only the structure factors amplitudes of the anomalous scatterers will show their distance vectors and their positions can be determined.

For experimental phases, selenium positions in Selenomethionine peptidyl-dipeptidase Dcp were found with RSPS (CCP4 suite, 1994). Phasing and following phase

improvement by electron density modification were carried out using MLPHARE (CCP4, 1994) and the program DM (Cowtan, 1994), respectively. In addition the heavy atom positions were found and phased with CNS (Brunger et al., 1998). Solvent flattening (CNS), taking into account the solvent crystal content, was used to improve the resulting electron density.

3.5.7 Modeling and Refinement

Atom model building was done on an SGI graphics workstation using the program MAIN (Turk, 1996). Changes in the model were minimized with the same program according to the stereochemical parameters and the electron density displayed.

The model was improved by iterative rounds of model building and refinement with CNS (Brunger et al., 1998) using overall anisotropic B-factor correction for *F_{obs}*, torsion angles molecular dynamics, bulk solvent correction, and maximum likelihood or standard crystallographic residual for individual temperature-factors refinement. In the final stages of refinement solvent molecules were automatically built with CNS and sulphate ions were introduced manually in *F_{obs} - F_{calc}* maps with 2.5 σ contouring level.

3.5.8 Structural Analysis and Graphical Representation

Secondary structure assignment and quality assessments of the refined model were carried out with PROCHECK (Laskowski et al., 1993), and superposition of coordinate files was done with the help of the programs TOP3D (CCP4, 1994) and the SiliconGraphics (SGI) TURBOFRODO Version 5.5 (Roussel and Cambillau, 1991). Interactions at the linker peptide interface with the catalytic domain were calculated with the Protein-Protein Interaction Server (<http://www.biochem.ucl.ac.uk/bsm/PP/server>), while accessibility calculations were performed with NACCESS (Hubbard and Thornton, 1993).

Illustrations were prepared using Molscript (Kraulis, 1991) or Bobscript (Esnouf, 1997; Esnouf 1999) and subsequently rendered with Raster3D (Merritt and Bacon, 1997). Some active site representations were carried out with Dino (<http://www.bioz.unibas.ch/~xray/dino>). Electrostatic surface potentials and molecular

surfaces were calculated using the program GRASP (Nicholls et al., 1993) and rendered with Raster3D (Merritt and Bacon, 1997). Sequence alignments were performed with the Wisconsin GCG Package and plotted with Alscript (Barton, 1993). The programs ChemDraw (CambridgeSoft, Cambridge, USA) and Microsoft PowerPoint (Microsoft Corporation, USA) were used for schematic representations. All images were edited using PHOTOSHOP (Adobe System Inc., USA).

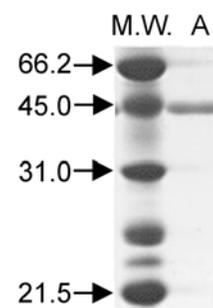
4 Results

4.1 Kumamolisin a Serine-Carboxyl Proteinase from *Bacillus novosp.* MN-32

4.1.1 Protein Crystallization

Recombinant kumamolisin serine-carboxyl proteinase was over-expressed and purified by Dr. K. Oda and co-workers (Department of Applied Biology, Faculty of Textile Science, Kyoto, Japan) and kindly provided for the crystallization assays. Two crystals forms, both of the monoclinic space group $P2_1$ and suitable for X-ray diffraction studies, were obtained by sitting-drop vapor-diffusion procedures. Crystals of form A containing one molecule per asymmetric unit could only be grown from solutions of the uninhibited protein. Form B crystals, with two molecules in the asymmetric unit, were obtained from the native protein and by co-crystallization of the enzyme with two different inhibitors, namely *N*-acetyl-isoleucyl-alanyl-phenylalaninal (Ac-IAF-CHO) and *N*-acetyl-isoleucyl-prolyl-phenylalaninal (Ac-IPF-CHO) (see section 3.5.2.1). The corresponding crystals are therefore denoted as B-IAF and B-IPF, respectively (Table 1).

Figure 4.1: SDS-polyacrylamide gel electrophoresis of purified recombinant kumamolisin. Denaturated samples were loaded onto a 12.5% SDS-polyacrylamide gel, electrophoresed, and stained with Coomassie Brilliant Blue R-250. The Molecular Weight Marker (M.W.) indicates the position of proteins with apparent molecular weights from 66.2 to 21.5 kDa. Lane A shows the purified enzyme (3 μ g).



First crystallization conditions (Figure 4.2.A), found after initial crystallization screening, were optimized by decreasing the ammonium sulphate concentration and varying the pH of the buffer. Long single crystals (Figure 4.2.B) were grown by mixing 1 μ l of a 9 mg/ml kumamolisin solution in 25 mM sodium chloride, 50 mM sodium acetate, pH 5, with 1 μ l reservoir solution, consisting of 1.2 M ammonium sulphate, 0.1 M sodium acetate (pH 5.2) for crystal form A, and of 0.4 M ammonium sulphate, 0.1 M

sodium acetate (pH 4.5) for crystal form B. Inhibitor stock solutions were prepared by dissolving the inhibitors in DMSO at a concentration of 40 mg/ml, and diluting them 10-fold in the crystallisation buffer. For co-crystallisation experiments, 0.3 μ l 40% (v/v) acetonitril and 0.3 μ l inhibitor solution were added to the drops, giving raise to a final inhibitor concentration of \sim 10 mM in the droplet. Additionally, one crystal of the inhibited protein was soaked overnight in a solution buffered to pH 3 (B-IPF pH 3) containing 0.75 M ammonium sulfate, 0.1 M sodium acetate and 25 mM glycine, before freezing and collecting a complete data set at 1.9 \AA . Crystals grew within one week at 21 $^{\circ}$ C to maximum dimensions of 0.2 x 0.05 x 0.05 mm³.

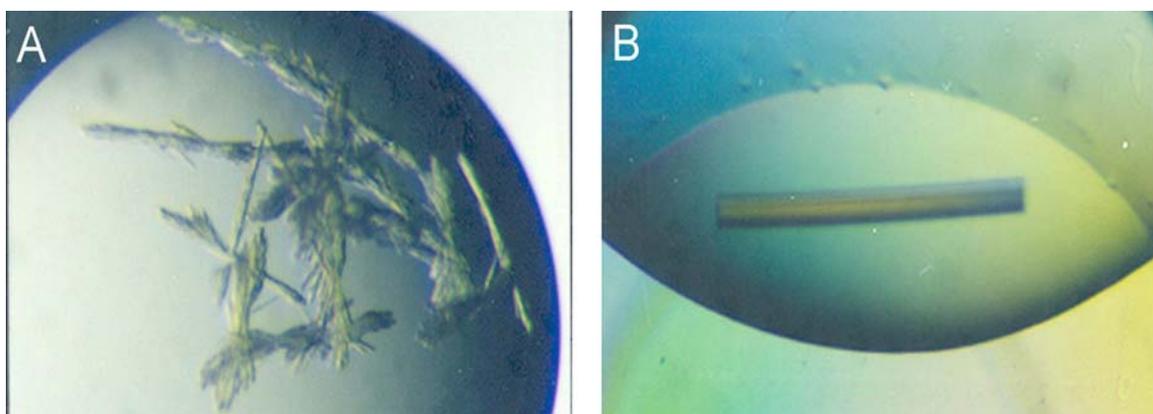


Figure 4.2: Crystals of kumamolisin from *Bacillus novosp.* MN-32. **A)** “Needles” in the initial crystallization condition (0.1 M sodium acetate pH 4.6, 2.0 M ammonium sulphate). **B)** Long single crystal of the P2₁ space group with 2 molecules per asymmetric unit (crystal form B, see Table 1).

4.1.2 Data Collection and Refinement

Addition of 30% (v/v) glycerol to the respective crystallization buffers allowed shock freezing of the crystals. Crystals were soaked for 10 s in the cryobuffer before freezing them under a nitrogen stream at 100 K (Oxford Cryosystems Cryostream). High-resolution data to 1.4 \AA resolution were collected for native kumamolisin (crystal form B) at the DESY BW6 beamline (Deutsches Elektronen Synchrotron, Hamburg, Germany) at a 1.050 \AA wavelength (Figures 4.3.A and B). From crystals A (2.3 \AA), B-IAF (1.8 \AA) and

B-IPF (2.8 Å), diffraction data were collected with our in-house MAR Research image-plate system (Hamburg, Germany) mounted on a Rigaku rotating-anode X-ray generator (Tokyo, Japan).

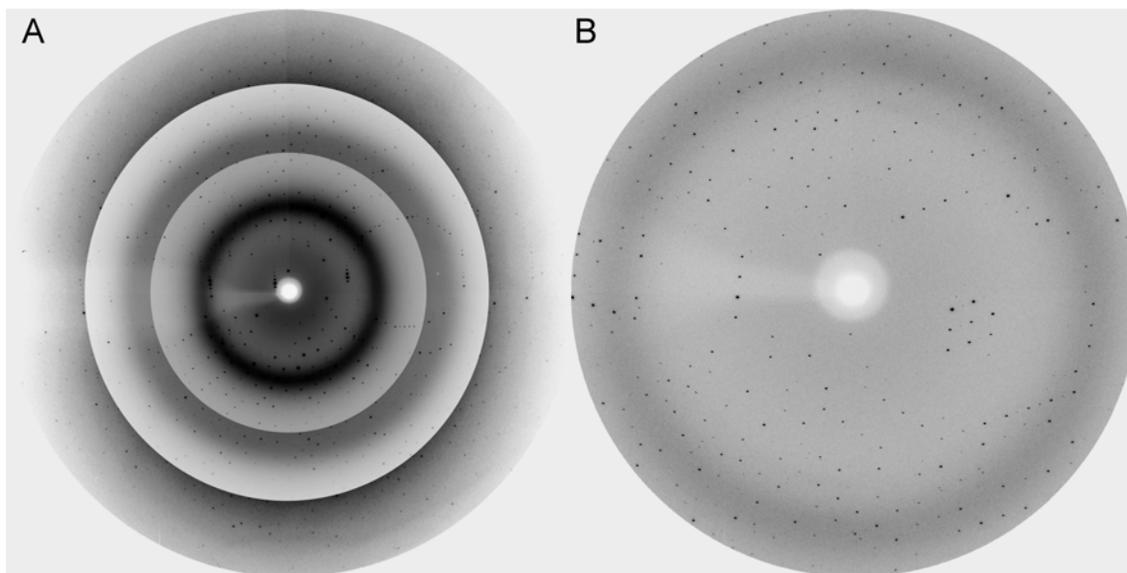


Figure 4.3: Diffraction images of kumamolisin crystal form B. **A)** High-resolution data set taken from 3 Å up to 1.38 Å (rotation 0.5°, mosaicity 0.18°). Contrast and brightness were separately adjusted in the outer shells in order to visualize the higher resolution spots. **B)** Low-resolution data set of the same crystal up to 3 Å (rotation 2°, mosaicity 0.14°). Both data sets were merged before integration.

Indexing and integration of the diffraction data was carried out with DENZO (Otwinowski and Minor, 1993). The data were merged, scaled and truncated using programs supported by the Collaborative Computational Project No. 4 (The CCP4 suite, 1994). Taking into account the systematic extinctions found in $0k0$ reflections (being $k=2n$), which indicated the presence of a two fold screw axis, the data was processed in the space group $P2_1$.

While searching for heavy atom derivatives of kumamolisin crystals, the three-dimensional structure of the homologous protein sedolisin was solved by A. Wlodawer and co-workers (Wlodawer et al., 2001). The structure of crystal form A (corresponding to the native kumamolisin monomer) was therefore solved using data collected in-house with the program suite AMoRe (Navaza, 1994) and a modified sedolisin search model

Results

(Wlodawer et al., 2001). The search model consisted of the residues conserved in both enzymes, while all other residues were truncated to alanines. After rigid body fitting, the best solution had a correlation coefficient and R-factor of 23.1% and 49.5% for 15.0 to 3.5 Å data. The refined structure of crystal A was used as the search model to solve all crystals of form B, with two kumamolisin molecules per asymmetric unit, yielding a correlation coefficient and an R-factor of 83.8% and 33.5%, respectively, for the same resolution range.

Table 1: Statistics for Data Collection and Refinement

Crystal	A	B	B-IPF	B-IAF	B-IPF pH 3
Content	Native kumamolisin monomeric	Native kumamolisin dimeric	kumamolisin Ac-IPF-CHO dimeric	kumamolisin Ac-IAF-CHO dimeric	kumamolisin Ac-IPF-CHO dimeric
Data Collection					
Space group	P2 ₁	P2 ₁	P2 ₁	P2 ₁	P2 ₁
Cell Constants (Å)					
a	42.63	54.86	55.03	54.76	54.96
b	78.32	78.25	78.30	78.18	78.23
c	49.00	73.53	73.17	72.98	73.36
β	106.33	98.28	98.45	98.09	98.13
Resolution (Å)	2.3	1.4	2.8	1.8	1.9
Outermost shell (Å)	2.39-2.27	1.41-1.38	2.95-2.80	1.84-1.78	1.93-1.90
Reflections measured	26014	603967	26276	114754	102389
Unique reflections	13692	120406	13434	55533	45324
^a R _{merge} (%)	9.1	6.9	9.9	4.1	8.6
Completeness overall/ outermost shell (%)	95.6/87.4	95.5/79.0	89.6/89.6	97.3/90.6	93.8/93.8
Refinement					
Reflections used for refinement	12659	108310	12871	53281	43023
Resolution range (Å)	36.3-2.27	19.52-1.38	12.00-2.80	24.87-1.78	28.28-1.90
^b R _{factor} (%) / ^c R _{free} (%)	20.6/25.8	19.4/20.8	21.4/28.4	19.5/23.1	19.8/24.1
R.m.s.d.					
Bonds (Å) / Angles (°)	0.010/1.65	0.011/1.63	0.010/1.58	0.010/1.59	0.009/1.55
Average B (Å ²)					
Protein+Ca ²⁺ / Inhibitor	16.1/-	23.8/-	15.7/13.3	10.3/16.0	15.4/16.9

$$^a R_{\text{merge}} = \frac{\sum_{\text{hkl}} \langle |I| \rangle - I}{\sum_{\text{hkl}} I}$$

$$^b R_{\text{factor}} = \frac{\sum_{\text{hkl}} |F_{\text{obs}}| - |F_{\text{calc}}|}{\sum_{\text{hkl}} |F_{\text{obs}}|}$$

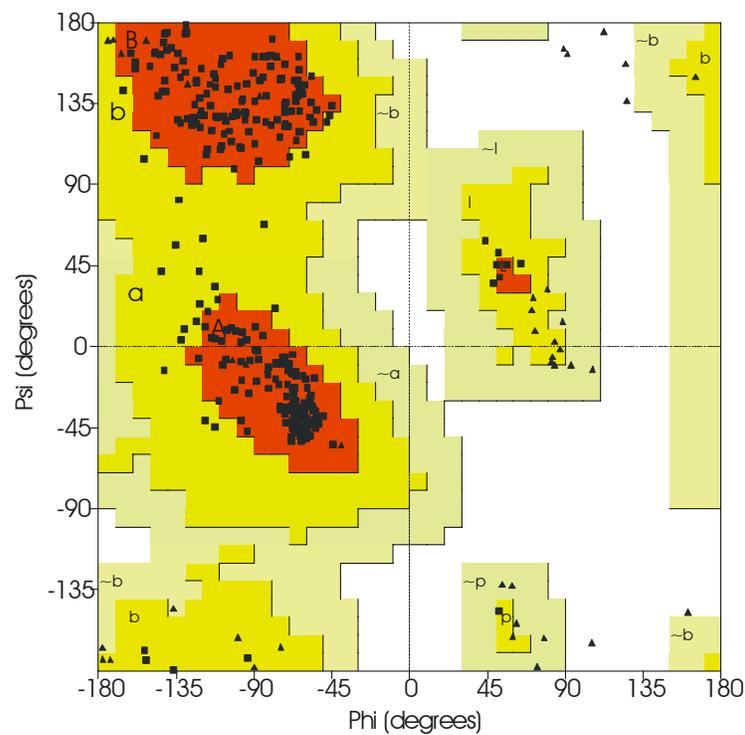
^cR_{free} is the R-value calculated with 500 reflections not used in refinement.

Model building was performed on an SGI graphics workstation using the program MAIN (Turk, 1992). Positional and individual temperature factor refinement was carried out with the program CNS (Brunger et al., 1998) until convergence with and without imposing non-crystallographic symmetry constraints for the dimeric and the monomeric forms, respectively, yielding a final R-factor of 19.4% for the native 1.4 Å structure (Table 1).

The crystal structure analysis of purified kumamolisin showed an inconsistency of the electron density with the last C-terminal protein residues of the deposited sequence. Subsequent analysis proved that deletion of the third base of the Phe351 codon in the expression plasmid had led to a single base frame-shift, the concomitant change of the following residues (from 351-FASGAAAERFTGPA-364 to 351-LLQALLPSASQAQP-364), and the premature termination at P364 leading to the loss of the 20 C-terminal amino acids. The truncated kumamolisin, whose structure is described below, exhibits an almost 2.5 fold higher activity compared with the wild-type enzyme, and seems to be equally stable. The defined residues Leu351 to Leu356 of this construct are expected to have a similar structure as those of the wild-type enzyme, while the last eight residues are disordered.

In the dimeric structures, the polypeptide chains for two independent molecules was defined by electron density from the N-terminal Ala1 to Leu356, except for the main chain segment 241-245 and a few side chains. In crystals B-IAF and B-IPF an unambiguous electron density indicated the presence of the expected inhibitors. In all structures one Ca^{2+} ion coordinated by six ligands was identified and refined with full occupancy. Additional electron density near Gly200 was interpreted as a sulfate ion in all structures. During the final steps of refinement, two alternative conformations were identified and built for a few side chains of structures B and B-IAF. Plot of the main chain angles (Figure 4.4), calculated with PROCHECK (Laskowski et al., 1993), show that all residues in all structures fall into the most favoured or additionally favoured regions of the Ramachandran plot (Ramachandran and Sasisekharan, 1968).

Figure 4.4: Ramachandran plot of native kumamolisin (monomeric crystal form). The diagram represents ϕ angle (axis of abscisses) *versus* ψ angle (axis of ordinates) for each residue. The color code is red for most favored regions, yellow for additional allowed regions, pale yellow for generously allowed regions and white for disallowed regions. As glycine (\blacktriangle) lacks any side chain, this residue is also allowed in white regions. For other residues (\blacksquare), the statistics show 88.8% located in most favored regions and 11.2% in additional allowed regions.



4.1.3 Description of the Structure

4.1.3.1 Overall Structure of Kumamolisin

Two crystal forms of kumamolisin, which belong to the monoclinic space group $P2_1$, were obtained in two different crystallization conditions. Crystals of form A obtained with the free enzyme contain one molecule per asymmetric unit (Table 1). Crystal form B, with two molecules in the asymmetric unit, was obtained from the native protein as well as from complexes with two different inhibitors, *N*-acetyl-isoleucyl-alanyl-phenylalaninal (Ac-IAF-CHO) and *N*-acetyl-isoleucyl-prolyl-phenylalaninal (Ac-IPF-CHO), referred to as B-IAF and B-IPF, respectively (Table 1). In both the monomeric and the dimeric crystal forms, the kumamolisin molecules are tightly packed, consistent with a 35% (v/v) solvent content. In line with the similar packings, the structures do not differ significantly, reflected by the low r.m.s. deviations of 0.17 Å and 0.23 Å for all 357 α -carbon atoms between the two independent molecules within the dimer and between the monomeric and dimeric forms, respectively. Thus, the following description of the overall fold of kumamolisin is virtually independent of both crystal packing and complexation state and will be mainly based on the 1.4 Å native dimeric structure.

Kumamolisin is a single domain protein with overall dimensions 48 x 44 x 31 Å³ (Figures 4.5.A and B). The globular structure exhibits hollow at the front in the *standard orientation* (with the substrate or peptide inhibitor running from the N-, left, to the C-terminal, right, as shown in Figure 4.5.A), which accommodates the substrate binding region and the active site (see section 4.1.3.3). The core of the molecule consists of a highly twisted β -sheet composed of eight parallel β -strands arranged in the sequential order s2, s3, s1, s4, s5, s6, s7 and s12 (Figure 4.5.A). This β -sheet is flanked by five helices arranged on both sides running parallel to each other and antiparallel to the core strands. Four of these helices (h2a, h4, h5, h8) and helix h1 are mostly exposed at the molecular surface, while the remaining helices h3 and h6 are buried inside the hydrophobic core of the molecule. In addition, kumamolisin exhibits a β -hairpin made by the antiparallel strands s6-s6a. Helix h7 and two other solvent exposed antiparallel pairs of strands, situated to the right of the active site (s6b-s6c and s8-s9) complete the overall fold of the enzyme.

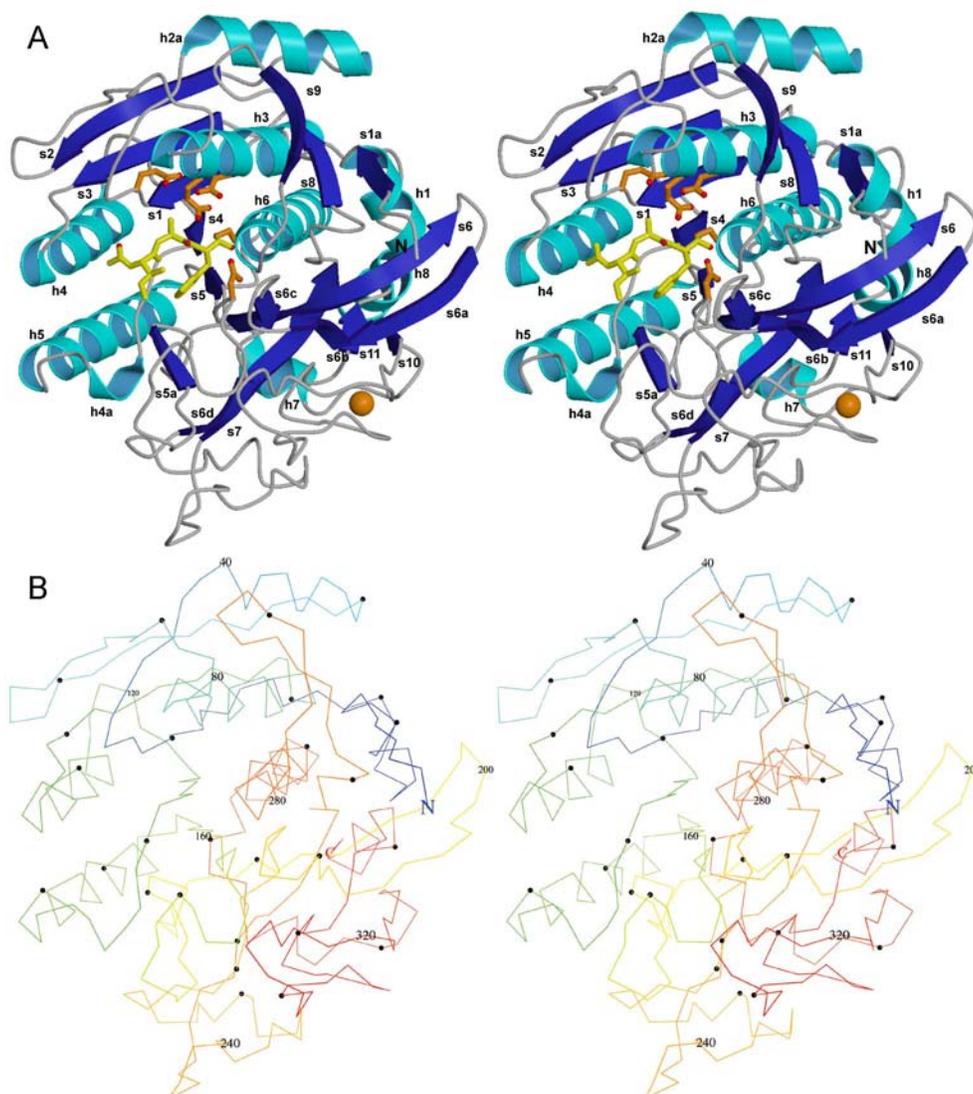


Figure 4.5: Crystal structure of the catalytic domain of kumamolisin. **A)** Stereo ribbon plot of kumamolisin in the standard orientation, with the covalently bound Ac-IAF-CHO inhibitor shown as a yellow stick model. The catalytic residues and the Ca^{2+} ion are shown as orange stick models and an orange sphere, respectively. **B)** Stereo view of the Ca trace of kumamolisin shown with identical orientation as used for the ribbon plot. The protein chain is colored by rainbow colors from the N- (blue) to the C-terminal (red). The main chain trace is marked at every 10th residue and numbered every 40.

One Ca^{2+} binding site, created by the long open loop connecting strands s10 and s11 (toward the bottom right, Figure 4.5.A), was identified during refinement. The calcium ion is bound by the loop segment Gly334-Asp338 (Figure 4.6) on one side and segments Asp316-Ile317 and Thr341-Gly342 on the other. It is coordinated in an almost ideal

octahedral manner by one carboxylate oxygen from each Asp316 and Asp338, the main chain carbonyl groups of Ile317, Gly334, and Gly336, and an internal water molecule (Figure 4.6). The B-factors of the calcium ion and the coordinating oxygen atoms are below the average indicating tight binding and full occupancy even at pH 3 (Table 1).

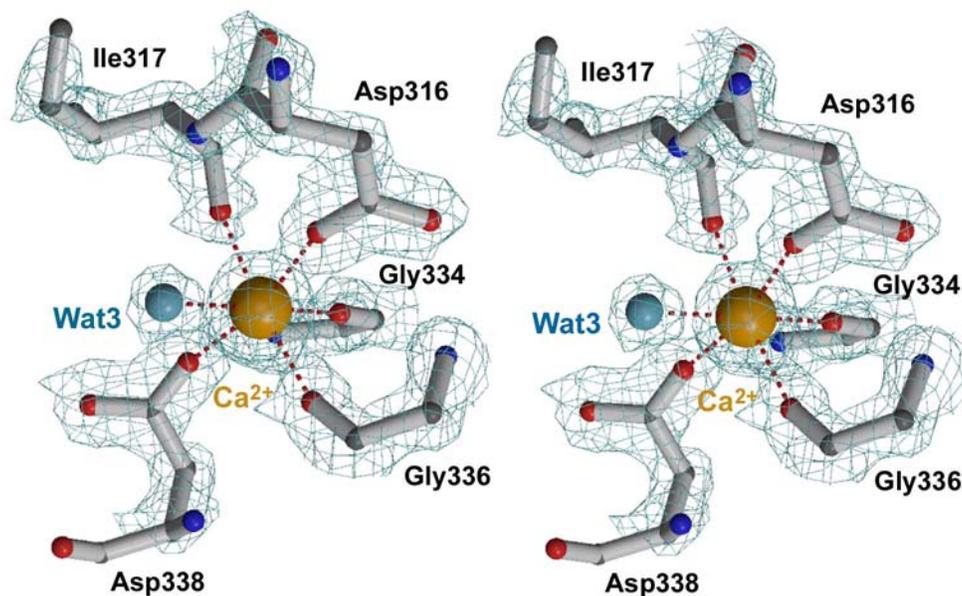


Figure 4.6: Stereo view on the calcium site, superimposed with the final electron density map. Ball-and-stick models are shown for the ligand-providing residues (grey), the bound Ca^{2+} ion (orange) and the liganding water molecule (blue). The final 2Fo-Fc electron density is contoured at 1.8σ . The ion is octahedrally coordinated by one carboxylate oxygen of Asp316 (2.35 Å) and Asp338 (2.35 Å) each, the three main chain carbonyl groups of Ile317 (2.39 Å), Gly334 (2.36 Å) and Gly336 (2.33 Å), and one water molecule (2.42 Å).

Residues Pro181, Pro251 and Tyr331 exhibit a *cis* conformation. The internal *cis*-Pro181 provides a change in the main chain direction of the loop connecting strands s5 and s6, while *cis*-Pro251 introduces a displacement between sub-strands s7a and s7b. Both the amino nitrogen and the carbonyl oxygen of the surface located *cis*-Tyr331 are directed inwards. This allows formation of two adjacent hydrogen bonds with two internal polar groups, namely the carboxamide oxygen of the conserved Asn322 and the

kumamolisin-specific Ser167 O γ , which thus stabilize the *cis* conformation. The unusual *cis*-Tyr331 favors solvent exposure of the polar side chains of residues Arg327 and Arg329.

A previous biochemical analysis had suggested the absence of free thiol groups in kumamolisin (Muraio et al., 1993), implicating that the four cysteine residues might form two disulphide bridges. It was therefore surprising that no disulphide bridges are found in the crystallized kumamolisin construct, which contains only three of the four cysteines present in wild-type proteinase due to lack of the last 28 amino acid residues (see section 4.1.2). Of these cysteines, Cys190 and Cys340 are deeply buried in the hydrophobic core, with their S γ atoms being 15 Å apart from each other, thus precluding any disulfide bridge formation. Both cysteine residues might not be accessible to thiol-modifying compounds such as Ellman's reagent, thus explaining previous failures to detect free thiol groups. The third cysteine, Cys27, is located close to the molecular surface, with its S γ atom essentially buried in an internal cavity. Formation of a disulphide bridge with the missing Cys377 would require (besides the proper bridging of the corresponding C-terminal segment) a side chain rotation of Cys27 and some rearrangements of surrounding residues, making disulphide formation with Cys377 unlikely.

4.1.3.2 Structural Comparison with Sedolisin and Subtilisin

The optimal superposition of kumamolisin with the previously solved sedolisin (Wlodawer et al., 2001) (Figure 4.7.A) reveals that 247 of the 357 α -carbon atoms are topologically equivalent, with an r.m.s deviation of 0.79 Å (using a threshold of 1.5 Å). Accordingly, the overall fold of kumamolisin resembles that of sedolisin, with most secondary structure elements and the Ca²⁺ binding site being conserved (Figure 4.7.B). Major differences include helix h2 of sedolisin, which is missing in kumamolisin, strand s1a that is only present in the latter enzyme, and several β -hairpin structures. The calcium site and the coordination geometry in kumamolisin resemble those in sedolisin. The *cis*-Pro182 and 251 residues of kumamolisin have equivalents in sedolisin, embedded in similar environments. The sedolisin tyrosine equivalent to the *cis*-Tyr331 of kumamolisin, exhibits the usual *trans* conformation, presumably because an internal polar residue equivalent to kumamolisin Ser167 is not present in sedolisin. The four

cysteine residues of kumamolisin have no equivalent in the other serine carboxyl proteinases of known sequence (Figure 4.7.B).

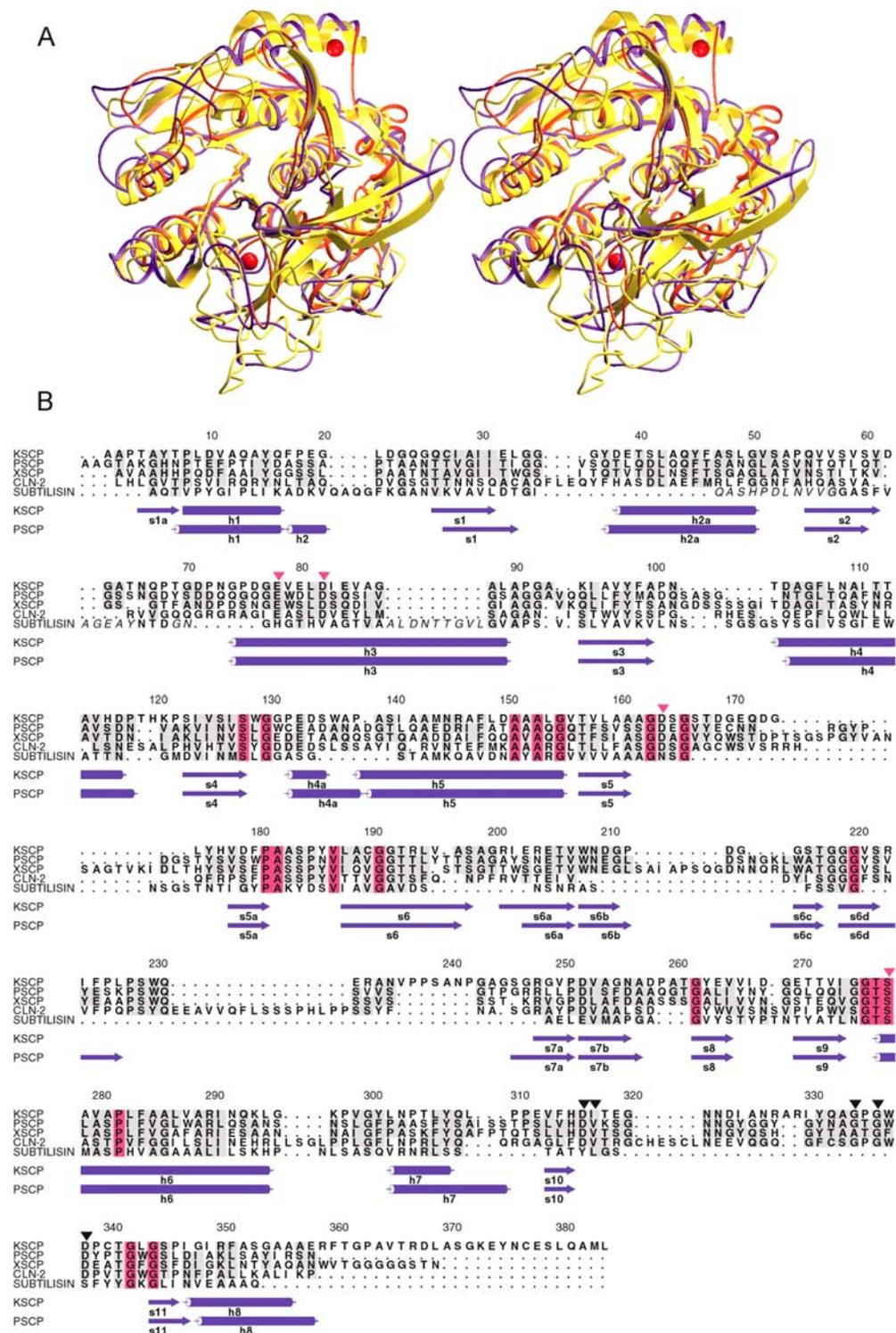


Figure 4.7: (Previous page) Topological and sequence comparison of kumamolisin with other members of the SCP family and subtilisin Carlsberg. **A)** Stereo ribbon plot of kumamolisin (gold) superimposed with sedolisin (Wlodawer et al., 2001) (purple) and subtilisin Carlsberg (Bode et al., 1987) (red). The view is in standard orientation. **B)** Structure-based alignment of the amino acid sequences of kumamolisin (KSCP), sedolisin (PSCP) (Wlodawer et al., 2001), XSCP (Oda et al., 1996), CLN2 (Ezaki et al., 1999) and subtilisin Carlsberg (Bode et al., 1987). The identities of these enzymes to KSCP based on topologically equivalence are: PSCP, 27.6%; XSCP, 23.7%; CLN2, 26.6%; and subtilisin, 13.3%. The grey shading indicates structurally equivalent residues that are conservatively substituted in at least three of the five proteins. Residues conserved in all five sequences are highlighted in red. Numbers refer to the kumamolisin sequence; β -sheets and α -helices found in kumamolisin and sedolisin are represented by arrows and cylinders, respectively. The catalytic residues are indicated with red arrows, and the Ca^{2+} ligands for kumamolisin and sedolisin are marked with black arrows. Parts of the subtilisin sequence not topologically equivalent to kumamolisin are shown in italics.

The topological similarity of kumamolisin with subtilisin is significantly lower, but the overall fold is conserved between the two enzymes (Figure 4.7.A). A structural comparison of kumamolisin with subtilisin Carlsberg (Bode et al., 1987) shows 157 topologically equivalent $\text{C}\alpha$ atoms, with a 0.92 Å r.m.s. deviation (threshold 1.5 Å). Both proteinases share about 13% structurally identical residues, a number of which are also conserved in the other SCPs (highlighted in red in Figure 4.7.B). The longer peptide chain of kumamolisin is accommodated in additional secondary structure elements. Interestingly, the number and positions of the calcium ion sites are not conserved between the two enzymes (Figure 4.7.A). The kumamolisin Ca^{2+} site is situated in a peptide insertion with no counterpart in subtilisin, while the molecular region of kumamolisin corresponding to the first subtilisin Ca^{2+} site differs considerably from subtilisin. In contrast, the second subtilisin calcium binding loop is topologically conserved in kumamolisin; however, the Ca^{2+} position is occupied by the guanidinium group of Arg248. This guanidinium group forms hydrogen bond interactions with surrounding oxygen atoms that are equivalent to the interactions of the calcium ion in subtilisin.

Compared with other proteinases of the subtilisin superfamily, the proportion of charged residues (15%) is not unusually high in kumamolisin. At neutral pH, its 23 aspartate and 15 glutamate residues are not charge-compensated by the 11 arginine, 4

lysine and 4 histidine residues giving rise to a strong negative overall charge (Figures 4.8.A and B). Though at pH 3.0, which is the one of maximum proteolytic activity due to the *Bacillus novosp.* MN-32 natural habitat, the net overall charge will be slightly positive. Most of the aspartate and glutamate residues cluster to the active-site cleft and its environment (Figure 4.8.A), while only a few polar residues are located on the opposite molecular surface (Figure 4.8.B). It is remarkable that several acidic residues are arranged as salt bridges (surface located: Asp179/Asp169, Asp134/Glu171, Asp210/Asp258; partially buried: Asp38/Glu80, Asp76/Asp268, and the Glu78/Asp82/Glu32 triplet), connected through quite short O···H···O inter-carboxyl hydrogen bonds of about 2.6 Å. Three more isolated acidic residues (Glu84, Glu205, and Asp338) are completely buried, with their carboxylate groups engaged in internal

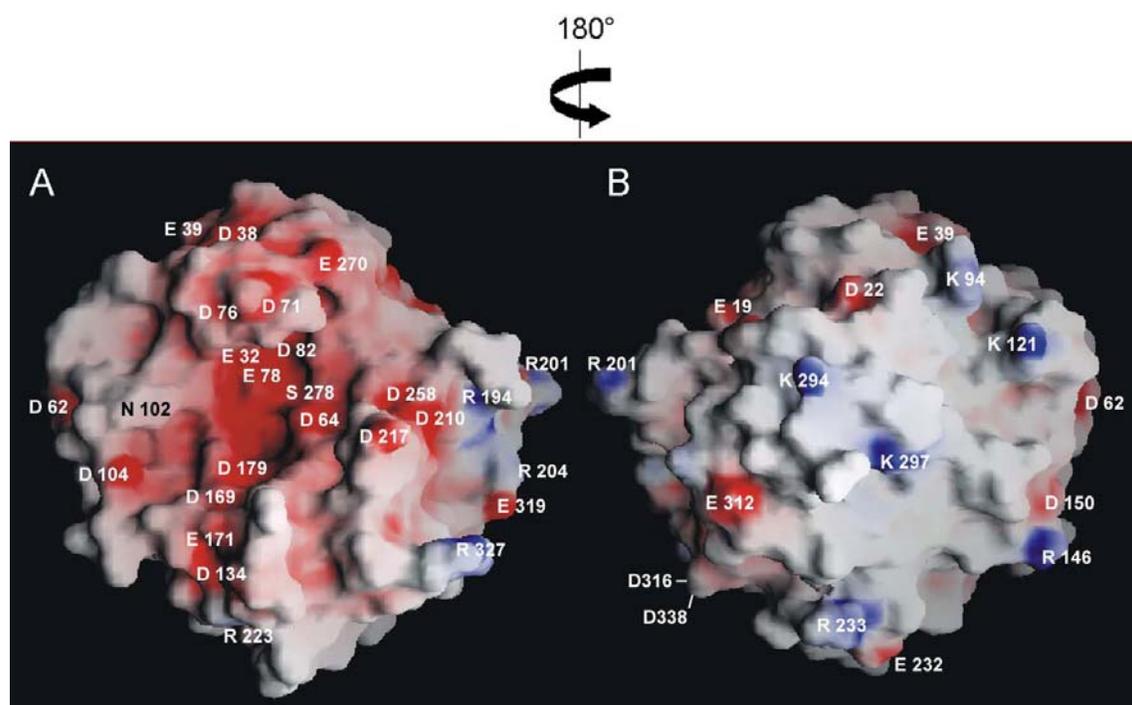


Figure 4.8: Electrostatic surface representation of kumamolisin. **A)** Solid surface representation of the “front” side of kumamolisin shown in the same orientation as in Figures 4.5.A and B. **B)** Solid surface representation of the “back” side, obtained after a rotation of 180° around a vertical axis. White lines indicate the approximate position of the Ca²⁺ binding residues Asp316 and Asp338. In **A)** and **B)** the colors indicate the electrostatic surface potential at pH 7 contoured from +20 kT/e (dark blue) to -20 kT/e (dark red).

hydrogen bonds, and the buried Asp252 forms an internal salt bridge with Arg248 (which mimics the bound calcium in subtilisin, see above).

4.1.3.3 Active Site and Substrate Binding Site

To define the binding geometry of a productively bound peptide substrate and to explore its interactions with kumamolisin subsites, two structures of the enzyme were solved in complexes with two different aldehyde inhibitors, which were designed based on the observed cleavage preference of the enzyme (Muraio et al., 1993; Oda et al., 2000). Both inhibitors, Ac-IAF-CHO and Ac-IPF-CHO, bind covalently through their aldehyde group to the O γ atom of Ser278. The hemiacetal group formed is clearly defined in the electron density, which unambiguously identifies its central carbon atom to exhibit *S*-chirality. This covalent binding confirms Ser278 as the primary catalytic nucleophile in the active site of kumamolisin. The active-site cleft of the enzyme is solvent-accessible, and forms a slight depression, which runs from left to right over the active site (Figures 4.5.A and 4.8.A).

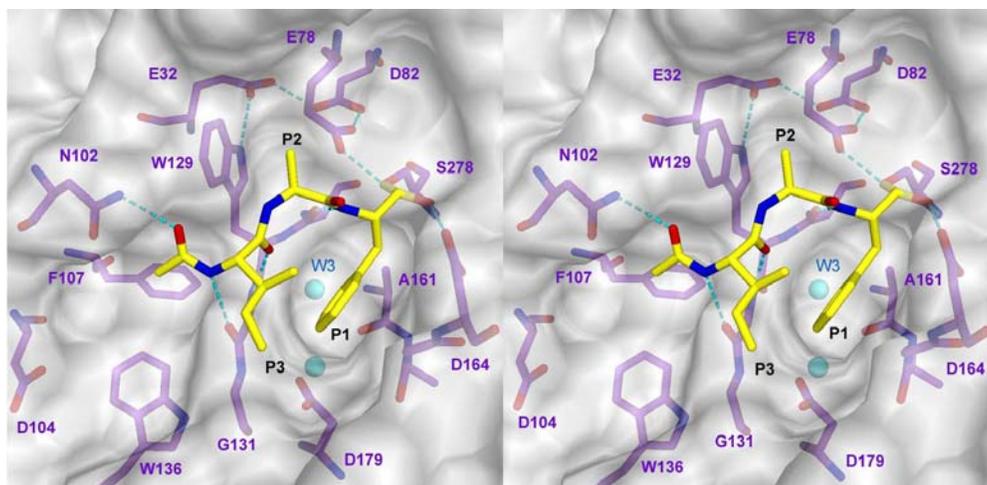


Figure 4.9: Stereo plot of the active site shown in standard orientation. Kumamolisin (represented as a purple stick model) is superimposed with a transparent Connolly surface. The Ac-IAF-CHO inhibitor and the water molecules are displayed as a yellow stick model and as cyan spheres, respectively. The dashed cyan lines indicate hydrogen bonds between the active site residues and between the enzyme and the inhibitor.

In the free enzyme, a network of quite short hydrogen bonds connects Ser278 O γ with Asp82, via the carboxylate group of Glu78 (Figure 4.9). In kumamolisin, this hydrogen bond system extends beyond Asp82 to the Glu32 carboxylate oxygens and Trp129 N ϵ completing a semicircular network around the S2 pocket. Additionally, Asp82 hydrogen bonds to an internal water molecule (2.97 Å), which is at its turn hydrogen bonded to Ile31 O, Ser128 N, and to a second localized solvent molecule. Close to the catalytic Ser278 another acidic residue, Asp164, is placed on the molecular surface of kumamolisin. The Asp164 O δ 1 atom together with the Ser278 amide nitrogen form the oxyanion hole, which in the inhibited structures accommodates the hemiacetal oxygen of the bound inhibitor (Figure 4.9).

The peptidyl moieties of the two bound inhibitors run antiparallel to kumamolisin segment Ser128-Pro132 forming a twisted two-strand antiparallel β -sheet with two main chain hydrogen bonds between Gly130 and P3-Ile, and one between Ser128 O and P1-Phe N, respectively (Figure 4.9). In addition, the carbonyl oxygen of the N-terminal acetyl group of the inhibitor accepts a hydrogen bond from the nitrogen of the Asn102 carboxamide group. The side chain of P1-Phe slots into the S1 groove mainly bordered by main chain segments 130-131 and 161-164, and delimited by the Asp179 side chain (Figure 4.9). The tight interaction (2.7 Å) between the partially positively charged phenyl ring edge and the partial negative charge of Asp179 carboxylate might further facilitate accommodation of a phenylalanine (Figure 4.9). The two water molecules placed at the bottom of the S1 subsite seem to be integral structural elements of kumamolisin, since they are also present in the uninhibited structure. The P2 residues of the inhibitors, proline in Ac-IPF-CHO and alanine in Ac-IAF-CHO, fit well into the small and hydrophobic S2 enzyme depression, which is demarcated on one side by the aliphatic part of the catalytic Glu78 side chain, and on the other side by the indole moiety of Trp129. This bulky tryptophan side chain narrows the S2 subsite of the enzyme, allowing only the accommodation of small hydrophobic groups. Finally, the isobutyl side chain of P3-Ile extends into the active-site cleft, without being fitted in a defined S3 pocket.

The probable location of subsites S1'-S3', S4 and S5 and their interactions with a bound substrate have been studied by modeling a substrate octapeptide to kumamolisin. The model was based on the crystal structure of eglin in its complex with subtilisin

Carlsberg (Bode et al., 1987). The sequence of the eglin reactive-site loop was replaced by Lys-Pro-Ile-Pro-Phe↓Tyr-Arg-Leu (Figure 4.10), which was found to be the most effectively hydrolyzed substrate out of 74 tested (Oda et al., 2000). The central part of this peptide substrate (P1-P3) fits the conformation observed in kumamolisin structure with the Ac-IPF-CHO inhibitor. The P4-P3 peptide bond was modeled according to the N-acetyl group interactions of the inhibitor, with the P4 carbonyl hydrogen bonding to the Asn102 carboxamide nitrogen (Figures 4.9 and 4.10).

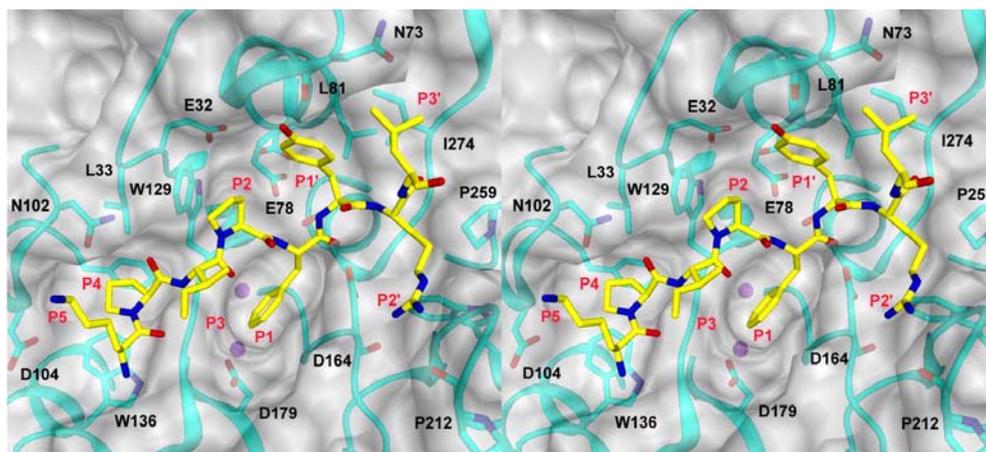


Figure 4.10: Stereo image of the active-site cleft of kumamolisin. A Lys-Pro-Ile-Pro-Phe↓Tyr-Arg-Leu peptide substrate (yellow stick model) is modeled as probable binding mode to kumamolisin. The proteinase, shown as a cyan ribbon with a few important subsite residues highlighted as stick models, is superimposed with a grey half-transparent Connolly surface.

According to the model, the hydrophobic S4 groove is shaped by the main chain segment 102-104 and the side chains of residues Asn102, Leu33 and Phe107 (Figure 4.10). The P5-Lys side chain has been arbitrarily modeled to extend away from the molecular surface. To the right of the active Ser278, the active-site cleft expands along segment 274-277, allowing hydrogen bond formation between main chain atoms of P2' and Gly275. The phenolic side chain of P1'-Tyr nestles into the large hydrophobic S1' groove flanked by the aliphatic side chain parts of Glu78 and Leu81 and the main chain of segment 74-78. This disposition allows the P1'-Tyr side chain O ϵ atom to hydrogen bond to the Asn73 carbonyl group. The P2'-Arg side chain runs along a cleft bounded by

Asn73, Pro259 and Gly214 favoring hydrogen bond interactions between its guanidinium group and the adjacent carbonyl groups. The P3'-Leu side chain is directed into a hydrophobic surface pocket, which is large enough to accommodate more bulky side chains.

4.1.3.4 Site-Directed Mutagenesis Studies

In parallel to the structure elucidation of kumamolisin, site-directed mutagenesis studies were performed on five selected residues, Glu78, Asp82, Asp164, Ser278, and Asp316, which are conserved throughout the SCP family. The 64kDa proenzymes were produced in *E. coli* JM109 cells harboring the wild-type plasmid (pS3-A1) and mutant plasmids (pE78A, pD82A, pD164A, pS278A, and pD316A), respectively, and are all immunoreactive against anti-kumamolisin antibodies (Figure 4.11). The D316A precursor was, as the recombinant wild-type proenzyme, converted to the 43-kDa active form by incubation under acidic conditions, yielding a proteinase with about 3% activity of wild-type kumamolisin. The E78A, D82A, D164A and S278A proforms, in contrast, did not auto-activate under these conditions, and did not exhibit any measurable proteolytic activity upon activation cleavage with wild-type enzyme.

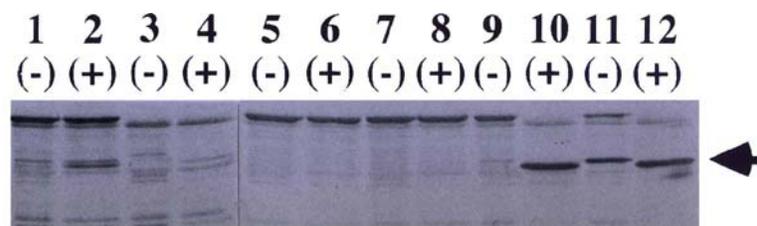


Figure 4.11: Western blot analysis of the expressed kumamolisin mutants in transformed *E. coli* cells. The different bands correspond to *E. coli* cell extracts before (-) and after (+) acidic activation, electrophoresed in a 12.5% SDS-polyacrylamide gel and analyzed by Western blotting using rabbit anti-kumamolisin antiserum. The arrowhead indicates the wild-type enzyme. The Western blot shows the cell-free extracts of cells harboring the following plasmids: pE78A (lanes 1 and 2), pD82A (lanes 3 and 4), pD164A (lanes 5 and 6), pS278A (lanes 7 and 8), pD316A (lanes 9 and 10) and pS3-A1 wild-type (lanes 11 and 12).

4.2 Pro-Kumamolisin and Mutants E32A and W129A

4.2.1 Protein Crystallization

Samples suitable for crystallization trials were provided by Dr. K. Oda and coworkers (Department of Applied Biology, Faculty of Textile Science, Kyoto, Japan). The purified recombinant mutants E32A (catalytic domain only), W129A (in complex with the prodomain as a separate entity) and S278A pro-kumamolisin (Oyama et al, 2002) were crystallized using sitting-drop vapor-diffusion procedures. Mutant E32A crystallized isomorphously with native kumamolisin (space group $P2_1$; Table 1) with one molecule per asymmetric unit (Table 2). The same packing form was found for crystals consisting of the W129A catalytic domain alone, grown from the complex of the W129A catalytic domain and the propeptide. The S278A pro-kumamolisin crystals belong to space group R32 and contain one molecule per asymmetric unit (Table 2).

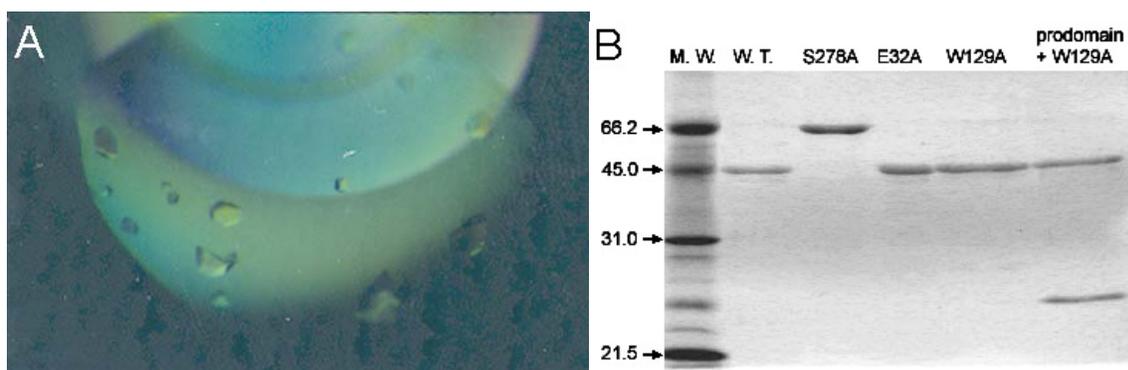


Figure 4.12: **A)** Crystals of pro-kumamolisin from *Bacillus novosp.* MN-32 grown at 21° C. Crystals appear in the condition n° 38 of Crystal Screen (Hampton Research, Laguna Hills, USA) and no optimization was required. **B)** Analysis of samples used in crystallization trials. Proteins were electrophoresed in a 12.5% SDS-polyacrylamide gel and were visualized by Coomassie staining. The Molecular Weight Marker (M.W.) indicates the position of proteins with apparent molecular weights from 66.2 to 21.5 kDa.

The $P2_1$ crystals were grown by mixing 1 μ l of a 9 mg/ml protein solution in 25 mM sodium chloride, 12.5 mM sodium acetate, pH 5.5, with 1 μ l reservoir solution, consisting of 0.4 M ammonium sulphate, 0.1 M sodium acetate (pH 4) and evaporation toward the reservoir. The R32 crystals grew from a mixture of 1 μ l protein solution (11.6

mg/ml enzyme in 25mM sodium chloride, 12.5 mM MOPS, pH 7.0) and 1 μ l of the reservoir buffer (1.4 M tri-sodium citrate dihydrate, 0.1 M HEPES-Na, pH 7.5). After one week at 21° C, the crystals reached a final size of $0.4 \times 0.2 \times 0.2$ mm³ in the case of full-length S278A kumamolisin and of $0.2 \times 0.05 \times 0.05$ mm³ for E32A and W129A.

4.2.2 Data Collection and Refinement

All crystals were soaked for 10 s in cryobuffers containing 25% (v/v) glycerol before freezing them in a nitrogen stream at 100 K (Oxford Cryosystems Cryostream). High-resolution data to beyond 1.2/1.3 Å resolution were collected for all kumamolisin mutants at the DESY BW6 beamline (Deutsches Elektronen Synchrotron, Hamburg, Germany) at a wavelength of 1.050 Å (Figure 4.13).

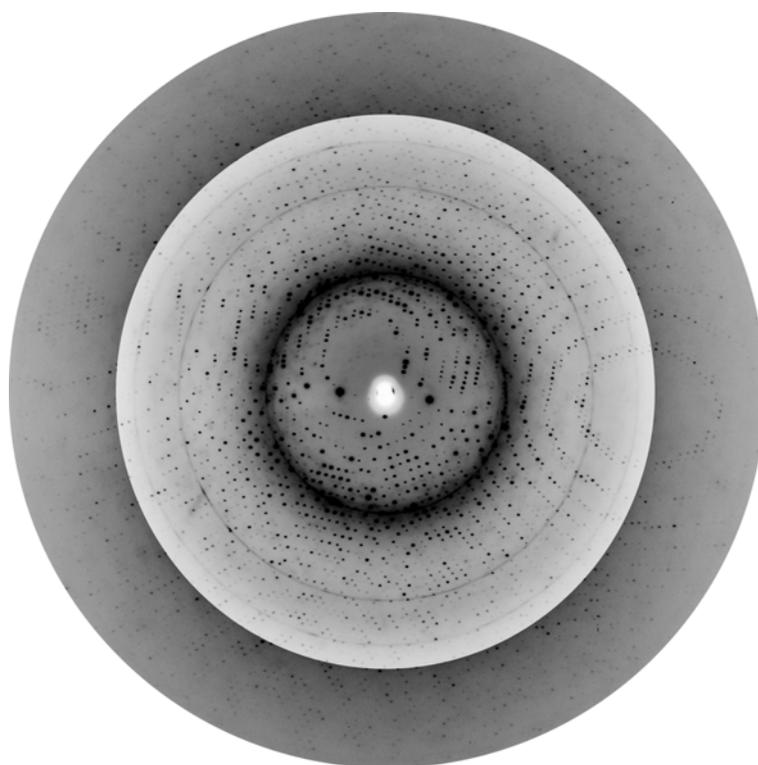


Figure 4.13: Diffraction image of pro-kumamolisin mutant S278A. The image was recorded on a Mar CCD detector on beamline BW6 at DESY (Hamburg) with rotation of 0.5° during exposure of the crystal (crystal mosaicity of 0.2°). The detector edge corresponds to a limiting resolution of 1.18 Å at a wavelength of 1.0500 Å. Contrast and brightness were separately adjusted in the outer shells in order to visualize the higher resolution spots.

Results

Indexing and integration of the diffraction data was carried out with MOSFLM6.11 (Leslie, 1991). The data were merged, scaled and truncated using programs supported by the Collaborative Computational Project No. 4 (The CCP4 suite, 1994). All structures were solved by molecular replacement using the program suite AMoRe (Navaza, 1994) and the native kumamolisin search model (Table 2). In the case of the pro-kumamolisin structure, the ambiguity between space groups R3 and R32 was resolved by comparing statistic values. For 15.0 to 3.5 Å data, correlation coefficient and R-factor values of 10.44% and 51.6%, respectively, for R3, and 36.7% and 46.9% for R32, established R32 as the correct space group. Thus, the first electron density map was calculated with the kumamolisin molecular replacement model in the space group R32, which displayed a nice density for the catalytic domain as well as for most of the prodomain. Model refinement was performed by iterative rounds of model building in MAIN (Turk, 1992) on an SGI graphics workstation and refinement using CNS (Brunger et al., 1998) until convergence. The final R-factors yielded were from 18.3 to 21.3% (Table 3).

Table 2: Statistics for Data Collection

Crystal	A	B	C
Content	Pro-kumamolisin mutant S278A monomeric	Kumamolisin mutant E32A monomeric	Kumamolisin mutant W129A monomeric
Space group	R32	P2 ₁	P2 ₁
Cell Constants			
a,b,c (Å)	158.54, 158.54, 105.72	42.72, 78.57, 49.99	42.89, 78.83, 49.03
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 105.48, 90.0	90.0, 105.41, 90.0
Resolution (Å)	1.18 (1.20-1.18)	1.18 (1.20-1.18)	1.28 (1.30-1.28)
Reflections measured	523625	207860	153826
I/σ (I)	24.0 (6.9)	13.1 (5.6)	9.5 (4.4)
Unique reflections	164763	98535	74340
^a R _{merge} (%)	5.3	7.7	5.4
Completeness (%)	99.3 (98.2)	97.0 (85.8)	92.7 (72.8)

$$^a R_{\text{merge}} = \frac{\sum_{hkl} |I| - I}{\sum_{hkl} I}$$

Data in brackets hold for the outermost shell.

Table 3: Refinement Statistics

Crystal	A	B	C
Content	Pro-kumamolisin mutant S278A	Kumamolisin mutant E32A	Kumamolisin mutant W129A
Reflections used for refinement	162816	98434	74287
Resolution range (Å)	99.00-1.18	99.00-1.18	99.00-1.28
^b R _{factor} (%) / ^c R _{free} (%)	21.3/24.4	20.2/22.2	18.3/20.7
R.m.s.d.			
Bonds (Å) / Angles (°)	0.010/1.73	0.011/1.42	0.009/1.49
B-factor r.m.s. deviation between bonded atoms (Å ²)	0.98	0.85	0.87
Averaged B-factors (Å ²)			
Protein + Ca ²⁺ / Water	19.17/32.31	14.24/28.00	13.10/28.42
Ramachandran core conformations (%)	91.1	88.8	89.5
Additionally allowed conformations (%)	8.9	11.2	10.5

$${}^bR_{\text{factor}} = \frac{\sum_{\text{hkl}} |F_{\text{obs}}| - |F_{\text{calc}}|}{\sum_{\text{hkl}} |F_{\text{obs}}|}$$

^cR_{free} is the R-value calculated with 500 reflections that were not used for refinement.

All mutant crystal structures show the same C-terminal truncation as the recombinant wild-type enzyme (see section 4.1.2) and the subsequent loss of the 20 C-terminal amino acids. In addition, sequencing of the crystallized E32A and W129A samples indicated heterogeneity N-terminal end. Analyzed samples started with Ala1e or Ser186p, with a clear predominance of the latter residue. Nevertheless, in the solved structures, the polypeptide chains are defined by electron density from the N-terminal Ala1e up to Pro357e, except for main chain segment 241e-245e and a few side chains. The pro-kumamolisin S278A molecule, which starts unambiguously with Ser2p, is fully defined by electron density from residue Glu12p up to Ala178p. Segment Glu179p-Ala184p is only partially defined, while the eight residues from Arg185p to Thr4e corresponding to the loop connecting the prodomain with the catalytic domain are fully

disordered. The rest of the catalytic domain shows adequate electron density up to Pro357e. In all structures, a single Ca^{2+} ion coordinated by six ligands is observed, and was refined with full occupancy. An additional electron density near Gly200 was interpreted as a sulfate ion in both catalytic domain structures, E32A and W129A. During the final steps of refinement, two alternative conformations were identified and built for several side chains in all three structures. An analysis of the main chain angles, performed using the program PROCHECK (Laskowski, 1993), show that all residues in all structures fall into the most favored or additionally favored regions of the Ramachandran plot (Ramachandran and Sasisekharan, 1968; Figure 4.14).

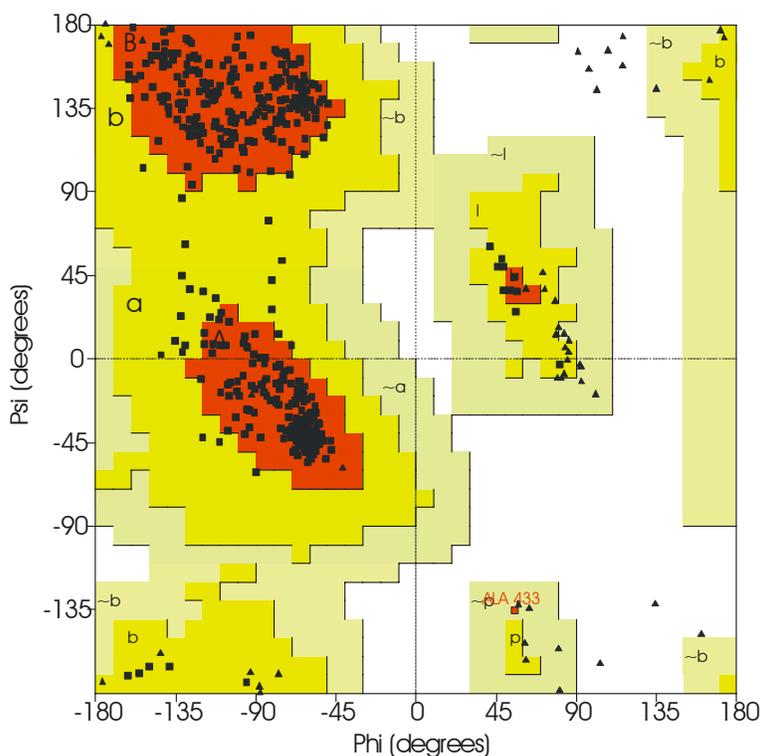


Figure 4.14: Ramachandran plot of pro-kumamolisin crystal structure. The figure plots ϕ angle (axis of abscisses) *versus* ψ angle (axis of ordinates) for each residue. All residues are situated in the most favored or additionally favored regions except for residue Ala234e that is located in a generously allowed region.

4.2.3 Description of the Structures

4.2.3.1 Structures of the E32A and W129A Mutants of Kumamolisin

Kumamolisin mutants E32A and W129A, with single amino acid replacements of Glu32e and Trp129e by alanine residues, consist of single-chain molecules with the same apparent molecular weight (Figure 4.12.B) and similar CD spectra as the mature enzyme (data not shown). However, their catalytic activity/specificity constant towards the peptide-like substrate Lys-Pro-Ile-Ala-Phe*Nph-Arg-Leu is reduced to 5.7 and 3.8% of the native enzyme, respectively (Table 4) (Oyama et al., unpublished data).

Table 4: Kinetics Parameters of Kumamolisin Wild-Type and Mutants

	k_{cat} (s^{-1})	K_{m} (μM)	$k_{\text{cat}}/K_{\text{m}}$ ($\mu\text{M}^{-1}\text{s}^{-1}$)	Relative $k_{\text{cat}}/K_{\text{m}}$ (%)
Wild-type	6.02	29	0.21000	100
S278A	0.08	84	0.00099	0.5
E32A	1.24	102	0.01200	5.7
W129A	0.67	89	0.00800	3.8

Substrate: Lys-Pro-Ile-Ala-Phe*Nph-Arg-Leu (*, cleavage site). Assay conditions: 60° C; 8.85, 17.9, 29.5, 43.1, and 61.4 μM substrate; 0.1 M sodium formate buffer (pH 3.5).

Both mutants crystallize isomorphously with the monomeric form A of native kumamolisin (see Table 1), characterized by a tightly packed proteinase molecules and consistent with the low (35% v/v) solvent content. The polypeptide chains are defined from the N-terminal Ala1e up to Pro357e, except for residues 127e-132e in the case of W129A, and are virtually identical with that of the native enzyme, with 361 α -carbon atoms showing r.m.s. deviations of 0.45 and 0.50 Å using a threshold of 1.5 Å. Like mature kumamolisin, E32A and W129A catalytic domains essentially consist of a central

eight-stranded parallel β -sheet, flanked by eight helices and a few short strand pairs arranged on both sides of the major sheet (Figure 4.5.A). Helices h3 and h6, carrying the active site residues Glu78e/Asp82e and Ser278e (see Figure 4.5.A), respectively, are mostly buried inside of the hydrophobic core. In contrast, the other helices, particularly h4 and h5 (see Figure 4.5.A), are basically exposed at the molecular surface. The single bound calcium ion is sandwiched between the multiple turn segments of a long open hairpin loop connecting strands s10 and s11.

As in the wild-type enzyme, the substrate binding region and the active site residues are located in a large cavity towards the front of the globular catalytic domain (showed in standard orientation in Figure 4.5.A). In the E32A mutant, the main chain is virtually unchanged compared with native kumamolisin, even in the catalytic center (Figure 4.15). The absence of the Glu32e side chain gives rise to an intra-molecular cavity occupied by a solvent molecule (Wat249), which hydrogen bonds to both the Asp82e carboxylic acid (O δ 1) and the Trp129e indole nitrogen (N ϵ 1). The latter indole moiety adopts two slightly different conformations, allowing only in one of them formation a favorable

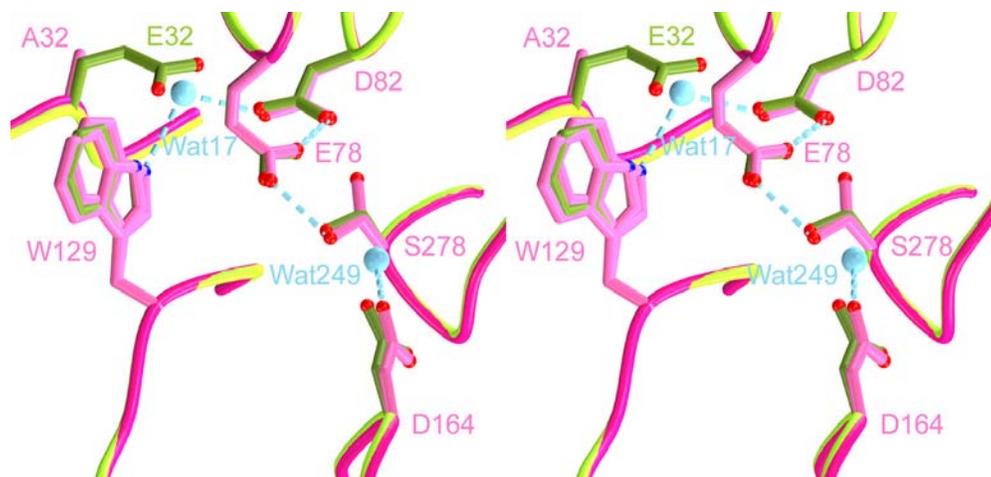


Figure 4.15: Stereo plot of a section around the active site of kumamolisin mutant E32A. Some loop segments (ropes) are shown together with a few side chains of particular importance (stick models). The wild-type enzyme colored in yellow and olive-green is superimposed with the mutant E32A, drawn in pink. Water molecules and possible hydrogen bonds (only for the mutant) are represented by cyan balls and dashed lines, respectively.

hydrogen bond to the fixed water molecule. The side chain of the catalytic Ser278e oscillates between two different positions, one in which the O γ atom can still hydrogen bond to the unchanged Glu78e carboxylic acid O ϵ 1, and a second one directed towards the interior of the catalytic domain i.e. away from Glu78e O ϵ 1. Therefore, the reduction in activity seems to be mainly caused by the lack of the Glu32e carboxylate group preventing formation of an optimal proton shuttle. The bound water molecule that replaces the absent carboxylate group may partially mimic its function, possibly explaining the remaining catalytic activity of mutant E78A (Table 4).

The active site region of the W129A mutant, which retains 3.8% of the original hydrolytic activity (Table 4), shows more drastic changes compared to native kumamolisin (Figure 4.16). Lacking of the bulky tryptophan side chain facilitates the main chain segment Ile127e-Gly131e, to deviate considerably from the wild-type conformation. In particular, the Ala129e side chain is directed away from the catalytic triad, while the hydrogen bond pair between both Asp179e carboxylate oxygens and the Gly130e and Gly131e amide nitrogens is essentially maintained. These structural changes

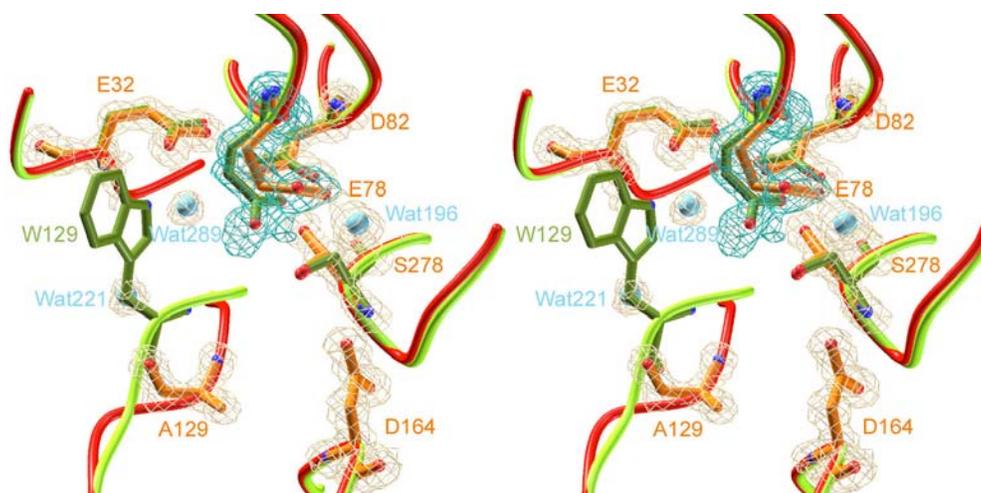


Figure 4.16: Stereo plot of a section around the active site of kumamolisin mutant W129A. Important residues are highlighted as stick models and main chains as ropes. Wild-type kumamolisin (olive-green) superimposed with mutant W129A (red and orange). The final 2Fo-Fc electron density of mutant E32A (beige) contoured at 1.5 σ is only shown around the emphasized side chains. The Glu78 residue is superimposed with the difference electron density omit map (blue) contoured at 3.0 σ .

and the loss of a hydrogen bond donor (Trp129e) do not affect the positioning of the side chains of Glu32e and Asp82e. However, these structural rearrangements cause the Glu78e side chain to adopt multiple conformations (see the omit density in Figure 4.16), resulting in the Ser278e side chain switching between the native and an alternate conformation. Thus, the lack of Trp129e and the consequent structural changes seem to be primarily responsible for the reduction of activity in the W129A mutant.

4.2.3.2 Overall Structure of Pro-Kumamolisin

To prevent auto-activation, pro-kumamolisin had been prepared and crystallized as an inactive mutant, with the catalytic Ser278e replaced by an alanine residue. This single-chain pro-enzyme, virtually lacking any hydrolytic activity (Table 4), crystallized in a different space group (R32) than the catalytic domain (see Table 2 and 3). As shown in Figures 4.17.A and B, the prodomain contacts the catalytic domain at the non-primed-side entrance to the active-site cleft. The connecting linker runs through the entire active-site cleft ending up in a relatively flexible loop, before the chain turns into the globular domain of the mature enzyme. This domain arrangement gives rise to a bowl-shaped pro-enzyme (Figure 4.18.B) with two laterally attached knobs of different size, the larger represented by the prodomain, and the smaller by the exposed, partially flexible C-terminal end of the linker.

A superposition with active kumamolisin reveals that the catalytic domain of the pro-enzyme exhibits, except for the first four N-terminal residues, a virtually identical structure, in agreement with an r.m.s. deviation of 0.49 Å for 359 α -carbon atoms (using a threshold of 1.2 Å). Thus, the catalytic domain is essentially folded into the catalytic competent form within the pro-enzyme. Small but significant differences are detectable in the loop Glu70e-Pro75e encompassing the S1' cavity, and in two loops (Pro101e-Asp104e and 134Aspe-Ala137e) in the interface toward the prodomain, where the peptide linker enters the active-site cleft. The prodomain essentially packs along the two surface-located helices h4 and h5 of the catalytic domain (Figure 4.17.A), grasping with its long α 1- α 2 loop far around the catalytic domain, almost reaching the C-terminal end of the mature enzyme. The contact surfaces between the two compact domains are overall complementary forming a combined interface of ~ 1650 Å² (calculated with the Protein-

Protein Interaction Server), leaving a few cavities where immobilized solvent molecules are observed. At the extreme tip, the partially flexible side chain of Arg69p extends into one of these interfacial cavities, and side chains of Arg48p and Asp150e form an inter-domain salt bridge. Deeper in the interface, both domains provide a number of aliphatic side chains forming a solvent-free intermolecular hydrophobic core.

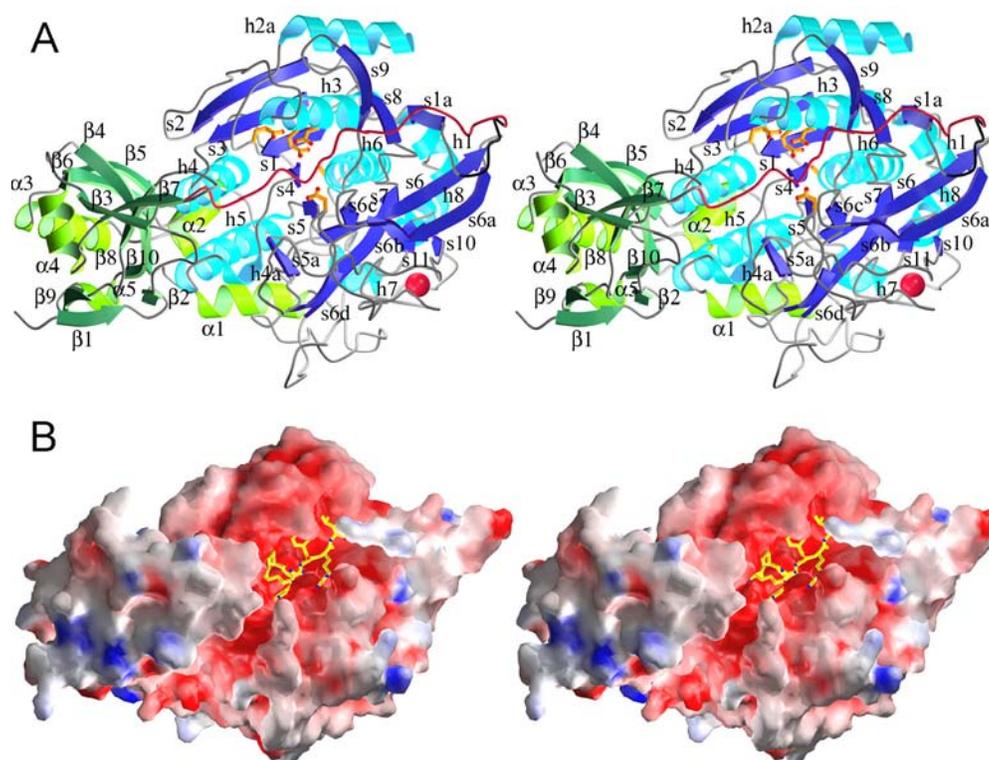


Figure 4.17: 3D structure of full-length pro-kumamolisin. **A)** Stereo ribbon plot of uncleaved pro-kumamolisin shown in standard orientation. The peptide linker between the catalytic domain (in blue) and the prodomain (in green) is colored in red for well-defined and in black for undefined residues. The catalytic residues and the Ca^{2+} ion are shown as orange stick models and a red sphere, respectively. Helices and β -strands are labeled with Greek (propeptide) and Latin letters (catalytic domain). **B)** Stereo solid surface representation of pro-kumamolisin shown in standard orientation as in Figure 4.17.A. The colors indicate the electrostatic surface potential at pH 7 contoured from +20 kT/e (dark blue) to -20 kT/e (dark red). The peptide linker is represented as a yellow stick model visible from P3-Arg169p to P3'-Leu174p.

4.2.3.3 Prodomain of Pro-Kumamolisin

The shell-shaped prodomain of pro-kumamolisin comprises three sub-domains: a compact core, which is flanked by a three-stranded-multiple turn entity at one end and by an elongated loop containing helices $\alpha 1$ and $\alpha 2$ at the opposite end (Figure 4.18). The globular core residues (39p-47p, 85p-126p, and 147p-161p) exhibits an open-sandwich anti-parallel- α /anti-parallel- β fold, made up by the two α -helices $\alpha 3$ and $\alpha 4$ and a twisted β -sheet consisting of the anti-parallel strands $\beta 5$, $\beta 6$, $\beta 4$ and $\beta 10$ (Figures 4.18 and 4.19). Topologically, the pro-kumamolisin core resembles the prodomains of subtilisin BPN' (Gallagher et al., 1995), subtilisin E (Jain et al., 1998) and the mouse pro-hormone convertase PC1 (Tangrea et al., 2001; Tangrea et al., 2002). In spite of the missing sequence homology, the best fitted overlay between these homologs and kumamolisin prodomain yield 40 (subtilisin BPN'), 39 (subtilisin E) and 39 (PC1)

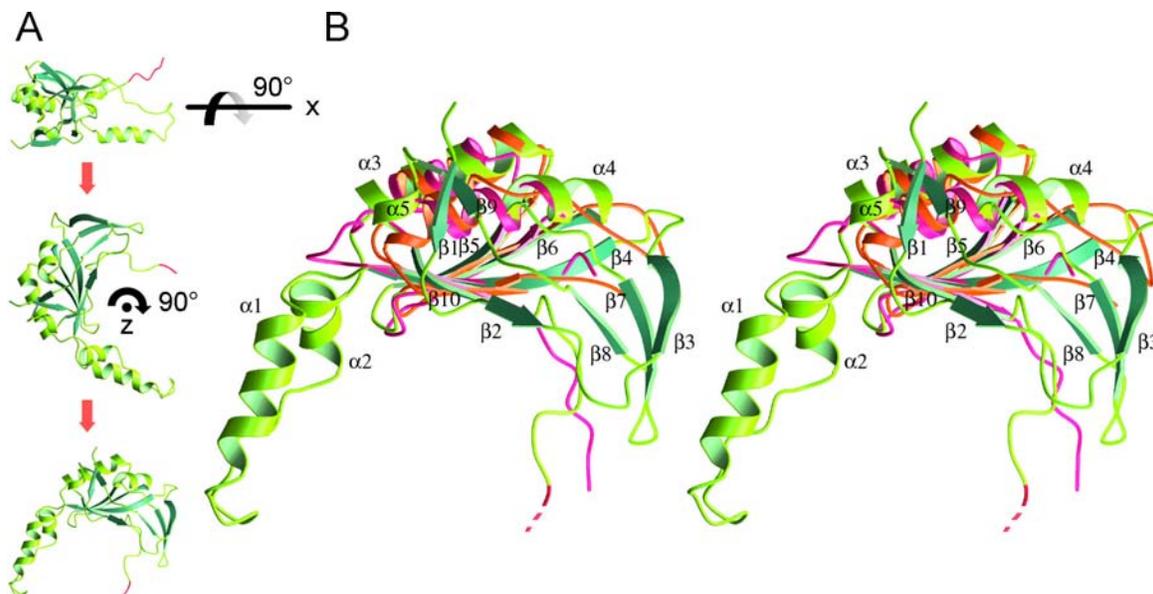


Figure 4.18 Superimposition of the prodomains of kumamolisin, subtilisins and the prohormone convertases. **A**) The new orientation is obtained starting from the standard orientation (Figure. 4.17) rotating 90° along the x-axis and 90° along the new z-axis. **B**) Stereo ribbon plot of kumamolisin prodomain (green) overlaid with the optimally fitted prodomains of subtilisin BPN' (pink) as docked to the mature subtilisin BPN' (Gallagher et al., 1995), and mouse prohormone convertase 1 (orange) as determined for the isolated propeptide (Tangrea et al., 2002). Helices and β -strands are labeled with Greek letters. The dashed red end of the pro-kumamolisin propeptide indicates linker positions P3 and P2.

structurally equivalent α -carbon atoms with r.m.s. deviations of 1.48 Å, 1.32 Å, and 1.46 Å, respectively, using 2 Å thresholds. If the propeptide central cores are optimally superimposed of subtilisin BPN' and pro-kumamolisin, their linkers deviated by a 25° angle from each other (Figure 4.18). This shift reflects slightly different arrangements of the prodomains with respect to the catalytic domains. Noteworthy, partially similar overall folds are also found in the activation domains of other functionally unrelated proteins (Gallagher et al., 1995), among them the prodomains of zinc pro-carboxypeptidases (Coll et al., 1991; Estebanez-Perpina et al., 2001; Gomis-Ruth et al., 1999; Guasch et al., 1992), which block the active sites of the carboxypeptidases in a quite different manner.

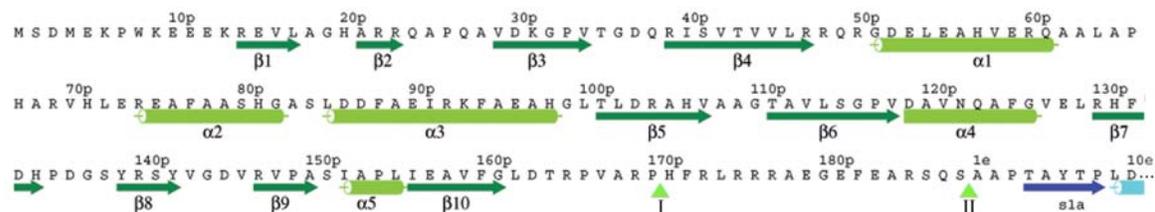


Figure 4.19: Amino-acid sequence of kumamolisin prosegment (Oyama et al., 2002) as stored under the Swiss-Prot accession number Q8RR56. Sequence numbers with the suffix “p” refer to the kumamolisin prodomain, while numbers with the suffix “e” correspond to ten first residues of kumamolisin catalytic domain. β -Sheets and α -helices observed in pro-kumamolisin are represented by arrows and cylinders, respectively (green for the prodomain and blue for the catalytic domain). The primary and the final procession sites are indicated by green arrow heads labeled with I and II.

The prodomain of pro-kumamolisin is considerably larger than its structurally homologous prodomains, giving rise to two additional sub-domains, which extend towards the front and the back of the molecule (in the standard orientation, Figure 4.17). The front-sided three-stranded multiple-turn appendix consists of the anti-parallel strands β 2, β 8 and β 7, and packs against the linker up to Pro166p, partially covering the active-site cleft (Figure 4.17.B). The back-side sub-domain, on the other hand, comprises the two- and the three-turn helices α 2 and α 3, which are connected by an exposed multiple-turn loop. These helices are arranged in a V-shaped manner along the catalytic domain

surface and frame helix h5 (Figure 4.17.A). The sequence homology of the propeptides is quite low among the sedolisins known so far. Noteworthy is a short segment of relatively strong similarity between kumamolisin, sedolisin and CLN2, mapping to the last three secondary structure elements of the prodomain ($\beta 9$, $\alpha 5$ and $\beta 10$), which are not in direct contact with the catalytic domain.

4.2.3.4 Activation Peptide or Linker

Leaving the prodomain core after $\beta 10$, the polypeptide chain of pro-kumamolisin continues into the linker, which forms a short six-residue kink (Leu160p-Pro166p) before entering the active-site cleft at P5-Val167p. From P5-residue Val167p down to P4'-residue Arg175p, the linker runs through the active-site cleft in an overall extended manner, forming seven inter-main chain hydrogen bonds and covering a surface

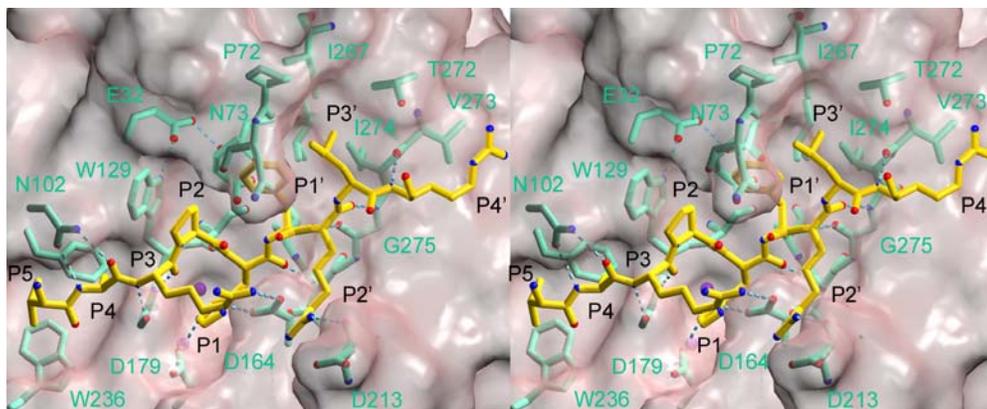


Figure 4.20: Stereo plot of the active-site cleft and the bound linker peptide of pro-kumamolisin. The pro-kumamolisin S278A mutant represented as a cyan stick model is superimposed with a half-transparent pink-colored Connolly surface. The linker peptide is displayed as a yellow stick model, with nitrogen and oxygen atoms given in blue and red. The dashed cyan lines indicate hydrogen bonds between the active site residues and the linker peptide.

of $\sim 1067 \text{ \AA}^2$ (Figure 4.20). The arrangement of the activation peptide side chains along the active-site cleft indicates the subsites S5 to S4', and fits well with the previously modeled bound peptide substrate (Figure 4.10). In the active-site cleft on either side of the scissile P1-His171p—P1'-Phe172p peptide bond (Figure 4.19), each linker residue

exhibits a distinct main chain conformation. These particular conformations are characteristic for the peptide bonds of *canonically binding* protein inhibitors in their complexes with trypsin-like and subtilisin-like serine proteinases (Bode and Huber, 1992) (Figure 4.20; Table 5). Thus, the conformation of P3 and P2' residues is typical of parallel/antiparallel β -pleated sheets, P2 and P1' residues have the conformation of a polyproline II helix, and the P1 amino acid residue attains the conformation of a 3_{10} -helix.

Table 5: Main-chain conformational angles of the linker, compared with the reactive-site loop of the proteinase inhibitor eglin c in its complex with subtilisin (Bode *et al.*, 1986; Bode *et al.*, 1987)

	Pro-kumamolisin linker		Reactive-site loop of eglin C	
	ϕ	ψ	ϕ	ψ
P4	-142.59	157.29	-71.07	140.08
P3	-114.48	139.01	-138.53	167.98
P2	-81.56	123.39	-62.17	143.06
P1	-80.94	42.62	-115.41	44.67
P1'	-116.28	159.39	-96.72	168.83
P2'	-140.08	141.08	-117.41	109.86
P3'	-86.06	139.06	-121.08	111.94
P4'	-63.27	139.06	-75.63	-1.90

The P5-side chain of Val167p slots into the shallow S5 subsite formed by the side chains of Trp136e and Phe107e together with the Thr103e-Asp104e main chain atoms. The P4-Ala168p residue, hydrogen bonded through its amido nitrogen and carbonyl atoms to the side chain of Asn102e, occupies the hydrophobic S4 pocket formed by the side chains of Phe107e, Leu33e and Trp129e (Figure 4.20). The side chain of P3-Arg169p extends away from the surface. Its guanidyl group, however, forms two N-O hydrogen bonds with the Asp164e carboxylate group, which normally contributes to oxyanion hole formation. Identical to the proline residue of the Ac-IPF-CHO inhibitor

(see section 4.1.3.3; Figure 5.3), the P2-Pro170p pyrrolidine ring fits into the hydrophobic S2 cavity, demarcated on one side by the aliphatic part of the side chain of the catalytic Glu78e, and on the other side by the indole moiety of Trp129e. The amido nitrogen of P1-His171p hydrogen bonds to Ser128e carbonyl oxygen, while its imidazole group is hydrogen bonded to the carboxyl carbon of Asp179e. The His171p side chain slots into the S1 groove, which is bordered by main chain segments 130e-131e and 161e-164e, two ordered water molecules at the bottom, and the Asp179e side chain. The P1 carbonyl group extends into the oxyanion hole accepting a hydrogen bond from the Ser278e amido nitrogen. Though, the oxyanion hole lacks the second usual hydrogen bond, since the Asp164e carboxylate group rotates away to form the above-mentioned salt bridge with the P3-Arg169p side chain.

The P1-P1' scissile peptide bond spans the catalytic center in a similar manner to that known for other inhibitor-proteinase complexes (Figure 4.20). Due to lack of the O γ hydroxyl in the S278A mutant, there is no direct contact between the *catalytic* residue Ala278e and the P1-P1' peptide bond. However, the O γ atom of a superimposed Ser278e residue would be only 2.3 Å away from the His171p carbonyl carbon, i.e. ready to attack the scissile bond. The P1'-side chain of Phe172p nestles into the large S1' pocket, flanked by the aliphatic side chain parts of Glu78e and Leu81e and main chain segment 73e-78e (Figure 4.20). The P2'-Arg173p main chain is clamped to the Gly275e main chain atoms via two hydrogen bonds, while its side chain extends along a surface groove bounded by Pro259e, Gly214e, and Gly215e. The distal guanidyl group of Arg173p forms a hydrogen bond to the Asp164e carbonyl oxygen. The Leu174p side chain inserts into the large hydrophobic surface subsite, S3', delimited by residues Pro72e and Asn73e and by the side chains of Ile267, Thr272 and Ile274. Finally, P4'-Arg175p extends its side chain along Val273e. The following 17 propeptide residues, from Arg176p to Thr4e (Figures 4.17.A and 4.19), form a wide loop, which expands into a large intermolecular crystal cavity. Ala178p is the last linker residue rigidly placed, while the segment Glu179p-Ala184p is only defined for its main chain and the eight residues from Arg185p to Thr4e are fully disordered. From Ala5e onwards, the pro-kumamolisin peptide chain is well defined by electron density, exactly following the course observed for the mature enzyme.

4.3 Peptidyl-Dipeptidase Dcp from *E. coli*

4.3.1 Protein Over-Expression and Preparation

Escherichia coli native Dcp and selenomethionine-substituted Dcp (SeMet-Dcp) were over-expressed in different *E. coli* strains with the plasmid pDcp (see section 3.1.3), which contains the OmpA signal sequence that exports the enzyme to the periplasm. Purification to apparent homogeneity of the enzyme succeeded after a three-step purification, combining an anionic exchange DE52 and a phenyl sepharose chromatographies with a gel filtration on a Superdex 200 column (see section 3.4.3.1; Figures 4.21.A and B).

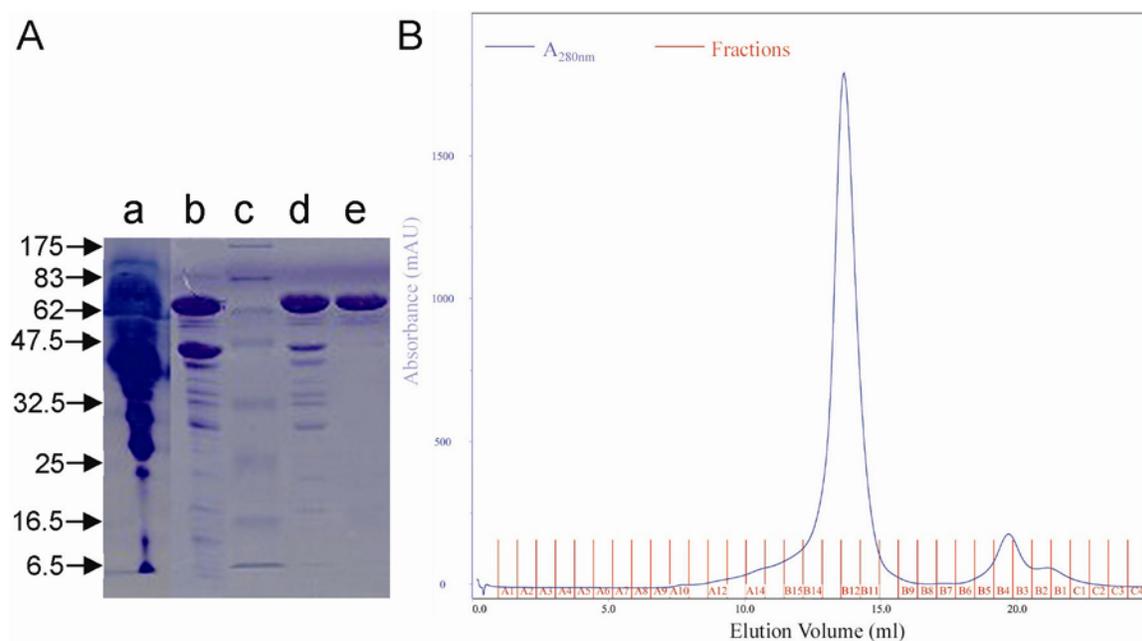


Figure 4.21: **A)** Coomassie stained SDS-PAGE gels of the Dcp purification steps. Lane **a** corresponds to the supernatant obtained after cell disruption (the sample was electrophoresed in a different gel and scaled to the other lanes). Lanes **b**, **d** and **e** show the Dcp sample after anionic exchange DE52, phenyl sepharose and size exclusion chromatographies, respectively. The molecular weight marker was applied in lane **c** and indicates the position of proteins with apparent molecular weights from 175 to 6.5 kDa. **B)** Gel filtration profile of Dcp in complex with the inhibitor N-1450 (H-PYHIKWGD-OH; Bachem, Weil am Rhein) carried out on a Superdex 200 column at 4° C. The first pick corresponds to the enzyme-inhibitor complex and the second one to non-bound inhibitor peptide.

The protein was characterized by N-terminal sequencing, which identified the N-terminus of Dcp preceded by residues ATVAQAL:

- Native sequence: MTTMN... (Heinrich et al., 1993)
- Recombinant sequence: *ATVAQAL* TTMN...

The leucine residue at -1 position was introduced via the cloning strategy at the *Stu* I restriction site of the plasmid inserted in the vector pRBI-DsbC (Maskos, 1995). The prior sequence ATVAQA corresponds to the last six residues of the OmpA signal sequence, which was not processed at the expected site by signal peptidase, presumably, due to the presence of the leucine residue at the C-terminus.

4.3.2 Protein Crystallization

The initial needles bundles from native peptidyl-dipeptidase Dcp, which were obtained by sitting-drop vapor-diffusion procedures at 21° C and a 4° C, could not be improved by any subsequent crystallization parameter screenings, neither by the addition of additives or detergents. In order to facilitate the crystallization of the enzyme new screenings were carried out with the sample obtained after gel filtration of Dcp in complex with different Dcp inhibitors, namely: N-1450, H-2215, H-2225, captopril, and lisinopril (see Figure 3.1). Only in droplets of Dcp-N1450 complex small crystals plates appeared in a crystallization condition containing 0.2 potassium nitrate and 20% (w/v) PEG 3350. After optimization, well-diffracting crystals were grown by mixing 1 µl of a 15 mg/ml Dcp-N1450 complex solution in 50 mM sodium chloride, 20 mM Tris, pH 8.0, with 1 µl reservoir solution consisting of 0.2 M potassium nitrate and 18% (w/v) PEG 3350. Crystals of P2₁2₁2₁ space group with one molecule per asymmetric unit reached a maximal size of about 0.3 x 0.1 x 0.1 mm³ in a period of two to three months at 21° C (Figure 4.22.A). All attempts to soak crystals with heavy atom compounds to obtain derivatives resulted in the loss of the diffraction power after one hour soaking or the lack of heavy atom incorporation in the case of shorter soaking times.

Microseeding with native Dcp crystals turned out to be critical procedure in crystallization of SeMet-Dcp (see section 3.5.3). Derivative crystals were obtained from a mixture of 1µl of SeMet-Dcp-N1450 complex at a concentration of 6mg/ml in 50

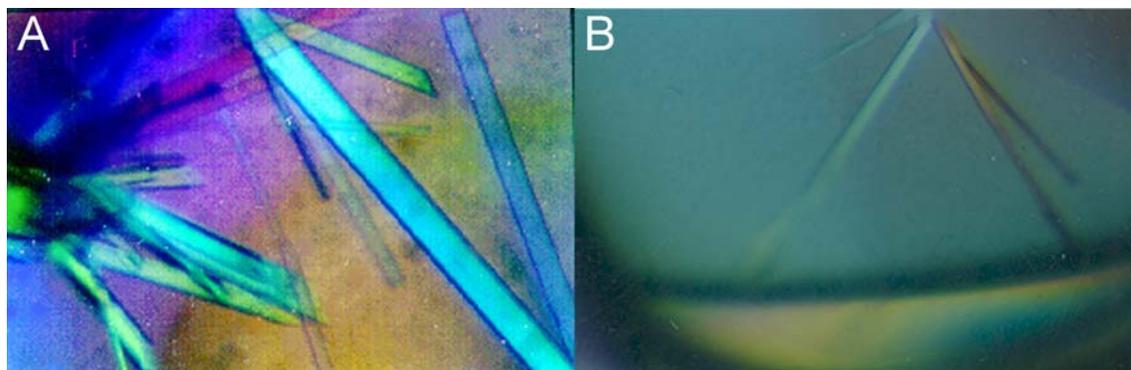


Figure 4.22: Crystals of peptidyl-dipeptidase Dcp-N1450 complex. Crystals grew in the space group P212121 with one monomer per asymmetric unit. **A)** The picture shows the crystals obtained after optimization of initial conditions, which diffracted to 2 Å. **B)** Derivative crystals of SeMet-Dcp in complex with N-1450 inhibitor grew after microseeding and diffracted to a resolution up to 2.4 Å.

mM sodium chloride, 20 mM Tris, pH 8.0 with a reservoir buffer containing 0.13 M potassium nitrate and 16% (w/v) PEG 3350. The crystal plates obtained with SeMet-Dcp after three months had similar morphology as those used for microseeding but with smaller dimensions (Figure 4.22.B).

4.3.3 Native and MAD Data Collection

Flash cooling of the crystals was possible by soaking them for 10 s in a cryobuffer consisting of 15% (v/v) MPD in the respective crystallization buffers and freezing them in a nitrogen stream at 100 K (Oxford Cryosystems Cryostream). All frozen crystals were measured at the DESY BW6 beamline (Deutsches Elektronen Synchrotron, Hamburg, Germany) on a Mar Research CCD detector, where a native data set to 2.0 Å resolution was collected for Dcp-N1450 complex at a wavelength of 1.050 Å (Figure 4.23; Table 7).

As a prerequisite for MAD data collection, an X-ray fluorescence scan was carried out to determine the exact position of the K-shell selenium absorption edge of SeMet-Dcp crystals. According to this scan the wavelengths for MAD data collection were chosen, $\lambda_1 = 0.97880$ Å to maximize the anomalous f'' contribution of the selenium atoms and $\lambda_2 = 0.97910$ Å at the peak of the f' inflection (Table 6). A third, remote dataset was collected at $\lambda_3 = 0.95000$ Å. The rotation angle for every single image was

1.0°. A previous trial to measure a MAD data set at the K-shell zinc absorption edge of native Dcp crystals showed no anomalous contribution in the collected data being the signal of one zinc ion as anomalous scatterer not sufficiently large relative to the total scattering from anomalous plus non-anomalous scatterers.

Figure 4.23: Diffraction image of the native Dcp-N1450 complex. The image was recorded on a Mar CCD detector on beamline BW6 at DESY (Hamburg) with rotation of 1.0° during exposure of the crystal. The detector edge corresponds to a limiting resolution of 2.0 Å at a wavelength of 1.0500 Å.

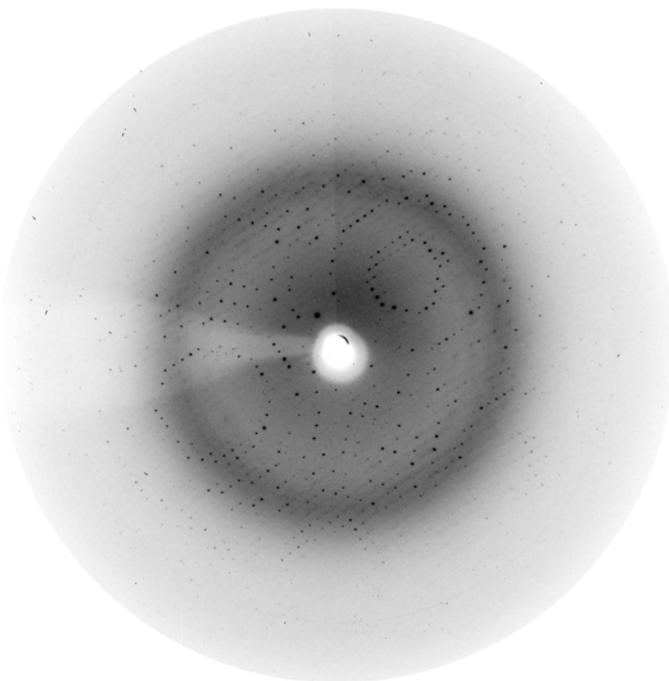


Table 6: Statistics for MAD Data Collection and Phasing

Crystal content	SeMet-Dcp-N1450 complex (monomer)		
	f'' - peak (λ_1)	f'' - inflection (λ_2)	remote (λ_3)
Wavelength (Å)	0.97880	0.97910	0.95000
Resolution range (Å)	20.0-2.4	20.0-2.4	20.0-2.4
Reflections measured	348262	95014	104153
I / σ	7.8	6.8	6.9
Unique reflections	54325	45170	49473
^a R _{merge} (%)	7.8	6.8	6.9
Completeness (%)	99.7 (99.7)	82.9 (82.4)	82.9 (82.1)

$$^a R_{\text{merge}} = \frac{\sum_{hkl} |I| - I}{\sum_{hkl} I}$$

Data in brackets hold for the outermost shell.

4.3.4 Phasing and Refinement

All data sets collected were indexed and integrated using DENZO (Otwinowski & Minor, 1993) and SCALEPACK, respectively. The data were merged, scaled and truncated with programs supported by the Collaborative Computational Project No. 4 (The CCP4 suite, 1994). Collection of MAD data was necessary for phasing because the programs AMORE (Navaza, 1994) and MOLREP (The CCP4 suite, 1994) were unable to obtain interpretable electron density maps for model building by molecular replacement

Table 7: Statistics for Data Collection and Refinement

Data Collection	
Crystal content	Dcp-N1450 complex (monomer)
Space group	P2 ₁ 2 ₁ 2 ₁
Cell Constants (Å)	a = 63.41; b = 67.92; c = 153.59
Resolution range (Å)	20.0 - 2.0
Reflections measured	276452
I/σ	12.6
Unique reflections	48916
^a R _{merge} (%)	8.0
Completeness (%)	99.9 (99.8)
Refinement	
Reflections used for refinement	45498
Solvent content (%)	42
Residue number per asymmetric unit	680
Water molecules per asymmetric unit	629
Zinc ions per asymmetric unit	1
^b R _{factor} (%) / ^c R _{free} (%)	21.3/27.3
R.m.s.d. Bonds (Å)	0.010
R.m.s.d. Angles (°)	1.40
Average B-factors (Å ²) protein atoms	33.6
Average B-factors (Å ²) water molecules	44.8
Average B-factors (Å ²) zinc ion	23.4
Average B-factors (Å ²) inhibitor atoms	38.4
^a R _{merge} = $\frac{\sum_{hkl} (I - I)}{\sum_{hkl} I }$	
^b R _{factor} = $\frac{\sum_{hkl} F_{obs} - F_{calc} }{\sum_{hkl} F_{obs} }$	
^c R _{free} is the R-value calculated with 2000 reflections not used in refinement.	
Data in brackets hold for the outermost shell.	

using neurolysin or ACE as models. The program package CNS (Brunger et al., 1998) was used for scaling and merging of MAD data and for all further phasing steps. Eighteen selenium positions were obtained with a phasing power of 0.92 and a figure of merit (*fom*) of 0.56 over the whole resolution range. A better electron density was obtained after solvent flattening, using only data to 3.0 Å (phasing power 1.11 and *fom* 0.63), where characteristic features such as α -helices and β -strands but no side chains could be distinguished. The initial density allowed building of a first poly-Ala model, which was combined with the experimental phases for new electron density calculation. Subsequently cycles of building, refinement of the model, and combination of experimental and model phases led to improvement of the density and the identification of the side chains.

The refined SeMet-Dcp to 2.4 Å resolution was used as a search model for molecular replacement with native Dcp data using the program AMORE (Navaza, 1994). The best solution yielded a correlation coefficient and an R-factor values of 38.6% and 45.2%, respectively (for 15.0 to 3.5 Å data). Model refinement was performed by consecutive rounds of model building on an SGI graphics workstation and refinement using CNS (Brunger et al., 1998) until convergence. The final R_{free} and R_{factor} obtained were 21.3 and 27.3% respectively (Figure 4.24; Table 7).

N-terminal sequencing of a dissolved Dcp-N1450 crystal revealed that the protein crystallized consists of a 1:1 mixture of Dcp starting at N-TTMN... (consistent with the sequence deposit by Henrich et al., 1993) and Dcp modified with part of the OmpA sequence: N-ATVAQALTTMN... (recombinant sequence, see section 4.3.1). Nevertheless, electron density defines only the complete Dcp molecule from residue Thr1 up to the C-terminal Ile680, from which only the residues segment Ile45-Gln50 is partially defined and 5% of the side chains are fully disordered. A single zinc ion was identified and refined with full occupancy in every Dcp molecule. Water molecules were automatically built with CNS (Brunger et al., 1998) during the last steps of refinement. An additional electron density near the active site was interpreted as two segments of the cleaved inhibitor N-1450 (H-PYHIKWGD-OH). The first part of the density was

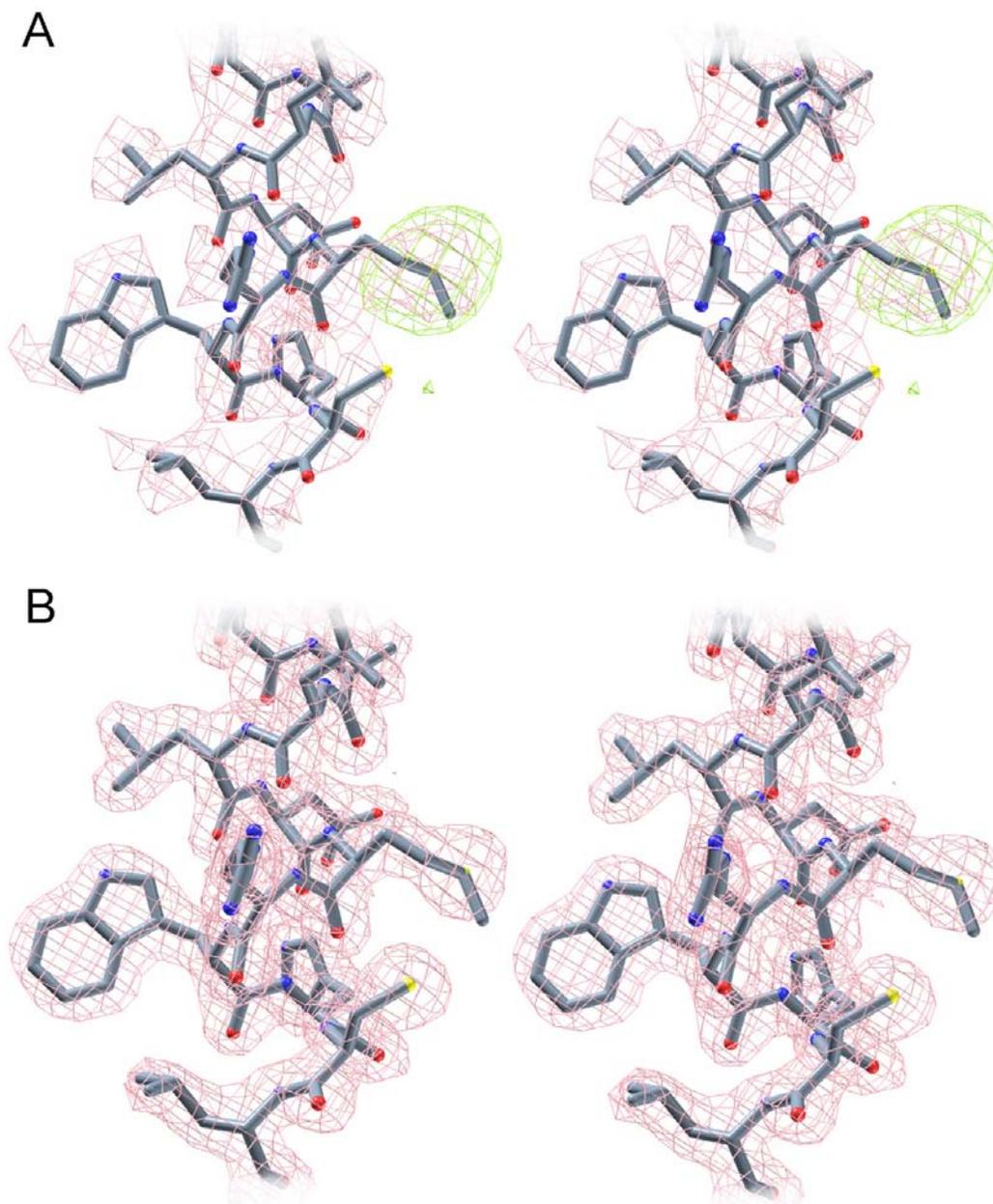


Figure 4.24: Exemplary electron densities. **A)** The experimental density of Dcp contoured at 1.0σ after MAD phasing and solvent flattening is shown in pink, while the anomalous difference Fourier map contoured at 2.5σ is shown in green. This density was used to build the poly-alanine model. **B)** The same region of the protein with the final, refined $2Fo-Fc$ electron density contoured at 1.0σ . The pictures show the end of helix $\alpha 20$.

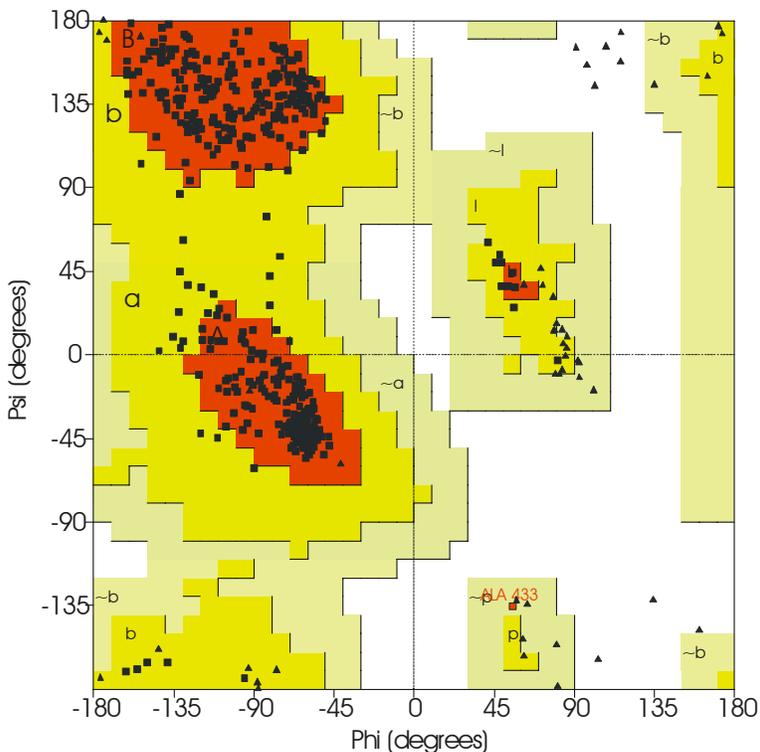


Figure 4.25: Ramachandran plot of native Dcp-N1450 complex crystal structure. The figure plots ϕ angle (axis of abscisses) versus ψ angle (axis of ordinates) for each residue. All residues are situated in the most favored, additionally favored regions except for three residues, which fall into generously allowed regions.

interpreted as the dipeptide H-Gly-Asp-OH, which corresponds to the C-terminal dipeptide of N-1450 (see section 3.5.2.2; Figure 3.1), and was refined with full occupancy except for the side chain of the aspartate that is half occupied. Following the sequence of the octapeptide N-1450, the second segment was built as *N*-Lys-Trp-OH, but only the main chain atoms and the CB of the tryptophan side chain were refined with full occupancy. The side chains of the tryptophan and the lysine as well as the rest of the inhibitor are not visible in the electron density. Main chain angles of the refined structure were analyzed by the program PROCHECK (Laskowski, 1993) and plotted in a Ramachandran diagram (Ramachandran and Sasisekharan, 1968), which shows most favored or additionally favored conformations for 99.5% of the residues (Figure 4.25).

4.3.5 Description of the Structure

4.3.5.1 Overall Structure of Peptidyl-Dipeptidase Dcp

E. coli Peptidyl-dipeptidase Dcp in complex with the inhibitor N-1450 (H-PTHILWGD-OH; Figure 3.1) crystallized as a monomer with dimensions of 64 x 63 x 59 Å in the orthorhombic space group P2₁2₁2₁. Dcp adopts an overall prolate ellipsoid shape that forms a large cavity (Figure 4.26), which extends from the upper tight to the bottom with about 46 Å length and 18 Å diameter in the widest section. This cavity divides vertically the molecule into two different domains, named domain I and domain II (Figures 4.26 and 4.27). The internal surface of domain I accommodates the buried active site of Dcp roughly at the midway along the length of the channel (see section 4.3.5.3).

The protein fold consists basically of α -helical secondary structure elements that account for 60.4% of all residues. The only structural β sheets, which build up just the 5.6% of the molecule, occur as a two-strand entity on the bottom of domain I and a

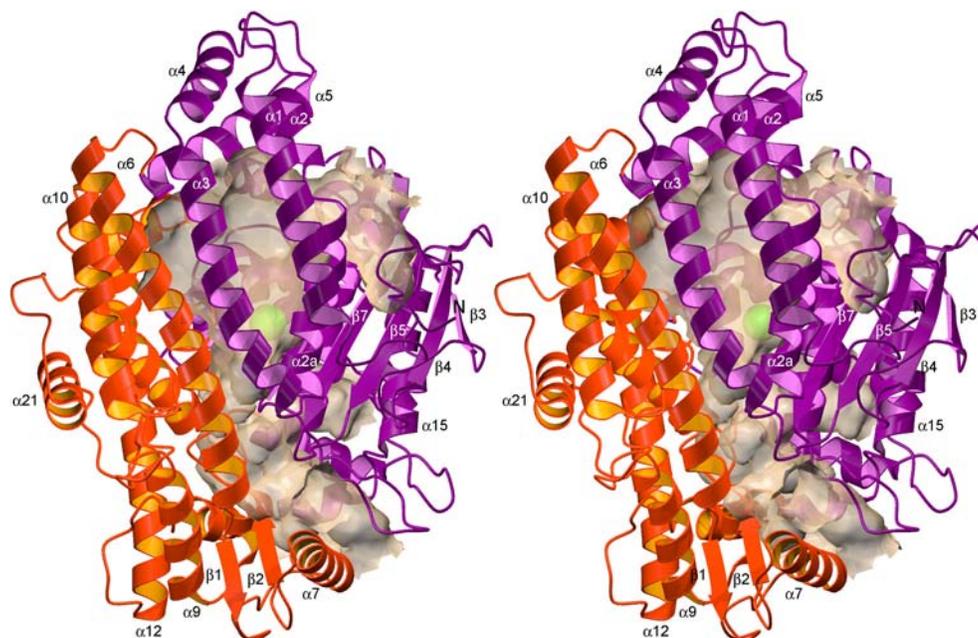


Figure 4.26: Stereo ribbon plot of peptidyl-dipeptidase Dcp in complex with the inhibitor N-1450. A front view of the *E. coli* enzyme with the cavity, which is vertically oriented, displayed as a brown internal surface. Domain I is shown in purple and domain II in orange. The zinc ion of the catalytic site is represented as a green ball, whereas the inhibitor is not shown for clarity.

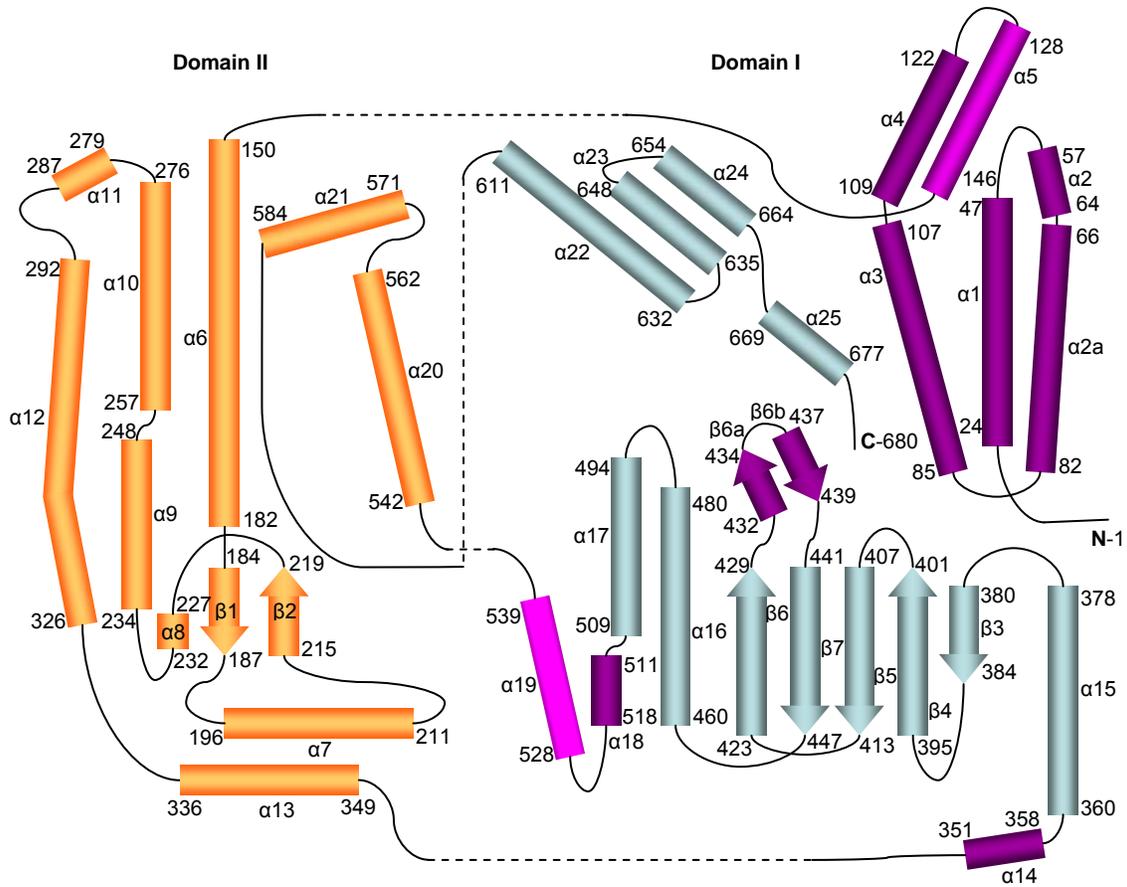


Figure 4.27: Topology diagram of peptidyl-dipeptidase Dcp. The α -helices represented as rectangles and the β -strands drawn as arrows are numbered sequentially from the N-terminus. The residue numbers are indicated for the beginning and end of the secondary structure elements. Domain I (drawn in purple and cyan) and domain II (colored in orange) are separated by dashed lines for clarity. Secondary structure elements of domain I shown in cyan are conserved in other families of metallopeptidases, such as the thermolysin family.

five-stranded sheet situated on the lower half of domain I (Figures 4.26 and 4.27). In addition a small two-strand element is found in the loop connecting strands $\beta 6$ and $\beta 7$ of the later β -sheet. The N-terminal loop segment connects to a three helix bundle made up by $\alpha 1$ flanked on the left side by $\alpha 3$ and on the right side by $\alpha 2a$ - $\alpha 2$. The later two helices, which run roughly parallel to $\alpha 1$ and $\alpha 3$, are interconnected by a short loop that introduces a kink in its direction (Figures 4.26 and 4.27). The first part of domain I, comprising the helical bundle and two subsequent helices ($\alpha 4$ and $\alpha 5$), shapes the front

wall and the top of the Dcp cavity, respectively. After $\alpha 5$, the polypeptide chain forms a number of helices, the longest of which ($\alpha 6$, $\alpha 9$, $\alpha 10$, and $\alpha 12$) run approximately parallel to the channel of the enzyme. This α -helical arrangement, together with the first short two-stranded sheet, constitutes the major part of domain II. The main chain trace then, passes across again to domain I and folds into the core of the molecule that comprises the remaining five and two-stranded β -sheets flanked on the back side (according to the orientation in Figure 4.26) by three helices, including the two helices that carry the active site ($\alpha 16$ and $\alpha 17$). C-terminal to the active site sequence, following $\alpha 18$ and $\alpha 19$, the chain crosses back to complete domain II with the large internal $\alpha 20$, which in addition to loop residues 593-610 mainly contour the left wall of the molecule hollow, and $\alpha 21$ allocated at the outer surface of Dcp perpendicular to the cavity. Finally, back to domain I, the fold of the enzyme ends in four helices situated at the backside of the molecule behind the active site core.

Domain I and II are held together by very short loops between helices that cross the channel at the backside of the molecule ($\alpha 5$ - $\alpha 6$, $\alpha 13$ - $\alpha 14$, and $\alpha 19$ - $\alpha 20$; Figure 4.27), and by a large internal loop between $\alpha 21$ and $\alpha 22$, which interacts directly with the active site and the inhibitor residues. In the front face of the enzyme, in contrast, domain I contacts with domain II through a number of hydrogen bonds between residues of the close helices $\alpha 3$ and $\alpha 6$. Within both domains, the internal channel of Dcp displays four small openings to the molecular surface of diameters ranging from 2.5 to 5.3 Å. In agreement with the tight packing of the secondary elements in each domain, only one orifice, which is the smallest one, opens in the middle of domain I, above the five-stranded sheet. The remaining three apertures emerge in the interface between domain I and II, one at the top of the backside, a second one at the bottom right and the largest one at the front bottom of the molecule (Figure 4.26).

4.3.5.2 Structural Comparison with other Metallopeptidases

E. coli enzyme peptidyl-dipeptidase Dcp belongs to the M3A family of metallopeptidases together with other members, such as saccharolysin from *Saccharomyces cerevisiae* (Buchler et al., 1994), or the mammalian homologues thimet

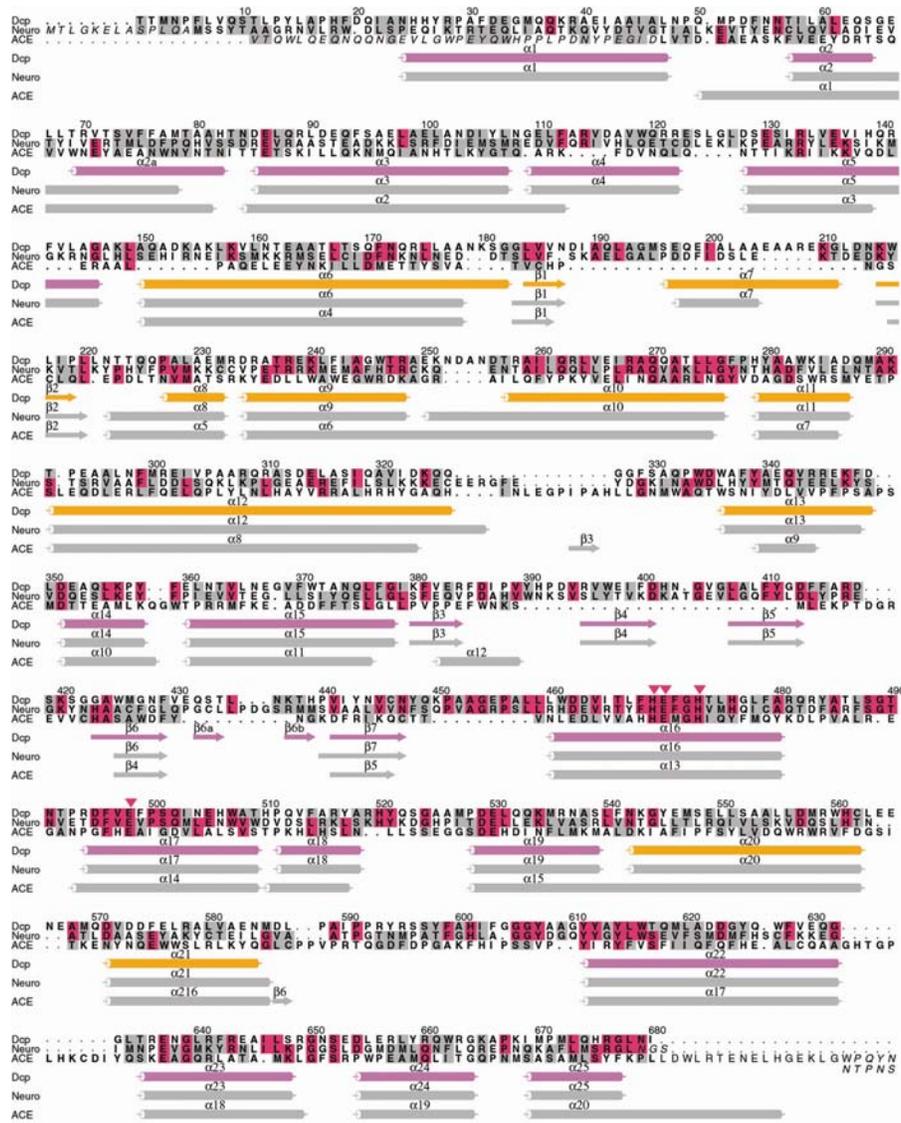


Figure 4.28: Topological and sequence comparison of peptidyl-dipeptidase Dcp with other members of the SCP family and subtilisin Carlsberg. Structure-based alignment of the amino acid sequences of *E. coli* peptidyl-dipeptidase Dcp (Dcp), rat neurolysin (neuro) (Brown et al., 2001) and human angiotensin-converting enzyme (ACE) (Natesh et al., 2003). The identities of these enzymes to Dcp based on topologically equivalence are 24.0% for neurolysin and 6.9% for ACE. The grey and red shading indicates structurally equivalent residues that are conservatively substituted and conserved residues, respectively, in at least two of the three proteins. The sequence numbers refer to Dcp and its secondary structure elements follow the same color code as in Figure 4.26 (domain I in purple; domain II in orange); β -sheets and α -helices are represented, for the three structures by arrows and cylinders, respectively. The catalytic residues are indicated by red arrows, and the inhibitor-binding residues are marked by black arrows. Parts of the neurolysin and ACE sequences not ordered in their crystal structures are shown in italics (the final five residues completing the ACE sequence are written at the end of its last line).

oligopeptidase (McKie et al., 1993; Pierotti et al., 1990) and neurolysin (Dauch et al., 1995). The crystal structure of native recombinant rat neurolysin, which shares 24.0% sequence identity with Dcp (Figure 4.28), is the only three-dimensional structure of the family reported so far (Brown et al., 2001). Superposition of the whole Dcp molecule with that of neurolysin reveals a virtually conserved overall fold, exhibiting a counterpart in the mammalian homologue for almost every structural components of Dcp. The only two exceptions are $\alpha 2$, which in Dcp is split in $\alpha 2/\alpha 2a$, and the two stranded β -sheet $\beta 6a/\beta 6b$ not found in neurolysin. In agreement with the highly conserved three-dimensional structure of the enzymes, their global superimposition gives rise to an r.m.s. deviation value of 1.28 for 273 topologically equivalent α -carbons using a threshold of 2.0 Å. However, structural comparisons of each domain separately between both

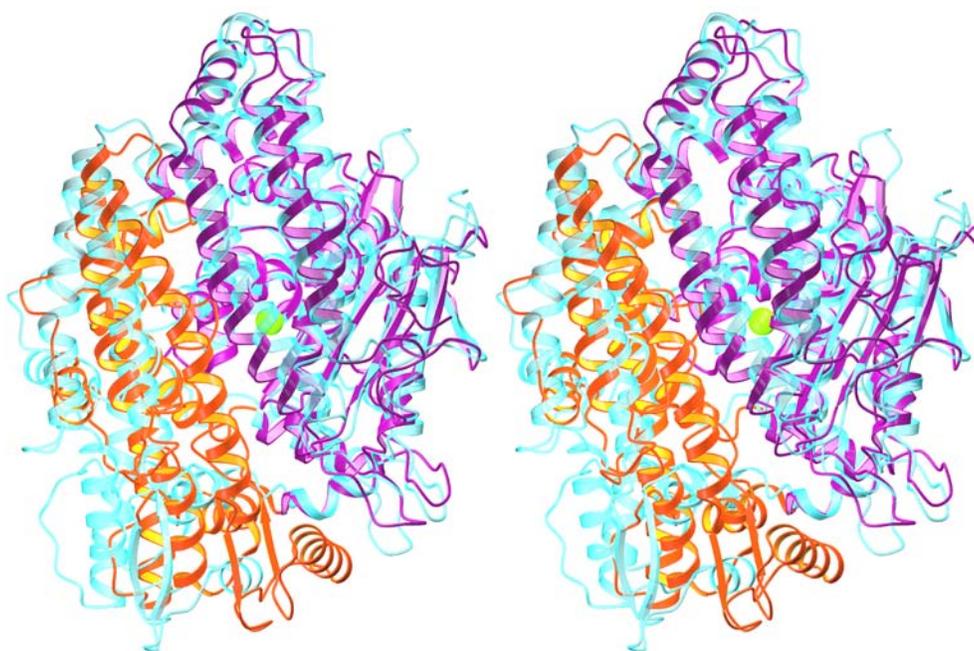


Figure 4.29: Superposition of inhibitor-bound peptidyl-dipeptidase Dcp and native neurolysin. Domains I of both enzymes were superimposed with an r.m.s deviation of 1.23 for 287 structurally related α -carbons. The stereo ribbon plot is shown maintaining the orientation and the color code (for domain I purple and orange for domain II) of Dcp in Figures 4.26 and 4.28. The related neurolysin structure is displayed as a light transparent cyan ribbon, while the zinc ion position conserved in both molecules is indicated as a green sphere.

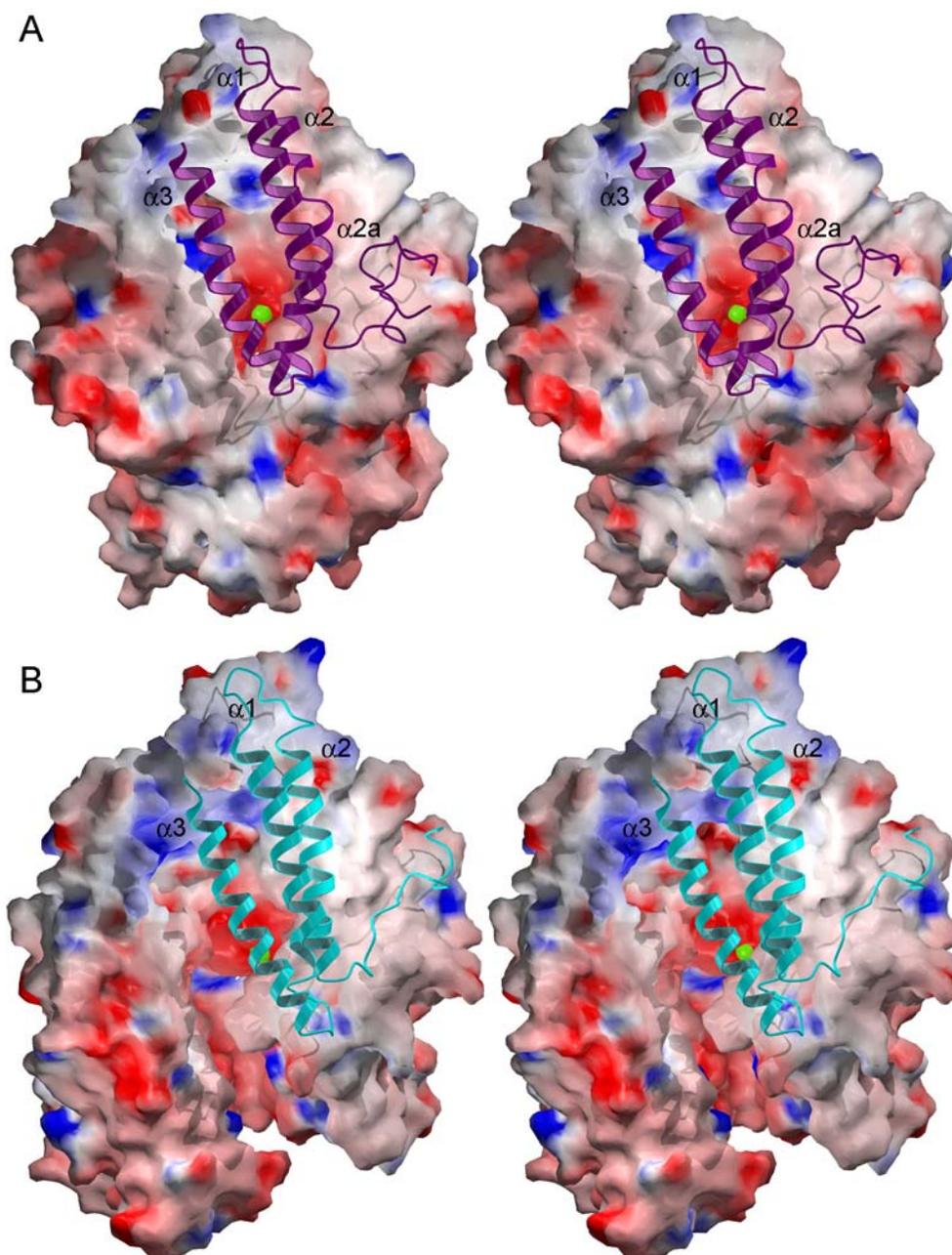


Figure 4.30: Stereo electrostatic molecular surface representations showing the active site grooves of **A)** inhibitor-bound Dcp and **B)** native neurolysin. Negative and positive potentials are colored in red and blue, respectively. The molecules are oriented in the same view as in Figures 4.26 and 4.29. For clarity, the molecular surfaces have been calculated without helices $\alpha 1$, $\alpha 2/\alpha 2a$ and $\alpha 3$ of Dcp and the corresponding $\alpha 1$, $\alpha 2$ and $\alpha 3$ of neurolysin, which are shown as purple and cyan ribbons, respectively. In both enzymes (**A** and **B**) the zinc ion is represented as a green sphere.

enzymes, yielded r.m.s values of 1.23 and 1.10 (threshold 2.0 Å) with 287 and 180 structurally related α -carbons for domains I and domains II, respectively (notice that in the present work the domain labeled as I is that including the N-terminal in Dcp and neurolysin, while Brown and coworkers named the domains in neurolysin the other way around without justification). Indeed, optimal superposition of domain I bearing the active site residues and the catalytic zinc ion, reveals that the domain II of inhibitor-bound Dcp is closer to domain I than domain II in the native neurolysin (Figure 4.29). The shift between both domains II reaches the maximal distance of 18 Å between equivalent C α carbons at the bottom of the molecules, whereas it is nearly negligible at their top (Figure 4.29). The move of domain II in Dcp with respect that of neurolysin accounts for the dramatic difference on their molecular surfaces (Figures 4.30.A and B) despite their conserved overall folds (Figures 4.28 and 4.29). The central channel in neurolysin appears open and accessible from the front side and the bottom (Figure 4.30.B), whereas in Dcp it is completely surrounded by the two domains and it remains buried as a cavity inside the molecule (Figures 4.26 and 4.30.A) with four small outer openings (see section 4.3.5.1).

Further structural comparison of peptidyl-dipeptidase Dcp with other known protein structures using the DALI server (Holm & Sander, 1999), placed testis angiotensin-converting enzyme (tACE, the isoform of ACE found specifically in testicles; Natesh et al., 2003) as the second match (after neurolysin) of structurally similar proteins. ACE is a member of the M2 family of metallopeptidases and, apart from the HEXXH zinc binding motif HEXXH, shares very little sequence identity to Dcp (6.9% after structural based sequence alignment) or to neurolysin (7.9%; Figure 4.28). Nevertheless, the best structural superimposition of tACE with Dcp revealed noteworthy homology showing a remarkable r.m.s. deviation of 1.35 Å for 212 C α atoms, using a threshold of 2 Å (Figure 4.31.A). In general, the core structures of both proteins are similar, with significant differences in loops, in the length of some topological elements and in the five-stranded β -sheet, which in ACE is a two-stranded sheet next to a short helix (α 12) (Figures 4.28 and 4.31.A). The crystal structure of native tACE and in complex with a potent inhibitor (lisinopril) exhibits, as Dcp, a central groove with small orifices of approximately 6.0 and 4.0 Å at the top and bottom of the molecule, respectively (Figure 4.31.B).

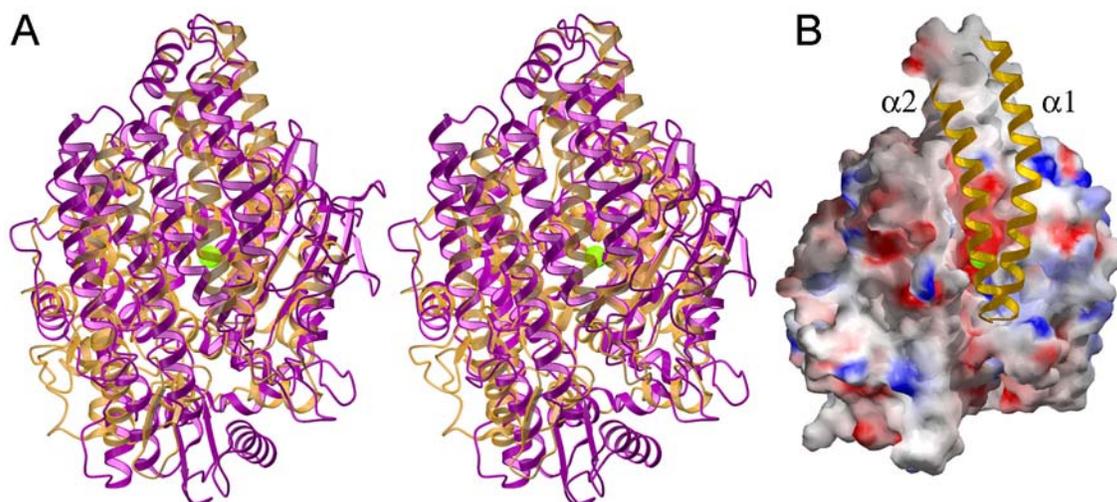


Figure 4.31: Structural comparison of peptidyl-dipeptidase Dcp and testis angiotensin-converting enzyme. **A)** Stereo superimposition of Dcp with tACE shown as a purple and light transparent yellow ribbons, respectively. **B)** The molecular surface of tACE displays its positive and negative electrostatic potentials with blue and red colors. For clarity, the surface has been sliced and helices $\alpha 1$ and $\alpha 2$ of tACE are represented as a yellow ribbon. In **A** and **B** the view is identical as the used for Figures 4.26 and 4.30. **A** and **B**, and the zinc ion is again symbolized by a green sphere.

Within the MA(E) clan characterized by the HEXXH active site motif, a more distant related structure of Dcp corresponds to that of thermolysin (Holmes and Matthews, 1982), which is a well studied member of the M4 family. Although the size of individual elements differs in the two enzymes, a large portion of thermolysin active site region have a counterpart in peptidyl-dipeptidase Dcp, including the five strands ($\beta 3$, $\beta 4$, $\beta 5$, $\beta 6$ and $\beta 7$), a helix ($\alpha 15$) N-terminal to the sheet, and the two active site helices ($\alpha 16$ and $\alpha 17$) (Figure 4.27). In addition, the C-terminal four helices in Dcp ($\alpha 22$ - $\alpha 25$) match those in thermolysin, with a small two β -strand insertion in the latter enzyme.

4.3.5.3 Active Site and Substrate Binding

The three-dimensional structure of Peptidyl-dipeptidase Dcp was solved as a complex with the peptidic inhibitor N-1450 (H-PYHIKWGD-OH; see section 4.3.4). Inspection of the interactions between N-1450 and Dcp revealed important residues responsible for inhibitor binding, catalysis and, presumably, for substrate binding. The

presence of two segments of N-1450, nearby the catalytic zinc ion and the HEXXH motive clearly identified the active site of peptidyl-dipeptidase Dcp, which is buried inside the groove of the enzyme on the internal surface of domain I (Figure 4.26; the residues of inhibitor N-1450 will be referred in the following numbering from the N- to C- terminal with the suffix “i” to distinguish them from the enzyme residues). The fully occupied catalytic zinc is coordinated by four ligands in a virtually tetrahedral coordination sphere (Figure 4.32). The two histidine ligands of the conserved HEXXH sequence, His469 and His473, arise from the third and fourth turn of helix α 16, respectively. Glu498 situated at the first half of helix α 17, 15 residues C-terminal from the typical zinc-binding motive, is the third coordination ligand provided by Dcp (Figure 4.32). Usually in native structures of gluzincins, a water molecule completes the coordination sphere of the zinc ion. In Dcp, this solvent molecule is replaced by one of the two carboxylate oxygens of Trp6i of inhibitor N-1450 (Figures 4.32 and 4.33).

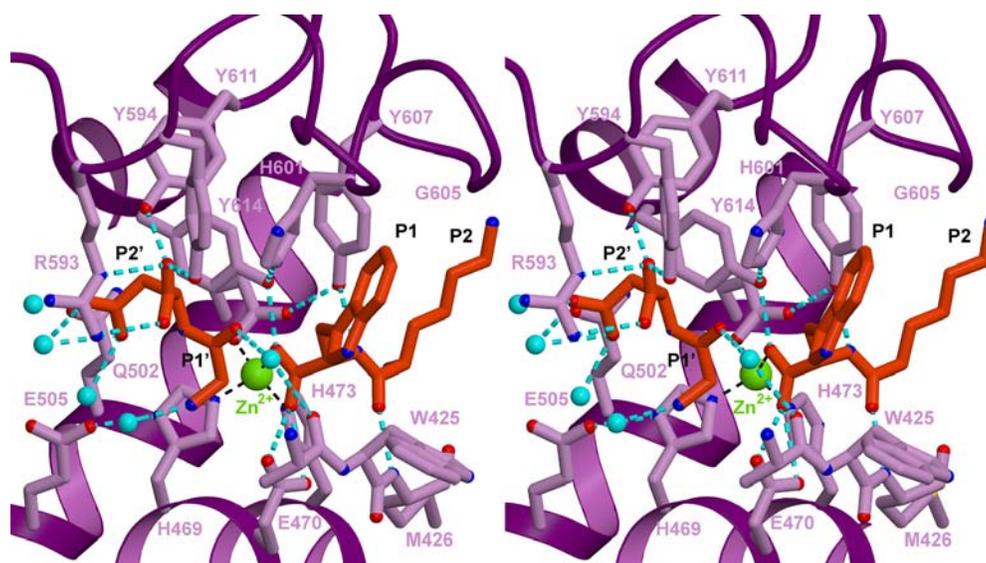


Figure 4.32: Stereo detail of the active site of the bound Dcp. Binding of N-1450 (orange) to Dcp (purple). The view corresponds to that of Figure 4.26 rotated clockwise about 90° around the z axis. Important secondary structure elements are marked. Selected residues are shown in a ball-and-stick representation with zinc ion in green and water molecules in cyan. Hydrogen bonds connecting N-1450 to Dcp or to water molecules are indicated by cyan dotted lines, whereas the zinc coordination is shown by thinner black dotted lines.

C-terminal carboxylate is hydrogen bonded to the side chains of Glu470, Tyr607 and Tyr614 (Figure 4.33). The later Tyr614 is, at its turn, hydrogen-bond connected to one imidazol nitrogen of His601 (Figures 4.32 and 4.33).

The prime side of the active site cleft is occupied by the cleaved dipeptide H-Gly-Asp-OH, which corresponds to the two C-terminal residues of the octapeptide inhibitor (see Mat and Met). The N-terminal nitrogen of Gly7i, which is situated at ~ 4.5 Å from C-terminal carbonyl carbon of Trp6i, hydrogen bonds to a water molecule that, at its turn, is within hydrogen bond distance of the side chain of Thr466 (Figure 4.33). Another solvent molecule interconnects the main chain of Ala424 to the carbonyl oxygen of Gly7i (Figures 4.32 and 4.33). The last visible inhibitor residue, Asp8i, is anchored to the enzyme through its C-terminal carboxylate. The phenolic groups of Tyr594 and Tyr611 donate each a hydrogen bond to one of the carboxylate oxygens. Both carboxylate oxygens interact with the side chain of Arg593 by hydrogen bonding two nitrogens of the guanidyl group. Finally, a water molecule donates a hydrogen bond to one of the C-terminal oxygens of Asp8i.

Besides the zinc coordination sphere and hydrogen-bond contributing residues, there are an additional eight residues of Dcp that make close contacts with N-1450 (< 4.5 Å), most of which are shown in Figure 4.33. These residues are not interacting with the inhibitor directly through hydrogen bonds, but provide important electrostatic and van der Waals interactions defining the active site cleft and the specificity pockets of the enzyme. The active site cleft of Dcp can be described as a channel with a “diabolo” shape. The diameter of the channel narrows gradually from the non-prime subsite S2 to the location of the catalytic zinc, in which the channel achieves its smaller dimensions, and from that point the diameter grows again along the prime side (Figure 4.34). In the non-prime side of the cleft, the structure of Dcp possesses wide open pockets as subsites. The fact that S2 and S1 are rather spacious agrees with the lack of density for the side chains of P2-Lys5i and P1-Trp6i that indicates weak interactions with the enzyme (see section 4.3.4). Following the main chain conformation of P2-Lys5i, the S2 subsite would extend along the aliphatic part of Tyr607 side chain being only limited at the top by the main chain of Ala606 (Figures 4.33 and 4.34). The last visible atom, CB, of the Trp6i side chain, suggests an arrangement of P1 as shown in Figure 4.34. The topology and chemical

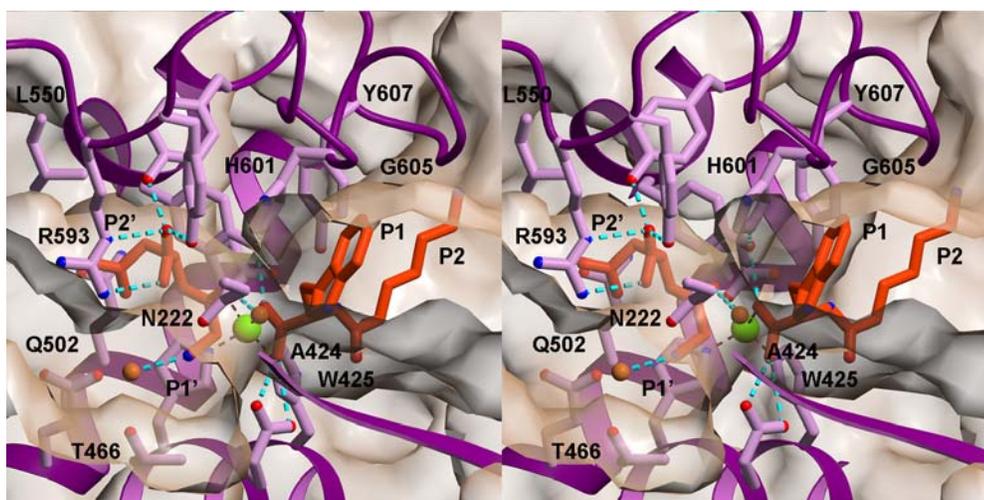


Figure 4.34: Stereo representation of the active site cleft of the inhibitor-bound Dcp. Same orientation as Figure 4.32. Peptidyl-dipeptidase Dcp residues, which is shown as a purple ball-and-stick model, is superimposed with a half-transparent Connolly surface. The surface has been sliced for clarity of the figure. Bound inhibitor N-1450 is displayed as a dark orange stick model, while the zinc ion is represented by a green sphere. The dashed cyan lines and the small brown spheres indicate hydrogen bonds and water molecules, respectively. Residues involved in the formation of Dcp subsites are marked with black labels.

of S1 are dictated by five residues that form a large cavity able to harbor bulky P1 side chains. Laterally, S1 is shaped by the side chains of His601 and Tyr607, whereas the main chains of His601 and Gly605 limit the top of S1 by forming a lid over the P1 side chain. The bottom of the S1 subsite is lined by the main chains of Trp425 and Ala424.

Since in Dcp structure the P1' position is occupied by Gly7i, the absence of a P1' side chain makes the precise description of the subsite difficult. Nevertheless, the narrow surroundings of the Gly7i main chain suggests that the P1' side chain would be directed towards the P1' position indicated in Figure 4.34. Hence, S1' is contoured at the top by the side chains of Asn222 and Arg593, at the bottom by Thr466 side chain, and at the right side by the main chains of Gly423 and Ala424. The P1' side chain would extend towards the cavity as no Dcp residues cover this side S1' pocket. Finally, the side chain of Asp8i packs nicely into the S2' subsite of the enzyme. The rather small S2' cavity is formed by the side chains of five residues, namely: Gln502, Leu550, Arg593, Tyr611 and Tyr614.

5 Discussion

5.1 Kumamolisin a Serine-Carboxyl Proteinase from *Bacillus novosp. MN32*

Most of the serine proteinases can be classified into three evolutionarily unrelated clans: SA (chymotrypsin-like) (Lesk and Fordham, 1996), SB (subtilisin-like) (Siezen and Leunissen, 1997) and SC (α/β -hydrolase fold) (Ollis et al., 1992), which differ in their overall fold but conserve the Ser-His-Asp catalytic triad. The recently identified sedolisin proteinases maintain a subtilisin-like fold, but possess a novel catalytic triad consisting of a Ser/Glu/Asp triad, which is able to confer proteolytic activity at acidic pH values. The crystal structures of native kumamolisin and of its complexes with tripeptidyl aldehyde inhibitors bound at pH 3.0 and 4.5, are the second example of serine-carboxyl proteinase studied by X-ray crystallography, and the first case of a thermostable enzyme belonging to this family. In kumamolisin, the two inhibitors bind covalently to Ser278, identifying this residue as the nucleophilic element. In addition, the superposition of kumamolisin structure with those of sedolisin and subtilisin Carlsberg (Figure 4.7.A) shows that Ser278 corresponds to the catalytic residues Ser287 (sedolisin) and Ser221 (subtilisin). This functional and topological equivalence is in agreement with our site-directed mutagenesis studies in kumamolisin, which show that the substitution of Ser278 by an alanine leads to the complete loss of both auto-processing and proteolytic activity. The equivalent mutants of sedolisin, XSCP (Oda et al., unpublished) and CLN2 (Lin et al., 2001) are likewise inactive.

Similar to sedolisin, the active site Ser278 O γ of kumamolisin is hydrogen bonded to Asp82 via the Glu78 carboxylate (Figures 4.9 and 5.3). Unexpectedly, however, the Asp82 carboxylate of kumamolisin is linked to another acidic residue, Glu32, which in turn is hydrogen bonded to the indole N ϵ atom of Trp129 (Figure 5.1). The many acidic groups clustering within and around the active site are expected to cause a slight increase of the pK $_a$ values of one or both, Glu78 and Asp82 (from their intrinsic pKs of about 4.2 and 3.9). On the other hand, the pK $_a$ of Glu32 might be lowered, due to the additional hydrogen bond donated from Trp129 and the positive charge moment of its indole ring

edge (Figures 4.9 and 5.1). The very short O \cdots O distances (of around 2.6 Å, see Figure 5.1) within this active site network seem to indicate the presence of strong, low-barrier hydrogen bonds (Frey et al., 1994; Flocco et al., 1995). Such short O \cdots H \cdots O interactions between acid/base groups of similar pK sharing a common hydrogen, in the low dielectric environment of protein active sites, are associated with high free energies and might therefore stabilise an overall negatively charged resonance network in the free enzyme at acidic pH values.

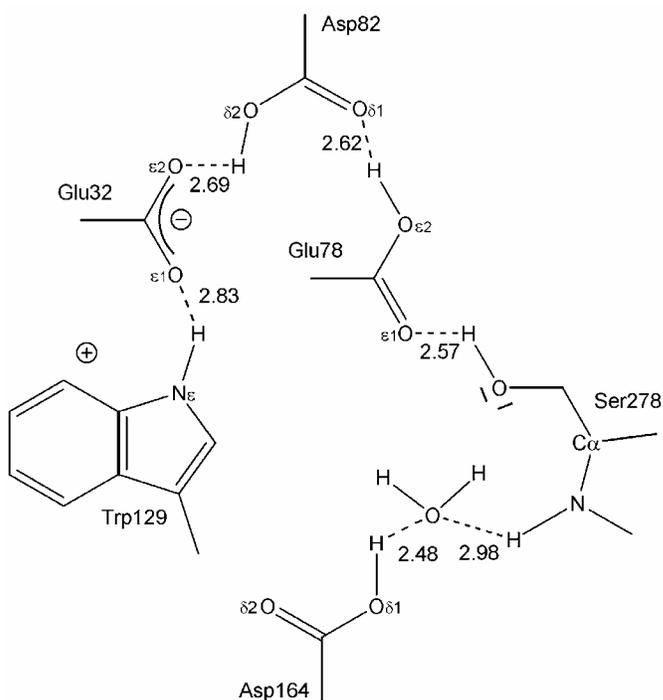


Figure 5.1: Proton network in native kumamolisin. The proteinase residues are shown as thin lines and the O \cdots O and O \cdots N distances indicated in Å are given as observed for the free enzyme at pH 4.5.

During the catalytic process, the Ser278 hydroxyl group becomes nucleophilic upon substrate binding, and releases the O γ proton. The only possible proton acceptor is the carboxyl group of Glu78. We therefore assume that during substrate approach at acidic pH the protons might be located as schematically shown in Figure 5.2. As indicated by the short inter-carboxyl distances, the bridging hydrogens might be placed more symmetrically between the carboxylic group partners, which implicates that the negative charge (in Figures 5.1 and 5.2 tentatively assigned to the Glu32 carboxylate) would rather be delocalized over all three carboxylic groups (Glu78, Asp82 and Glu32).

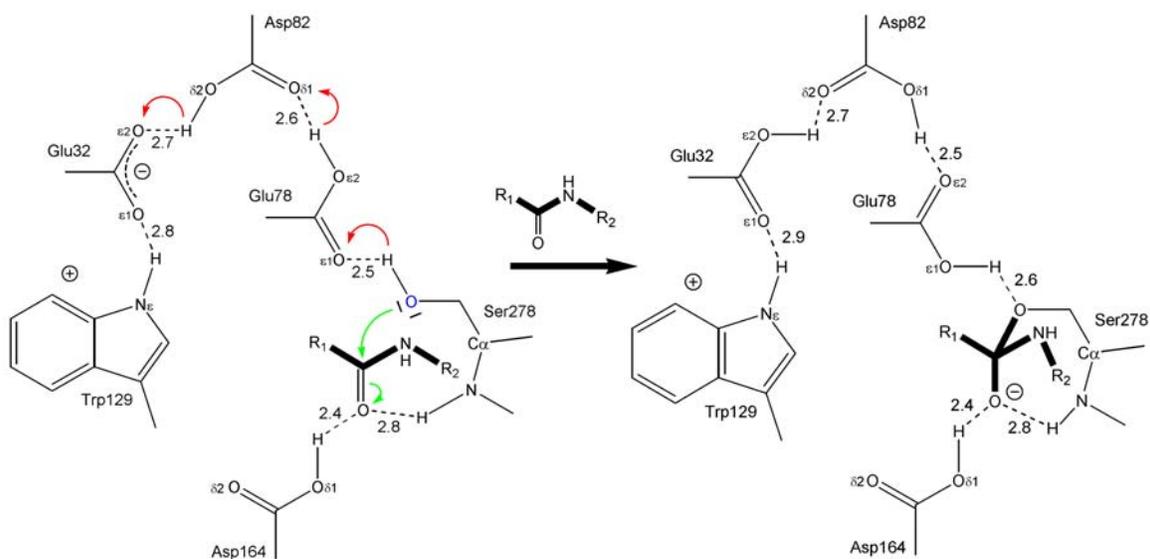


Figure 5.2: Suggested proteolytic mechanism during the first acylation step in kumamolisin. The enzyme residues are shown as thin lines, while the peptide substrate is highlighted by bold lines. The O \cdots O and O \cdots N distances indicated in Å are given as observed for the free enzyme at pH 4.5 / refined at a resolution of 1.4 Å (left side) and for the kumamolisin - Ac-IPF-CHO complex at pH 3.0 / 1.8 Å (right side). The electron movements are represented by green arrows and the probable proton transfer is indicated with a red arrow.

The three-dimensional structures of the kumamolisin enzyme-inhibitor complexes suggest the following acylation scenario (Figure 5.2): upon approach of a polypeptide substrate, the scissile peptide bond will be presented towards the Ser278 O γ with its carbonyl group inserting into the oxyanion hole, fixed by the subsite interactions of flanking peptidyl moieties. This would allow the nucleophilic Ser278 O γ to attack the polarized carbonyl group of the scissile peptide bond and formation of a tetrahedral intermediate after passing the transition state. Simultaneously, the protons would be transferred from the acidic Ser278 oxonium cation to the Glu78 carboxylic group, from the latter to Asp82, and finally from Asp82 to Glu32. In a second acylation step (not shown in Figure 5.2), this proton would be transferred back to the leaving group nitrogen of the substrate, allowing cleavage of the P1-P1' scissile bond, formation of the intermediate acyl ester, and dissociation of the C-terminal substrate fragment. The following deacylation step would occur in an inverted manner, with a fixed water molecule occupying the position of the leaving nitrogen. The critical function of Asp164,

as part of the oxyanion hole, and of Glu78 and Asp82, as proton acceptors after nucleophilic attack, has been confirmed by the results our mutagenesis, where the mutants D164A, E78A and D82A were not autoactivated under acidic conditions, and lack measurable activity upon “activation” by the wild-type enzyme. The importance of residues Glu32 and Trp129 for the catalytic efficiency of kumamolisin is discussed in more detail in the next section (5.2 Kumamolisin Mutants E32A and W129A).

The related bacterial sedolisin (Wlodawer et al., 2001) exhibits Ser287, Glu80, Asp84 and Asp170 as structural homologues of kumamolisin residues Ser278, Glu78, Asp82 and Asp164 (Figure 5.3). Thr34 in sedolisin corresponds topologically to kumamolisin residue Glu32, but is too far away from the carboxylic acid group of Asp84 to allow any direct hydrogen bonding interaction. However, close to its Asp84 carboxylate group, sedolisin also exhibits an internal water molecule, which is placed in a favorable hydrogen bonding distance. This water molecule thus could take over in sedolisin (and possibly also in other serine-carboxyl proteinases) the general base/acid function putatively assigned in kumamolisin to Glu32. More likely, however, the Glu80/Asp84 pair of sedolisin (as well as the equivalent Glu/Asp pairs of the other

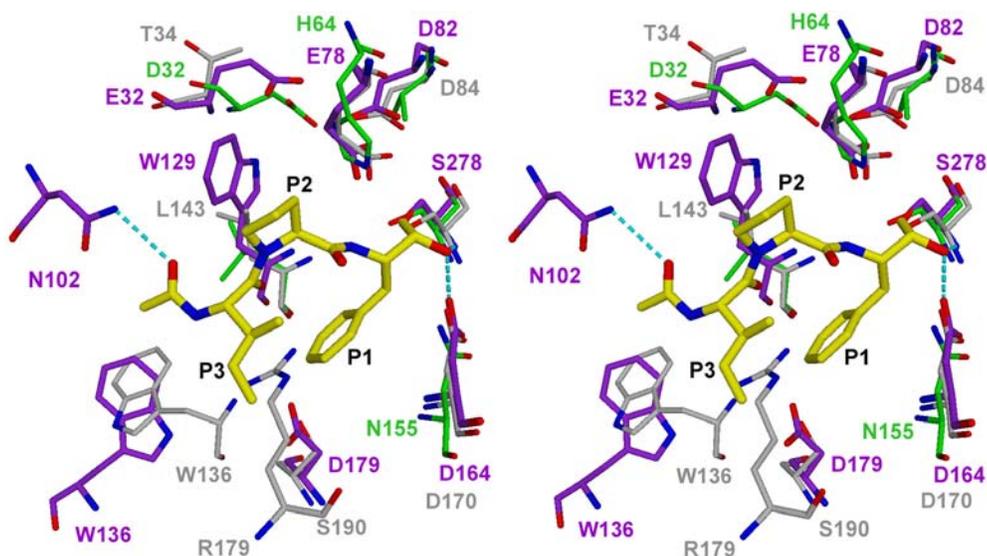


Figure 5.3: Stereo plot of the active site shown in standard orientation. The catalytic residues and surrounding residues of kumamolisin (purple) are superimposed with sedolisin (grey) (Wlodawer et al., 2001) and subtilisin Carlsberg (green) (Bode et al., 1987).

SCPs, including kumamolisin) exists as a stable carboxylic acid-carboxylate pair that could act as a general base/acid. We speculate that the additional Glu32/Trp129 pair, which is not conserved throughout the sedolisin family, might be important in the thermostable kumamolisin for keeping the catalytic apparatus functional also at high temperatures.

Not surprisingly, the similarity of the kumamolisin active site with that of subtilisin, which exhibits a Ser221/His64/Asp32 triad, is lower compared to the analogy with sedolisin (Figure 5.3). Kumamolisin residues Ser278 and Glu78 are topologically and functionally equivalent to Ser221 and His64 of subtilisin. Noteworthy, the topologic equivalent of kumamolisin residue Glu32 is Asp32 in subtilisin. The latter aspartate residue is thus not structurally equivalent to the third residue of the SCPs, but might have similar functions, namely to support the intermediate storage of a proton and to anchor the second active site residue. The Asp32 carboxylate group of subtilisin is hydrogen bonded to the Ser221 O γ nucleophile via the His64 imidazole side chain and, in analogy to the equivalent triad in trypsin, believed to remain in the ionized state during catalysis (Carter and Wells, 1988). The carboxylate-imidazol(ium) pair of subtilisin is an excellent proton acceptor/donor around neutral pH, but becomes not active at low pH values, at which the histidine will not be able to donate a proton. In contrast, the Glu/Asp acidic pair in the active site apparatus of the serine-carboxyl proteinases can activate the primary nucleophilic serine at lower pH values, but might be non-functional in the neutral pH range.

The structure of the substrate binding region of kumamolisin can nicely explain the determined specificity profile of the enzyme as well as differences to the related enzymes sedolisin, XSCP and J-4 (Oda et al., 2000; Ito et al., 1996). The well-defined S1 pocket and the more open S2 subsite of kumamolisin can accommodate medium-sized and small hydrophobic side chains, respectively, in good agreement with the preferred cleavage after leucine and phenylalanine, and with the beneficial effect of alanine and proline at the P2 position of peptide substrates (Oda et al., 2000). The equivalent S1 and S2 pockets of sedolisin, in contrast, are shallower, reflecting the preference for tyrosine and for large residues, respectively. The P3 side chains of bound, extended polypeptide substrates would mainly point away from the kumamolisin surface, while P4 side chains would

extend into a medium-sized hydrophobic pocket, in agreement with the acceptance of P4-leucine, alanine and proline (Oda et al., 2000). The large, mainly hydrophobic S1' pocket is well shaped to accommodate large phenole-like side chains, in agreement with the finding of P1'-Tyr residues at both cleavage sites in the oxidized insulin B-chain. The cleft-like S2' subsite can easily bind long polar side chains such as those of arginines and (after protonation) glutamic acids. Based on systematic studies with octapeptides, the S3' subsite was predicted to be large and to preferentially accommodate bulky aromatic residues, which is consistent with the modeled peptide substrate (Figure 4.10), where P3'-Ile points to a large hydrophobic S3' subsite.

Among family members, the human homolog CLN2/tripeptidyl-peptidase I is of particular interest because of its involvement in a neurodegenerative disease (Sleat et al., 1997). The fully conserved catalytic residues (emphasized with red arrows in Figure 4.7.B) suggest a similar hydrolytic mechanism for the eukaryotic enzyme. Nevertheless, the lack of a CLN2 crystal structure restricts the understanding of its tripeptidyl peptidase activity to a hypothetical comparison with the known sedolisin and kumamolisin structures. In the primed-side direction, the kumamolisin active-site cleft does not show any barrier that would prevent the binding of longer extended substrates, identifying the enzyme as a typical endopeptidase. A preliminary model of CLN2, based on the sequence alignment of Figure 4.7.B, predicts the presence of the Asp132 side chain in the structurally equivalent position of Gly131 in kumamolisin. The carboxylate of this Asp132 would extend out of the active-site cleft, accurately placed to anchor the unsubstituted N-terminus of a binding substrate in such a manner that tripeptides could be released from the N-terminus (Doebber et al., 1978). This prediction is supported by the energy-minimized model of human CLN2 (Wlodawer et al., unpublished), which shows an r.m.s. deviation between the 215 corresponding C α coordinates of kumamolisin and CLN2 of about 1.5 Å using a threshold of 1.3 Å (Figure 5.4). However, this model features a quite free non-primed site of the cleft after the P3 position, leaving unexplained the tripeptidyl aminopeptidase specificity of the human homolog.

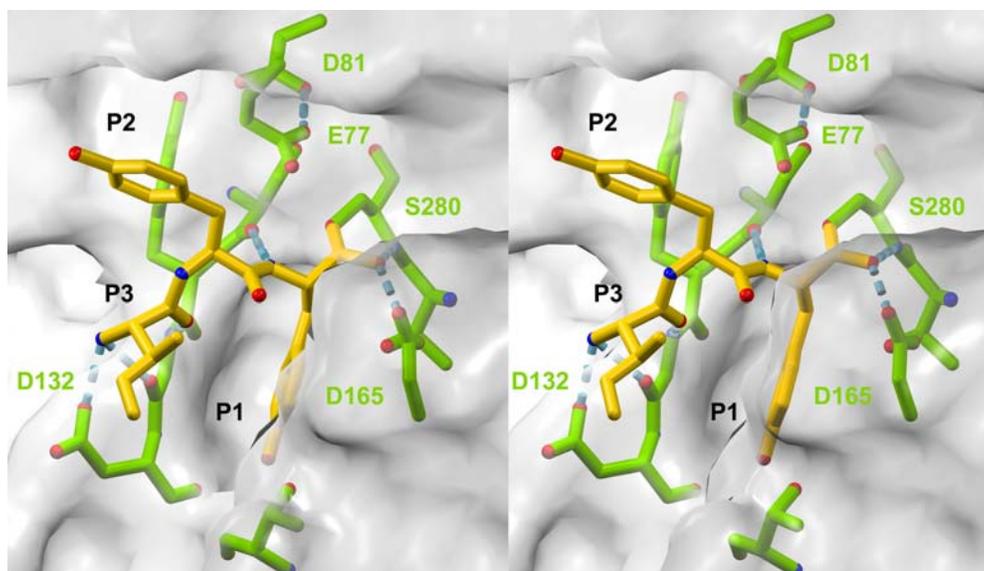


Figure 5.4: Active site model of human tripeptidyl-peptidase I, also called CLN2 (Wlodawer et al., unpublished; theoretical model access code in the PDB: 1R60). The enzyme residues are highlighted in green and superimposed with a semi-transparent Connolly surface. The tripeptidic inhibitor modeled in the active site is shown in yellow and possible hydrogen bonds are indicated by dashed cyan lines.

Bacillus novosp. MN-32 grows under quite acidic conditions at high temperatures. Not surprisingly therefore, kumamolisin is a thermostable enzyme achieving its maximal activity at 70° C and pH 3. Despite intensive investigations to unravel structural elements responsible for thermostability, a general explanation remains elusive (Jaenicke, 2000). Sequence and structure comparisons with sedolisin and subtilisin Carlsberg suggest that the high proline content of kumamolisin (7.5% compared to only 3.5% in sedolisin) might contribute to its thermal stability (Tanner et al., 1996). No significant differences could be detected between sedolisin and kumamolisin concerning the number of intramolecular salt bridges (Lim et al., 1997), the occurrence of labile amino acids (such as cysteines, asparagines, and aspartates) (Demirjian et al., 2001), the surface-to-volume ratio (Chan et al., 1995), and the number of hydrogen bonds (Tanner et al., 1996). Interestingly, an analysis of the hydrogen bonds reveals that kumamolisin and sedolisin possess 7 and 5 Asp/Asp or Glu/Asp pairs, respectively, characterized by carboxylate oxygen atom distances around and below 2.6 Å. These pairs presumably represent stable, low-barrier hydrogen bonds. These particularly stable hydrogen bonds, which are a

characteristic feature of these serine-carboxyl proteinases, constitute structure-stabilizing elements only in acidic environments.

The fully occupied calcium binding site of kumamolisin is equivalent to that of sedolisin (and also anticipated for the other SCPs), but not to those of subtilisin Carlsberg or the thermophile subtilisin-like thermitase (Gros et al., 1991). Our mutagenesis studies identified one of the Ca^{2+} ligands, Asp316, in kumamolisin and its equivalents in sedolisin and XCP (Oyama et al., unpublished) to be essential for the catalytic activity of these enzymes. In agreement with these data, this calcium site remained virtually identical in the crystal structures of kumamolisin resolved at pH 3 and pH 4.5 (crystals A, B, B-IPF and B-IAF), indicating a full Ca^{2+} coordination also by both aspartate carboxylate groups. The free energy needed to keep these carboxylate groups deprotonated at pH 3 must be provided by the stable molecular scaffold. The kumamolisin structure, however, does not explain if the energetic advantage of maintaining the Ca^{2+} site in the SCPs under acidic conditions. An interesting feature strictly conserved in all compared sequences belonging to the sedolisin family, is the presence of a buried ion pair, formed by the side chains of Arg248 and Asp252. The latter residue is conserved in subtilisin as well, where the aspartate participates in the formation of the low-affinity Ca^{2+} -binding site. It is thus probably that this ion pair enhances enzyme stability in the serine-carboxyl family, similar to the role of a Ca^{2+} -binding site in subtilisin.

5.2 Kumamolisin Mutants E32A and W129A

Kumamolisin exhibits, like the other sedolisins, an active site triad consisting of residues Ser278e, Glu78e and Asp82e. It seems to be unique, however, in that the catalytic Asp82e further hydrogen bonds to a Glu32e-Trp129e pair. In this study, we have tried to analyze the possible involvement of the latter residues in catalysis. The replacement of either of these residues by alanine resulted in a significant reduction in the catalytic activity/specificity constant (Table 4), suggesting that both residues are relevant for proton shuttling during catalysis, but are not essential, allowing residual activity. Structure analysis reveals that these mutations induced, in addition to the side chain replacement, perturbations in the local environment. The lack of the Glu32e side chain caused the catalytic Ser278e to adopt, besides the wild-type conformation, a second presumably inactive conformation. The water molecule that replaces the Glu32e carboxylate group might, however, allow a partial restoration of the wild-type conformation (see Figure 4.15).

The active site disturbance caused by the removal of the Trp129 side chain in the W129A mutant have much more dramatic structural effects in the active site as well as in the substrate anchoring segment Ser128e-Gly131e (Figure 4.16). Despite the lack of the indole moiety of the tryptophan, W129A shows a residual activity (Table 4), which is also manifested in the self-activation of the proW129A mutant and the further degradation of the propeptide upon incubation. These evidences suggest that the Glu32e-Trp129e pair is not essential for the activity of kumamolisin. Thus, it can be conclude that in wild-type kumamolisin the Ser-Glu-Asp triad, connected through short low-barrier hydrogen bonds (Frey et al., 1994; Flocco et al., 1995), together with an adjacent proton acceptor/donor (such as the Thr32e-water pair in sedolisin, or a water molecule in the E32A mutant) is sufficient to increase the Ser278e O γ nucleophilicity at low pH, to attract a proton from the reactive Ser278e, and to donate it afterwards to the leaving group of the cleaved substrate. Under the extreme grow conditions of the thermophilic *Bacillus* novosp. MN-32, the Glu32e-Trp129 pair might provide an additional advantage of stabilizing the hydrogen bond network of the catalytic residues.

5.3 Pro-Kumamolisin

Most proteins identified as serine-carboxyl proteinases possess a propeptide, comprising between 170 and 215 residues, which has to be cleaved off for activation of the enzymes. Experimental evidences in the cases of kumamolisin, sedolisin and tripeptidyl-peptidase I, showed that the enzymes are synthesized as a preproform, which is activated under acidic conditions to release the catalytic domain. To study the structural and functional features of the SCPs prodomain, the stable full-length mutant of pro-kumamolisin enzyme, Ser278Ala, was isolated and crystallized.

Despite a very low sequence identity of topologically equivalent residues (about 10%), the core of kumamolisin prodomain resembles, besides those of zinc pro-carboxypeptidases, the prodomains of pro-subtilisins and pro-protein convertases (see Figure 4.18). Also common to these families is the docking of the core prodomain to an equivalent site of the catalytic domain. The structural similarities between the Ser-His-Asp subtilases and the Ser-Glu-Asp sedolisins are a further indication of divergent evolution from a common ancestor. However, the relationships between both proteinase families are even closer. The absence of activity in several propeptide deletion mutants of sedolisin (Oda et al., 1994) has demonstrated that its prodomain is required for efficient refolding, presumably acting like a chaperonin. Similar observations have been reported for subtilisins (Gallagher et al., 1995) and pro-protein convertases (Bhattacharjya et al., 2001). In the absence of more detailed kinetic measurements, we can only presume that, similar to the latter cases, the prodomain in kumamolisin stabilizes a folding intermediate of the catalytic domain, by interacting with helices h4 and h5, and becomes essential to achieve proper three-dimensional structure. In the subtilisins, such an intermediate is required in the latter folding steps, to overcome a high energy barrier considered to be the price to be paid for the extreme stability of the mature enzyme (Eder and Fersht, 1995). In the pro-subtilisins, this refolding resistance of the mature peptidase has been shown to be associated with the first calcium site of subtilisin (Strausberg et al., 1995) (see Figure 4.7.A). Therefore, an equivalent association of the refolding kinetics with formation of the single calcium site in the sedolisins will have to be investigated in future.

In addition, the current pro-kumamolisin structure directly proves that the catalytic domain has completely adopted a mature-like conformation already in the proform

(Figure 4.17), and that the linker between the pro and the catalytic domains extends through the active-site cleft, with the His171p↓Phe172p peptide bond close to of Ser278 (Figure 4.20). This structural arrangement strongly suggests that the first activation cleavage will occur auto-catalytically between P1-His171p and P1'-Phe172p, via an intra-molecular attack. As the linker chain and the active site cleft are quite complementary, the structure mimics an efficient kumamolisin–substrate complex, with a typical substrate cleavage conformation, presumably not disturbed by the Ser278→Ala replacement (Carter and Wells, 1988). It is noteworthy that the linker sequence around the scissile peptide bond (shown in Figure 5.5) agrees well with the known cleavage preference of kumamolisin (Oda et al., 2000). Besides the lack of the Ser278e O γ atom and the neutral pH value of the crystallization buffer, the presence of the P3-Arg169p linker residue might increase the extreme stability of the S278A pro-kumamolisin mutant. The guanidyl group of P3-Arg169p attracts the Asp164e carboxylate side chain, preventing the formation of a functional oxyanion hole by substrate-assisted inactivation. In an acidic environment, protonation of the latter carboxylate group would cause the disruption of this salt bridge, rendering the catalytic apparatus in pro-kumamolisin more active. Thus, the P3-Arg might be a built-in switch, delaying self-activation of pro-kumamolisin until secretion into the acidic medium. Noteworthy, this arginine residue in the P3 position is not conserved among the propeptides of sedolisins.

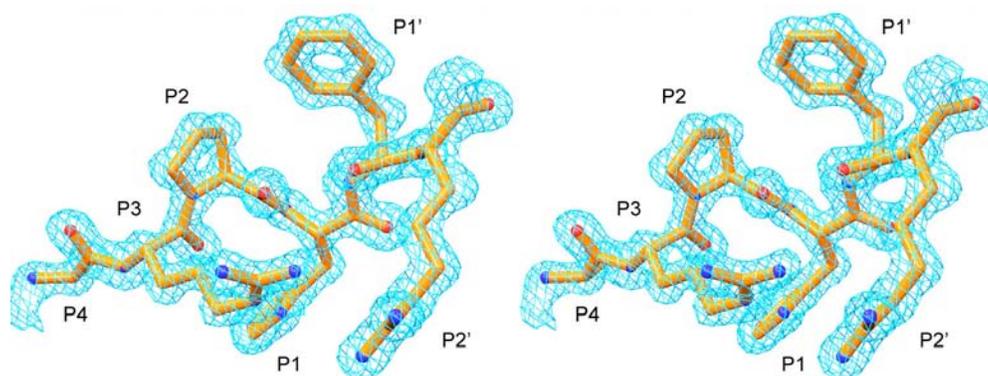


Figure 5.5: Stereo plot of the bound linker peptide of pro-kumamolisin. Central part of the linker peptide (orange) around the scissile peptide bond (P4 to P2') superimposed with its difference electron density omit map (blue) contoured at 3 σ .

Presumably, the first cleavage product upon pro-kumamolisin processing will be the inactive complex between the N-terminally elongated catalytic kumamolisin domain and its propeptide down to His171p. Co-elution in the cleaved pro-W129A mutant in gel filtration experiments (data not shown) proves that the propeptide and the catalytic domain of the initially cleaved pro-kumamolisin form a complex, with the propeptide presumably acting as a competitive inhibitor, analogous to the subtilisins (Ikemura and Inouye, 1988), and to furin (see Anderson et al., 2002). This hetero-dimeric complex will dissociate releasing the active, N-terminally elongated enzyme, perhaps facilitated by acidification or by additional cleavage(s) within the compact prodomain as observed in other subtilases. Because mature kumamolisin is found to start with Ala1e, the exposed, flexible linker segment Ser188p-Ala1e must be further truncated. This second cleavage is presumably not catalyzed by active kumamolisin, as the sequence around the Ser188p↓Ala1e cleavage site does not match kumamolisin specificity (Figure 4.19).

In conclusion: The pro-kumamolisin crystal structure is the first of an intact pro-subtilase, and therefore presumably a structural prototype for the whole subtilase superfamily. Thus, this structure has wider implications not only for the microbial and mammalian sedolisins, but for the subtilisins as well including furin and other pro-protein convertases.

5.4 Peptidyl-Dipeptidase Dcp from *E. coli*

Classification of peptidyl-dipeptidase Dcp has been very controversial, since the discovery of the enzyme on 1978 (Deutch and Soffer, 1978). Initially, the similarity of the peptidyl-dipeptidase activity with the human angiotensin-converting enzyme led to the classification of both metallopeptidases under the same entry, as dipeptidyl carboxypeptidases. Subsequently, the enzymes were separated and joined again depending on the new reported properties, until Dcp and ACE were definitely included in different families (M3(A) and M2, respectively) based on sequencing studies (Hamilton and Miller, 1992; Henrich et al., 1993). The crystal structure of Dcp confirms the final classification showing an overall fold closer related to another member of the M3(A) family, neurolysin, than to ACE. However, and despite the lack of sequence similarity, a major part of the topological elements are conserved between Dcp and ACE (Figure 4.28). This indicates a case of divergent evolution from a common ancestor between the families of peptidyl-dipeptidase Dcp and angiotensin-converting enzyme.

A detailed comparison between the structures of the inhibitor-bound Dcp and the native homologue neurolysin reveals an important difference with respect to the distance separating the two domains, which achieve a maximal value at the lower end of 18 Å (Figure 4.29). While, in neurolysin, the two domains are open and expose a wide cleft, the structure of Dcp in complex with inhibitor N-1450 shows a closed conformation of domain I and II that buries the active site cleft in the core of the molecule. This shift of one domain relative to the other suggests a large inhibitor-dependant hinge-bending movement ($\sim 16^\circ$) that would cause the deep open cleft in the native form of the enzyme to close around a substrate/inhibitor during catalysis/inhibition, respectively (Figures 4.30.A and B). The ligand-dependending rearrangement between the two domains would mimic the action of a closing clam shell as schematically shown in Figure 5.6, with the hinge situated at the top of the molecule (orientation as in Figures 4.26 and 4.29). Upon ligand binding, the loop segment Pro591- Tyr611 of Dcp presumably changes its conformation and closes over the active site due to the hydrogen-bond interactions

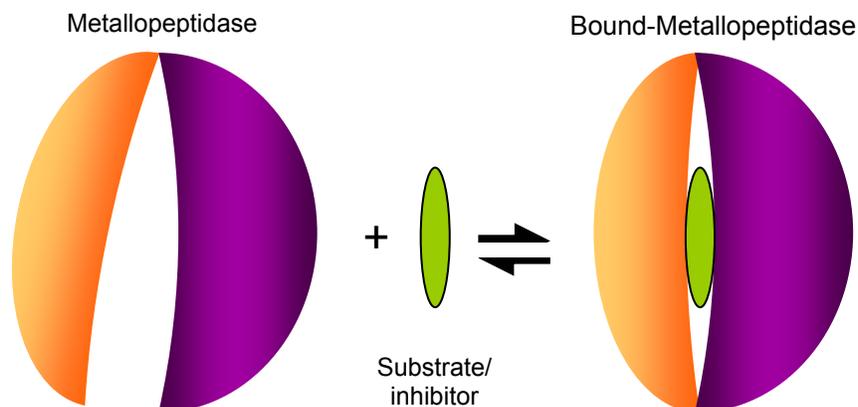


Figure 5.6: Schematic representation of the suggested inhibitor-dependant hinge-bending movement for Dcp and neurolysin. The two domains of the molecule are shown in purple (domain I) and in orange (domain II), while the substrate or inhibitor is indicated in green.

between the substrate or inhibitor and six residues arising from the loop. Two important facts support the proposed mechanism as a general catalytic process for enzymes of the M3(A) family. On one hand, four of these six hydrogen-bonding residues (His601, Tyr607, Tyr6011, and Tyr614) are conserved in neurolysin (Figure 5.7), whereas the non-conserved Arg593 and Tyr594 bind the C-terminal carboxylate of the cleaved dipeptide, in primed site of Dcp (Figure 4.34). Since neurolysin is an endopeptidase instead of dipeptidyl-peptidase, it is not surprising that these two residues at the prime site are not conserved in neurolysin. On the other hand, the loop segment Gly600-Tyr610 shows a significant inherent flexibility in native neurolysin, indicated by particularly high thermal factors (average all atoms *B* value of 48.7). In contrast, in the inhibitor-bound Dcp the corresponding loop is perfectly ordered with thermal factors for all atoms (*B* value of 25.8) below the average for the whole molecule (see Table 7).

Experimental evidences reported recently also reinforce the idea of a hinge-bending mechanism in M3(A) family members, upon ligand binding. Towler and coworkers observed similar rearrangements between the two subdomains in the crystal structures of native and inhibitor-bound ACE2 (Towler et al., 2004). ACE2 possesses only carboxypeptidase activity and is found to be an essential regulator of heart function (Crackower et al., 2002), as well as a functional receptor for coronavirus, which is linked

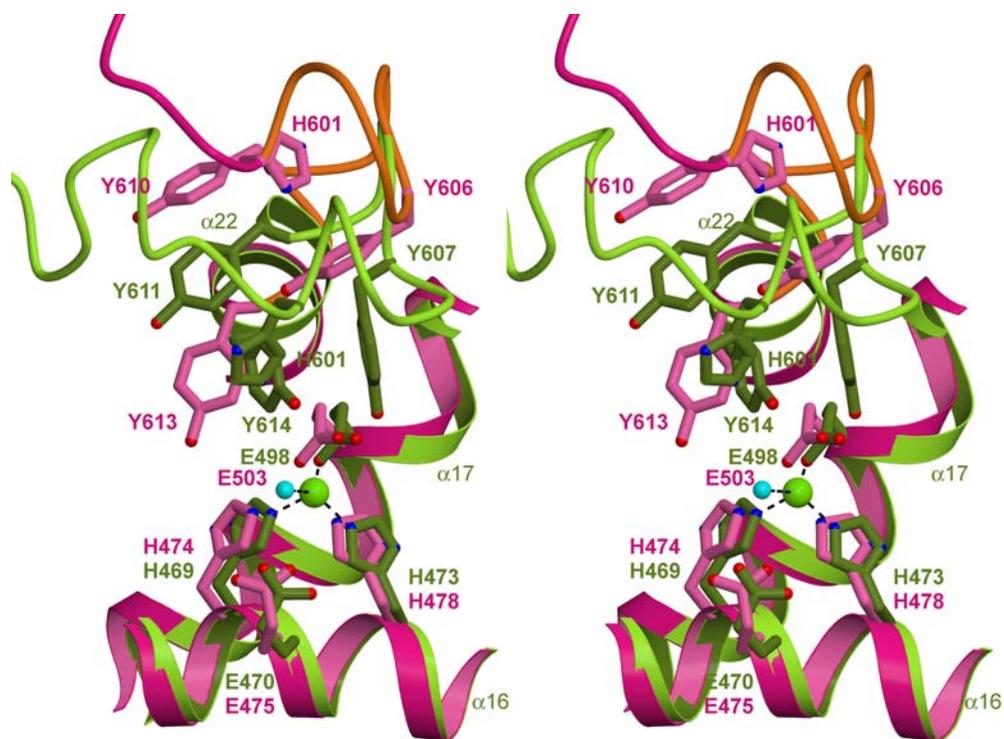


Figure 5.7: Superposition of the active sites of inhibitor-bound Dcp and free neurolysin. Important residues of the active site and the loop between $\alpha 21$ (domain II) and $\alpha 22$ (domain II) are indicated as ball-and-stick rendering for Dcp (green) and neurolysin (pink). In neurolysin, the flexible loop segment (601-612) is colored in orange. Significant secondary structure elements are marked for both enzymes. The zinc ion is indicated by a green sphere and the coordination sphere is shown as black dashed lines. The position the fourth ligand, a solvent molecule in the native structure of neurolysin and the C-terminal carboxylate in Dcp, is indicated by a small cyan sphere.

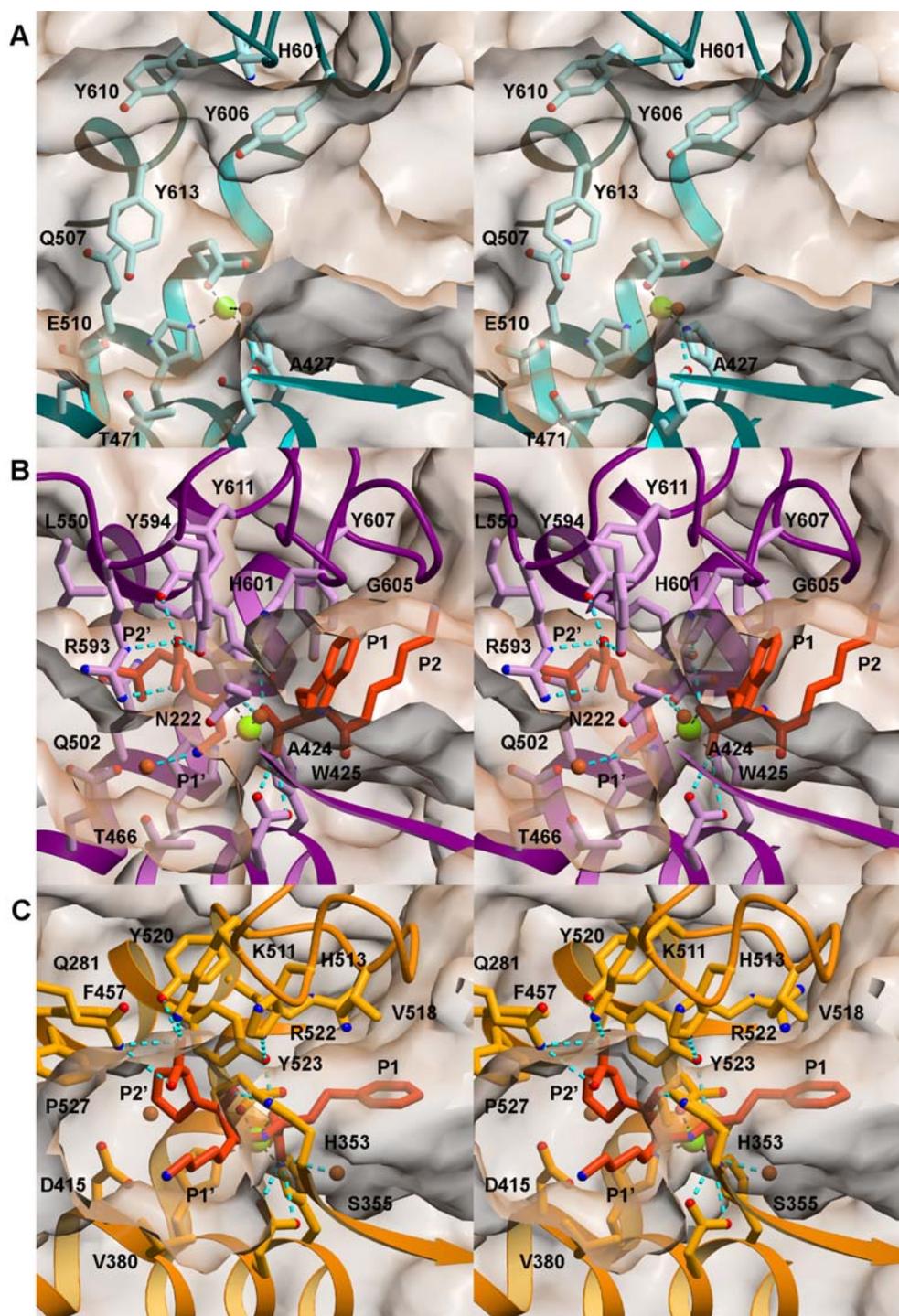
to the severe acute respiratory syndrome (SARS) (Li et al., 2003; Xiao et al., 2003). This ACE-related enzyme with an overall fold similar to Dcp, shows also an inhibitor-induced displacement of the subdomains of $\sim 16^\circ$ with a hinge location structurally equivalent to the proposed for the Dcp family. Furthermore the evidences observed in ACE2 indicated that the hinge-bending mechanism is presumably applicable to ACE and its related metallopeptidases. However, the fact that the native structures of human (Natesh et al., 2003) and *Drosophila* ACE (Kim et al., 2003) show the same close conformation as their respective inhibitor-bound structures would argue against the proposed mechanism. In fact, both enzymes were crystallized under extreme conditions: human ACE in sodium acetate buffer, pH 4.7, and *Drosophila* ACE in 7.2 M sodium formate buffer. In human

ACE, the “native” structure shows an acetate ion that coordinates the catalytic zinc mimicking the C-terminal carboxylate of the inhibitor (Natesh et al., 2003). The side chain of ACE residue Tyr532 can then be fixed by a hydrogen bond to one acetate oxygen and to one imidazole nitrogen of His513, favoring the close conformation of the protein (Figure 5.8.C). Similarly, in the crystal structure of “native” *Drosophila* ACE, the two water molecules coordinating the zinc ion are more probably a formate molecule (due to the high salt concentration in the crystallization buffer, 7.2 M), which might have been misinterpreted at the obtained resolution (2.6 Å). Like in human ACE, this ligand could induce the closed form of the enzyme. In contrast, the two reported structures in the open conformation, neurolysin (Brown et al., 2001) and ACE2 (Towler et al., 2004), display only one water molecule in the coordination sphere of the catalytic zinc, as usual for the known gluzincins. Therefore, we believe that the structures of human and *Drosophila* ACE represent ligand-bound conformations, and that their close conformational state is consistent with the proposed ligand-dependent hinge-bending mechanism for enzymes of families M3(A) and M2.

Other subdomain hinge-bending motions have also been observed for other zinc metallopeptidases, such as thermolysin and related enzymes belonging to the bacterial neutral protease family, M4 (Holland et al., 1992; Grams et al., 1996). The largest reported ligand-dependent movement for this metallopeptidases is a 14° hinge-bending subdomain motion demonstrated for *Pseudomonas aeruginosa* elastase (open versus closed) that resulted in a ~2-Å movement to close the N-terminal/C-terminal subdomain gap (Holland et al., 1992). Domain closure motions in proteins are a common mechanism for the positioning of critical groups around substrates and inhibitors (Gerstein et al., 1994; Gerstein and Krebs, 1998; Teague, 2003) and also for the trapping of substrate and reaction intermediates (Knowles, 1991).

Figure 5.8: (Next page) Comparison of the Dcp, neurolysin and ACE active sites. A stereo plot representation is shown for **A**) native neurolysin in cyan (Brown et al., 2001), **B**) Dcp (purple) in complex with N-1450 (dark orange) and **C**) ACE (pale orange) in complex with lisinopril (dark orange) (Natesh et al., 2003). The ball-and-stick models are superimposed with their corresponding half-transparent Connolly surface, which has been sliced, for clarity. Zinc ions and water molecules are represented by green and

brown spheres, respectively. Zinc coordination and hydrogen bonds are indicated by black and cyan dashed lines.



The strong structural homology of the active site of inhibitor-bound Dcp to catalytic motifs of evolutionarily related enzymes in metallopeptidases clans MA and MB, such as thermolysin (Grams et al., 1996) and astacin (Holden et al., 1987), indicates that many of these structural elements play similar catalytic roles in Dcp. These similarities have led to a suggested catalytic mechanism for peptidyl-dipeptidase Dcp that has many features in common with mechanisms proposed for other HEXXH zinc metallopeptidase clan members (Grams et al., 1996; Matthews, 1988; Arndt et al., 2002). In the absence of substrate, Dcp would show a wide active site cleft expected to be quite similar to the cleft of native neurolysin (Figure 5.8.A). The initial step for catalysis reveals probable enzyme-substrate (ES) complex formation that induces the hinge-bending movement of the two domains to close the active site cleft bringing important residues into a functional position for catalysis. Following the ES complex rearrangement, the nucleophile zinc-bound water attacks the carbonyl group of the scissile amide bond to form a tetrahedral intermediate, resulting in transfer of a proton from the attacking water to Glu470 (Figures 2.7 and 4.32). Simultaneously, a proton is transferred from the later glutamic acid, Glu470, to the leaving nitrogen atom of the P1' residue, to which remains hydrogen bonded. The modeled active site of Dcp bound to lisinopril inhibitor, which mimics the tetrahedral transition state (Figure 5.9), suggests that the P1' nitrogen with sp^3

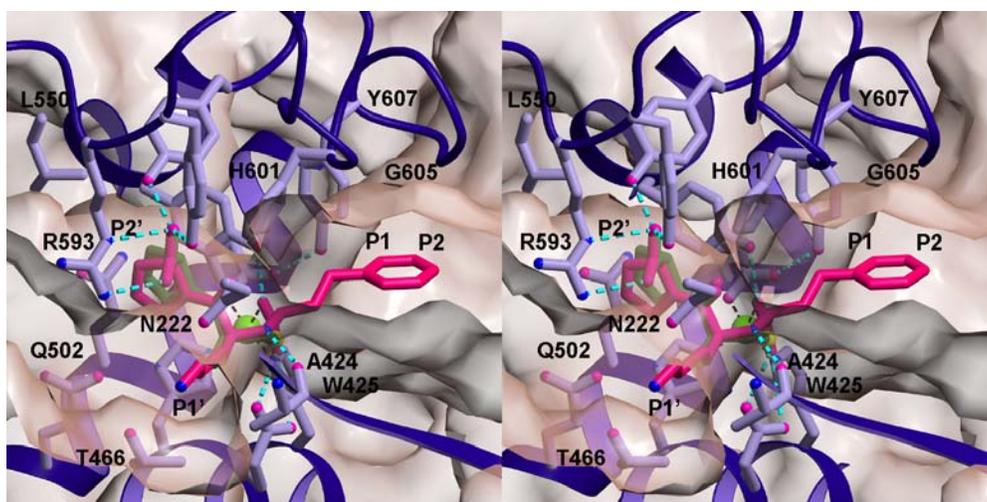


Figure 5.9: Modeled Dcp inhibition mode of lisinopril (pink) and captopril (olive green). The enzyme is shown as a ball-and-stick model in blue superimposed with a semi-transparent Connolly surface.

hybridization could be stabilized by an additional hydrogen bond from the main chain oxygen of Ala424. On the other hand, the tetrahedral intermediate stabilization most likely occurs through hydrogen bonding from the phenolic group of Tyr614 (Figure 4.23). The tetrahedral intermediate, presumably, decomposes to the products by breaking of the amide C-N bond together with abstraction of a proton from the tetrahedral carboxylate by the emerging free nitrogen of the cleaved dipeptide (Figure 5.8.B). In the non-prime side a new C-terminal carboxylate is formed and the substrate can be released while the enzyme adopts the open conformational state.

Although Dcp and ACE possess dipeptidyl-peptidase activity and broad substrate specificity, a detailed comparison of both active site clefts reveals significant differences in the formation of their subsites (Figures 5.8.B and C). Starting in the non-prime side, the cleft observed in human ACE is narrower than in peptidyl-dipeptidase Dcp. The P2 position of Dcp extends along the Tys607 side chain, whereas in ACE the S2 pocket is limited by the basic Arg522 side chain in agreement with the preference of the human enzyme for small residues in P2, such as serine (in bradykinin: RPPGFSP↓FR) or proline (in angiotensin I: DDVYIHPF↓HL) (Ehlers and Riordan, 1990). The S1 cavity, which possesses a rather hydrophobic character in both peptidases, is formed on the left side by a conserved histidine residue (His601 in Dcp and His513 in ACE). In contrast, the top of S1 in Dcp, Gly605, is substituted in ACE by the structurally equivalent Val518, forming a lower pocket (Figures 5.8.B and C). However, ACE accepts also bulky amino acid residues in the P1 position such as phenyl alanine or even tryptophan, since N-1450 is reported to be an ACE inhibitor, as well (Kohama et al., 1988). To determine the location of a P1' side chain in the active site cleft of Dcp, the lisinopril inhibitor was modeled based on its binding mode in the human homolog enzyme (Figure 5.9). Although the S1' cavity in Dcp is less open than in ACE due to the presence of the Asn222 side chain, the large hydrophobic side chain of P1'-Lys fits well in the corresponding S1' pockets of both enzymes (Figures 5.8.B and C and 5.9). Finally, the S2' subsite is the most similar pocket between the two peptidyl-dipeptidases. In both cases, the C-terminal carboxylate hydrogen bonds to three side chain residues, one of which, Tyr611 is conserved in ACE (Tyr520). Dcp Tyr594 is replaced by Lys511 in ACE, whereas Gln218 in the human enzyme substitutes the Arg593 in Dcp. The surroundings forming the S2' cavity are

rather hydrophobic and the pocket is able to accept middle to large side chains in the P2' position.

The crystal structure of peptidyl-dipeptidase Dcp shows for the first time the structural features of a ligand-bound enzyme belonging to family M3(A). The current work provides a detailed picture of the active site and the inhibitor/substrate binding mode of Dcp. Comparison with the native homologue neurolysin, strongly suggest a ligand-dependant hinge-bending movement as the catalytic mechanism for members of family M3(A). Furthermore, the resemblance of Dcp and neurolysin to ACE has shown a surprising structural conservation between these metallopeptidases and points to a similar hinge-bending domains motion during catalysis of ACE and related enzymes.

6 Abbreviations

Å	Ångström; $1\text{Å} = 10^{-10}\text{ m}$
AA	acrylamide
amp	ampicillin
APS	ammonium persulfate
AS	ammonium sulfate
bp	basepair
CCD	Charge-coupled device
CD	circular dichroism
Da	Dalton; $1\text{ Da} = 1\text{ g}\cdot\text{mol}^{-1}$
DCI	3,4-dichloroisocoumarin
DFP	diisopropyl phosphofluoridate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
ES	enzyme-substrate
<i>fom</i>	figure of merit
FPLC	Fast Performance Liquid Chromatography
h	hour
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
HPLC	High Performance Liquid Chromatography
IPTG	isopropyl-beta-D-thiogalactopyranoside
LINCL	classical late-infantile neuronal ceroid lipofuscinosis
LB	Luria-Bertani
mA	miliAmpers
MAD	Multiple-wavelength Anomalous Dispersion
MIR	Multiple Isomorphous Replacement
MOPS	3-N-morpholinopropane-sulfonic acid
MPD	2-methyl-2,4-pentanediol

Abbreviations

MR	Molecular Replacement
NMR	nuclear magnetic resonance
NMM	New Minimal Media
PAGE	polyacrylamide gel electrophoresis
PVDF	polyvinylidene fluoride
r.m.s.(d.)	root-mean-square (deviation)
rpm	revolutions per minute
RT	room temperature
SCP	serine-carboxyl proteinase
SDS	sodium dodecyl sulfate
TAE	tris-acetic-EDTA
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	α,α,α -trishydroxymethyl aminomethane
TTI	tripeptidyl-peptidase I
U	Unit (enzyme unit)
UV	ultraviolet
V	volts
v/v	volume per volume
w/v	weight per volume

7 References

- Anderson, E. D., Molloy, S. S., Jean, F., Fei, H., Shimamura, S., and Thomas, G. (2002). The ordered and compartment-specific autoproteolytic removal of the furin intramolecular chaperone is required for enzyme activation. *Journal of Biological Chemistry* *277*, 12879-12890.
- Arndt, J. W., Hao, B., Ramakrishnan, V., Cheng, T., Chan, S. I., and Chan, M. K. (2002). Crystal structure of a novel carboxypeptidase from the hyperthermophilic archaeon *pyrococcus furiosus*. *Structure* *10*, 215-224.
- Auld, D. S. (1995). Removal and replacement of metal-ions in metallopeptidases. In *Proteolytic enzymes: Aspartic and metallo peptidases*, pp. 228-242.
- Auld, D. S. (1997). Zinc catalysis in metalloproteases. In *Metal sites in proteins and models*, pp. 29-50.
- Babe, L. M., and Craik, C. S. (1997). Viral proteases: Evolution of diverse structural motifs to optimize function. *Cell* *91*, 427-430.
- Barton, G. J. (1993). Alscript - a tool to format multiple sequence alignments. *Protein Engineering* *6*, 37-40.
- Becker, J. W., Marcy, A. I., Rokosz, L. L., Axel, M. G., Burbaum, J. J., Fitzgerald, P. M. D., Cameron, P. M., Esser, C. K., Hagmann, W. K., Hermes, J. D., and Springer, J. P. (1995). Stromelysin-1 - 3-dimensional structure of the inhibited catalytic domain and of the c-truncated proenzyme. *Protein Science* *4*, 1966-1976.
- Betzl, C., Bellemann, M., Pal, G. P., Bajorath, J., Saenger, W., and Wilson, K. S. (1988a). X-ray and model-building studies on the specificity of the active-site of proteinase-k. *Proteins-Structure Function and Genetics* *4*, 157-164.
- Betzl, C., Pal, G. P., and Saenger, W. (1988b). 3-dimensional structure of proteinase-k at 0.15-nm resolution. *European Journal of Biochemistry* *178*, 155-171.
- Bhattacharjya, S., Xu, P., Xiang, H., Chretien, M., Seidah, N. G., and Ni, F. (2001). Ph-induced conformational transitions of a molten-globule-like state of the inhibitory prodomain of furin: Implications for zymogen activation. *Protein Science* *10*, 934-942.
- Blundell, T. L., and Johnson, L. N. (1976). *Protein crystallography* (London, Academic Press Inc.).
- Bode, W., Fehllhammer, H., and Huber, R. (1976). Crystal-structure of bovine trypsinogen at 1.8 a resolution .1. Data-collection, application of patterson search techniques and preliminary structural interpretation. *Journal of Molecular Biology* *106*, 325-335.

References

Bode, W., and Huber, R. (1992). Natural protein proteinase-inhibitors and their interaction with proteinases. *European Journal of Biochemistry* 204, 433-451.

Bode, W., Papamokos, E., and Musil, D. (1987). The high-resolution x-ray crystal-structure of the complex formed between subtilisin carlsberg and eglin-c, an elastase inhibitor from the leech *hirudo-medicinalis* - structural- analysis, subtilisin structure and interface geometry .2. *European Journal of Biochemistry* 166, 673-692.

Bode, W., Papamokos, E., Musil, D., Seemueller, U., and Fritz, H. (1986). Refined 1.2-Å crystal-structure of the complex formed between subtilisin carlsberg and the inhibitor eglin-c - molecular-structure of eglin and its detailed interaction with subtilisin. *Embo Journal* 5, 813.

Boldin, M. P., Goncharov, T. M., Goltsev, Y. V., and Wallach, D. (1996). Involvement of mach, a novel mort1/fadd-interacting protease, in fas/apo-1- and tnfr-receptor-induced cell death. *Cell* 85, 803-815.

Brown, C. K., Madauss, K., Lian, W., Beck, M. R., Tolbert, W. D., and Rodgers, D. W. (2001). Structure of neurolysin reveals a deep channel that limits substrate access. *Proceedings of the National Academy of Sciences of the United States of America* 98, 3127-3132.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998). Crystallography & nmr system: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D-Biological Crystallography* 54, 905-921.

Buchler, M., Tisljar, U., and Wolf, D. H. (1994). Proteinase yscd (oligopeptidase yscd) - structure, function and relationship of the yeast enzyme with mammalian thimet oligopeptidase (metalloendopeptidase, ep-24.15). *European Journal of Biochemistry* 219, 627-639.

Carter, P., and Wells, J. A. (1988). Dissecting the catalytic triad of a serine protease. *Nature* 332, 564-568.

Chan, M. K., Mukund, S., Kletzin, A., Adams, M. W. W., and Rees, D. C. (1995). Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. *Science* 267, 1463-1469.

Chen, P., and Hochstrasser, M. (1996). Autocatalytic subunit processing couples active site formation in the 20S proteasome to completion of assembly. *Cell* 86, 961-972.

Chen, P., Tsuge, H., Almassy, R. J., Gribskov, C. L., Katoh, S., Vanderpool, D. L., Margosiak, S. A., Pinko, C., Matthews, D. A., and Kan, C. C. (1996). Structure of the human cytomegalovirus protease catalytic domain reveals a novel serine protease fold and catalytic triad. *Cell* 86, 835-843.

- Coll, M., Guasch, A., Aviles, F. X., and Huber, R. (1991). 3-dimensional structure of porcine procarboxypeptidase-b - a structural basis of its inactivity. *Embo Journal* *10*, 1-9.
- Comellas-Bigler, M., Fuentes-Prior, P., Maskos, K., Huber, R., Oyama, H., Uchida, K., Dunn, B. M., Oda, K., and Bode, W. (2002). The 1.4 angstrom crystal structure of kumamolysin: A thermostable serine-carboxyl-type proteinase. *Structure* *10*, 865-876.
- Conlin, C. A., and Miller, C. G. (1995). Dipeptidyl carboxypeptidase and oligopeptidase-a from *escherichia-coli* and *salmonella-typhimurium*. In *Proteolytic enzymes: Aspartic and metallo peptidases*, pp. 567-579.
- Cowtan, K. D. (1994). 'dm': An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* *31*, 34-38.
- Crackower, M. A., Sarao, R., Oudit, G. Y., Yagil, C., Kozieradzki, I., Scanga, S. E., Oliveira-dos-Santos, A. J., da Costa, J., Zhang, L. Y., Pei, Y., *et al.* (2002). Angiotensin-converting enzyme 2 is an essential regulator of heart function. *Nature* *417*, 822-828.
- Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A. C., and Mort, J. S. (1996). Structure of rat procathepsin b: Model for inhibition of cysteine protease activity by the proregion. *Structure* *4*, 405-416.
- Dalbey, R. E., Lively, M. O., Bron, S., and VanDijl, J. M. (1997). The chemistry and enzymology of the type i signal peptidases. *Protein Science* *6*, 1129-1138.
- Dalbey, R. E., and Vonheijne, G. (1992). Signal peptidases in prokaryotes and eukaryotes - a new protease family. *Trends in Biochemical Sciences* *17*, 474-478.
- Das, M., and Soffer, R. L. (1976). Pulmonary angiotensin-converting enzyme antienzyme antibody. *Biochemistry* *15*, 5088-5094.
- Dauch, P., Vincent, J. P., and Checler, F. (1995). Molecular-cloning and expression of rat-brain endopeptidase-3.4.24.16. *Journal of Biological Chemistry* *270*, 27266-27271.
- Davie, E. W., Fujikawa, K., and Kisiel, W. (1991). The coagulation cascade - initiation, maintenance, and regulation. *Biochemistry* *30*, 10363-10370.
- Demirjian, D. C., Moris-Varas, F., and Cassidy, C. S. (2001). Enzymes from extremophiles. *Current Opinion in Chemical Biology* *5*, 144-151.
- Deutch, C. E., and Soffer, R. L. (1978). *Escherichia-coli* mutants defective in dipeptidyl carboxypeptidase. *Proceedings of the National Academy of Sciences of the United States of America* *75*, 5998-6001.
- Doebber, T. W., Divor, A. R., and Ellis, S. (1978). Identification of a tripeptidyl aminopeptidase in anterior- pituitary gland - effect on chemical and biological properties of rat and bovine growth hormones. *Endocrinology* *103*, 1794-1804.

References

- Drenth, J. (1994). Principles of protein x-ray crystallography (Heidelberg, Springer Verlag).
- Eder, J., and Fersht, A. R. (1995). Pro-sequence-assisted protein-folding. *Molecular Microbiology* 16, 609-614.
- Eder, J., Rheinneckner, M., and Fersht, A. R. (1993). Folding of subtilisin bpn' - role of the pro-sequence. *Journal of Molecular Biology* 233, 293-304.
- Edman, P., and Henschen, A. (1975). Sequence determination. In Protein sequence determination (Heidelberg, Springer Verlag), pp. 232-279.
- Ehlers, M. R. W., and Riordan, J. F. (1989). Angiotensin-converting enzyme - new concepts concerning its biological role. *Biochemistry* 28, 5311-5318.
- Ehlers, M. R. W., and Riordan, J. F. (1991). Angiotensin-converting enzyme - zinc-binding and inhibitor-binding stoichiometries of the somatic and testis isozymes. *Biochemistry* 30, 7118-7126.
- Esnouf, R. M. (1997). An extensively modified version of molscript that includes greatly enhanced coloring capabilities. *Journal of Molecular Graphics & Modelling* 15, 132-&.
- Esnouf, R. M. (1999). Further additions to molscript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallographica Section D-Biological Crystallography* 55, 938-940.
- Estebanez-Perpina, E., Bayes, A., Vendrell, J., Jongsma, M. A., Bown, D. P., Gatehouse, J. A., Huber, R., Bode, W., Aviles, F. X., and Reverter, D. (2001). Crystal structure of a novel mid-gut procarboxypeptidase from the cotton pest *helicoverpa armigera*. *Journal of Molecular Biology* 313, 629-638.
- Ezaki, J., Takeda-Ezaki, M., Oda, K., and Kominami, E. (2000). Characterization of endopeptidase activity of tripeptidyl peptidase-i/cln2 protein which is deficient in classical late infantile neuronal ceroid lipofuscinosis. *Biochemical and Biophysical Research Communications* 268, 904-908.
- Ezaki, J., Tanida, I., Kanehagi, N., and Kominami, E. (1999). A lysosomal proteinase, the late infantile neuronal ceroid lipofuscinosis gene (cln2) product, is essential for degradation of a hydrophobic protein, the subunit c of atp synthase. *Journal of Neurochemistry* 72, 2573-2582.
- Flocco, M. W., and Mowbray, S. L. (1995). Strange bedfellows - interactions between acidic side-chains in proteins. *Journal of Molecular Biology* 254, 96-105.
- Frey, P. A., Whitt, S. A., and Tobin, J. B. (1994). A low-barrier hydrogen-bond in the catalytic triad of serine proteases. *Science* 264, 1927-1930.

- Gallagher, T., Gilliland, G., Wang, L., and Bryan, P. (1995). The prosegment-subtilisin bpn complex - crystal-structure of a specific foldase. *Structure* 3, 907-914.
- Gerstein, M., and Krebs, W. (1998). A database of macromolecular motions. *Nucleic Acids Research* 26, 4280-4290.
- Gerstein, M., Lesk, A. M., and Chothia, C. (1994). Structural mechanisms for domain movements in proteins. *Biochemistry* 33, 6739-6749.
- Gill, S. C., and Vonhippel, P. H. (1989). Calculation of protein extinction coefficients from amino-acid sequence data. *Analytical Biochemistry* 182, 319-326.
- Goedert, M. (1984). Neurotensin - a status-report. *Trends in Neurosciences* 7, 3-5.
- Gomis-Ruth, F. X., Companys, V., Qian, Y., Fricker, L. D., Vendrell, J., Aviles, F. X., and Coll, M. (1999). Crystal structure of avian carboxypeptidase d domain ii: A prototype for the regulatory metallo-carboxypeptidase subfamily. *Embo Journal* 18, 5817-5826.
- Grams, F., Dive, V., Yiotakis, A., Yiallourous, I., Vassiliou, S., Zwilling, R., Bode, W., and Stocker, W. (1996). Structure of astacin with a transition-state analogue inhibitor. *Nature Structural Biology* 3, 671-675.
- Gray, T. E., Eder, J., Bycroft, M., Day, A. G., and Fersht, A. R. (1993). Refolding of barnase mutants and pro-barnase in the presence and absence of groel. *Embo Journal* 12, 4145-4150.
- Gros, P., Kalk, K. H., and Hol, W. G. J. (1991). Calcium-binding to thermitase - crystallographic studies of thermitase at 0,5, and 100 m-mu calcium. *Journal of Biological Chemistry* 266, 2953-2961.
- Guasch, A., Coll, M., Aviles, F. X., and Huber, R. (1992). 3-dimensional structure of porcine pancreatic procarboxypeptidase-a - a comparison of the a-zymogen and b-zymogen and their determinants for inhibition and activation. *Journal of Molecular Biology* 224, 141-157.
- Hamilton, S., and Miller, C. G. (1992). Cloning and nucleotide-sequence of the salmonella-typhimurium dcp gene encoding dipeptidyl carboxypeptidase. *Journal of Bacteriology* 174, 1626-1630.
- Hedstrom, L. (2002). Serine protease mechanism and specificity. *Chemical Reviews* 102, 4501-4523.
- Henrich, B., Becker, S., Schroeder, U., and Plapp, R. (1993). Dcp gene of escherichia-coli - cloning, sequencing, transcript mapping, and characterization of the gene-product. *Journal of Bacteriology* 175, 7290-7300.

References

- Henrich, S., Cameron, A., Bourenkov, G. P., Kiefersauer, R., Huber, R., Lindberg, I., Bode, W., and Than, M. E. (2003). The crystal structure of the proprotein processing proteinase furin explains its stringent specificity. *Nature Structural Biology* *10*, 520-526.
- Holland, D. R., Tronrud, D. E., Pley, H. W., Flaherty, K. M., Stark, W., Jansonius, J. N., McKay, D. B., and Matthews, B. W. (1992). Structural comparison suggests that thermolysin and related neutral proteases undergo hinge-bending motion during catalysis. *Biochemistry* *31*, 11310-11316.
- Holm, L., and Sander, C. (1999). Protein folds and families: Sequence and structure alignments. *Nucleic Acids Research* *27*, 244-247.
- Holmes, M. A., and Matthews, B. W. (1982). Structure of thermolysin refined at 1.6- \AA resolution. *Journal of Molecular Biology* *160*, 623-639.
- Hubbard, S. J., and Thornton, J. M. (1993). Naccess (London, Department of Biochemistry and Molecular Biology, University College London).
- Hubert, C., Houot, A. M., Corvol, P., and Soubrier, F. (1991). Structure of the angiotensin i-converting enzyme gene - 2 alternate promoters correspond to evolutionary steps of a duplicated gene. *Journal of Biological Chemistry* *266*, 15377-15383.
- Ikemura, H., and Inouye, M. (1988). Invitro processing of pro-subtilisin produced in escherichia-coli. *Journal of Biological Chemistry* *263*, 12959-12963.
- Ito, M., Dunn, B. M., and Oda, K. (1996). Substrate specificities of pepstatin-insensitive carboxyl proteinases from gram-negative bacteria. *Journal of Biochemistry* *120*, 845-850.
- Jaenicke, R. (2000). Do ultrastable proteins from hyperthermophiles have high or low conformational rigidity? *Proceedings of the National Academy of Sciences of the United States of America* *97*, 2962-2964.
- Jain, S. C., Shinde, U., Li, Y. Y., Inouye, M., and Berman, H. M. (1998). The crystal structure of an autoprocessed ser221cys-subtilisin e-propeptide complex at 2.0 \AA resolution. *Journal of Molecular Biology* *284*, 137-144.
- James, M. N. G., and Sielecki, A. R. (1986). Molecular-structure of an aspartic proteinase zymogen, porcine pepsinogen, at 1.8 \AA resolution. *Nature* *319*, 33-38.
- Karas, M., and Hillenkamp, F. (1990). Recent developments in matrix assisted laser desorption ionization mass-spectrometry of macromolecules. *Abstracts of Papers of the American Chemical Society* *200*, 140-ANYL.
- Khan, A. R., and James, M. N. G. (1998). Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Science* *7*, 815-836.

- Kim, H.M., Shin, D.R., Yoo, O.J., Lee, H. and Lee, L.O. (2003). Crystal structure of *Drosophila* angiotensin I-converting enzyme bound to captopril and lisinopril. *FEBS Letters* 538, 65-70.
- Knowles, J. R. (1991). To build an enzyme. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 332, 115-121.
- Kohama, Y., Matsumoto, S., Oka, H., Teramoto, T., Okabe, M., and Mimura, T. (1988). Isolation of angiotensin-converting enzyme-inhibitor from tuna muscle. *Biochemical and Biophysical Research Communications* 155, 332-337.
- Kojima, S., Minagawa, T., and Miura, K. (1997). The propeptide of subtilisin bpn' as a temporary inhibitor and effect of an amino acid replacement on its inhibitory activity. *FEBS Letters* 411, 128-132.
- Kraulis, P. J. (1991). Molscript - a program to produce both detailed and schematic plots of protein structures. *Journal of Applied Crystallography* 24, 946-950.
- Kwon, H. K., Kim, H., and Ahn, T. I. (1999). Pepstatin-insensitive carboxyl proteinase: A biochemical marker for late lysosomes in *amoeba proteus*. *Korean Journal of Biological Science* 3, 221-228.
- Laemmli, U. K., Molbert, E., Showe, M., and Kellenbe.E (1970). Form-determining function of genes required for assembly of head of bacteriophage-t4. *Journal of Molecular Biology* 49, 99-&.
- Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993). Procheck - a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* 26, 283-291.
- Lazure, C. (2002). The peptidase zymogen proregions: Nature's way of preventing undesired activation and proteolysis. *Current Pharmaceutical Design* 8, 511-531.
- Lee, Y. C., Koike, H., Taguchi, H., Ohta, T., and Matsuzawa, H. (1994). Requirement of a cooh-terminal pro-sequence for the extracellular secretion of aqualysin-i (a thermophilic subtilisin-type protease) in *thermus-thermophilus*. *Fems Microbiology Letters* 120, 69-74.
- Lesk, A. M., and Fordham, W. D. (1996). Conservation and variability in the structures of serine proteinases of the chymotrypsin family. *Journal of Molecular Biology* 258, 501-537.
- Leslie, A. (1991). Macromolecular data processing in crystallographic computing. In, V. D. Moras, A. D. Podjarny, and J. C. Thierry, eds. (Oxford, Oxford University Press), pp. 27-38.

References

- Li, P., Nijhawan, D., Budihardjo, I., Srinivasula, S. M., Ahmad, M., Alnemri, E. S., and Wang, X. D. (1997). Cytochrome c and datp-dependent formation of apaf-1/caspase-9 complex initiates an apoptotic protease cascade. *Cell* *91*, 479-489.
- Li, W. H., Moore, M. J., Vasilieva, N., Sui, J. H., Wong, S. K., Berne, M. A., Somasundaran, M., Sullivan, J. L., Luzuriaga, K., Greenough, T. C., *et al.* (2003). Angiotensin-converting enzyme 2 is a functional receptor for the sars coronavirus. *Nature* *426*, 450-454.
- Li, Y. Y., Hu, Z. X., Jordan, F., and Inouye, M. (1995). Functional-analysis of the propeptide of subtilisin-e as an intramolecular chaperone for protein-folding - refolding and inhibitory abilities of propeptide mutants. *Journal of Biological Chemistry* *270*, 25127-25132.
- Lim, J. H., Yu, Y. G., Han, Y. S., Cho, S. J., Ahn, B. Y., Kim, S. H., and Cho, Y. J. (1997). The crystal structure of an fe-superoxide dismutase from the hyperthermophile aquifex pyrophilus at 1.9 angstrom resolution: Structural basis for thermostability. *Journal of Molecular Biology* *270*, 259-274.
- Lin, L., Sohar, I., Lackland, H., and Lobel, P. (2001). The human cln2 protein/tripeptidyl-peptidase i is a serine protease that autoactivates at acidic ph. *Journal of Biological Chemistry* *276*, 2249-2255.
- Mach, H., Middaugh, C. R., and Lewis, R. V. (1992). Statistical determination of the average values of the extinction coefficients of tryptophan and tyrosine in native proteins. *Analytical Biochemistry* *200*, 74-80.
- Maskos, K. (1995) The bifunctional α -amylase/trypsin inhibitor from *ragi*: Three-dimensional structure, inhibitory properties and oxidative folding *in vivo* and *in vitro*, Swiss Federal Institute of Technology Zürich, Zurich.
- Massa, W. (1994). Kristallstrukturbestimmung (Stuttgart, Teubner Studienbücher: Chemie).
- Matthews, B. W. (1988). Structural basis of the action of thermolysin and related zinc peptidases. *Accounts of Chemical Research* *21*, 333-340.
- Matthews, B. W., Schoenbo.Bp, Dupourqu.D, Jansoniu.Jn, and Colman, P. M. (1972). 3-dimensional structure of thermolysin. *Nature-New Biology* *238*, 37-&.
- McDonald, J. K., Hoisington, A. R., and Eisenhauer, D. A. (1985). Partial-purification and characterization of an ovarian tripeptidyl peptidase - a lysosomal exopeptidase that sequentially releases collagen-related (gly-pro-x) triplets. *Biochemical and Biophysical Research Communications* *126*, 63-71.

- McKie, N., Dando, P. M., Rawlings, N. D., and Barrett, A. J. (1993). Thimet oligopeptidase - similarity to soluble angiotensin ii-binding protein and some corrections to the published amino-acid-sequence of the rat testis enzyme. *Biochemical Journal* *295*, 57-60.
- Merritt, E. A., and Bacon, D. J. (1997). Raster3d: Photorealistic molecular graphics. In *Macromolecular crystallography*, pt b, pp. 505-524.
- Mullereberhard, H. J. (1988). Molecular-organization and function of the complement-system. *Annual Review of Biochemistry* *57*, 321-347.
- Murao, S., Oda, K., and Matsushi, Y. (1972). New acid proteases from scytalidium-lignicolum m-133. *Agricultural and Biological Chemistry* *36*, 1647-&.
- Murao, S., Ohkuni, K., Nagao, M., Hirayama, K., Fukuhara, K., Oda, K., Oyama, H., and Shin, T. (1993). Purification and characterization of kumamolysin, a novel thermostable pepstatin-insensitive carboxyl proteinase from bacillus novosp mn-32. *Journal of Biological Chemistry* *268*, 349-355.
- Murao, S., Ohkuni, K., Nagao, M., Oda, K., and Shin, T. (1988). A novel thermostable, s-pi (pepstatin ac) insensitive acid proteinase from thermophilic bacillus novosp strain mn-32. *Agricultural and Biological Chemistry* *52*, 1629-1631.
- Natesh, R., Schwager, S. L. U., Sturrock, E. D., and Acharya, K. R. (2003). Crystal structure of the human angiotensin-converting enzyme-lisinopril complex. *Nature* *421*, 551-554.
- Navaza, J. (1994). Amore - an automated package for molecular replacement. *Acta Crystallographica Section a* *50*, 157-163.
- Nicholls, A., Bharadwaj, R., and Honig, B. (1993). Grasp - graphical representation and analysis of surface- properties. *Biophysical Journal* *64*, A166-A166.
- Oda, K., Ito, M., Uchida, K., Shibano, Y., Fukuhara, K., and Takahashi, S. (1996). Cloning and expression of an isovaleryl pepstatin-insensitive carboxyl proteinase gene from xanthomonas sp t-22. *Journal of Biochemistry* *120*, 564-572.
- Oda, K., Ogasawara, S., Oyama, H., and Dunn, B. M. (2000). Subsite preferences of pepstatin-intensitive carboxyl proteinases from prokaryotes: Kumamolysin, a thermostable pepstatin-insensitive carboxyl proteinase. *Journal of Biochemistry* *128*, 499-507.
- Oda, K., Takahashi, T., Tokuda, Y., Shibano, Y., and Takahashi, S. (1994). Cloning, nucleotide-sequence, and expression of an isovaleryl pepstatin-insensitive carboxyl proteinase gene from pseudomonas sp-101. *Journal of Biological Chemistry* *269*, 26518-26524.

References

- Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., *et al.* (1992). The alpha/beta-hydrolase fold. *Protein Engineering* 5, 197-211.
- Otwinowski, Z., and Minor, W. (1993). *Denzo: A film processing for macromolecular crystallography* (Yale, USA, New Haven: Yale University).
- Oyama, H., Abe, S., Ushiyama, S., Takahashi, S., and Oda, K. (1999). Identification of catalytic residues of pepstatin-insensitive carboxyl proteinases from prokaryotes by site-directed mutagenesis. *Journal of Biological Chemistry* 274, 27815-27822.
- Oyama, H., Hamada, T., Ogasawara, S., Uchida, K., Murao, S., Beyer, B. B., Dunn, B. M., and Oda, K. (2002). A *cln2*-related and thermostable serine-carboxyl proteinase, kumamolysin: Cloning, expression, and identification of catalytic serine residue. *Journal of Biochemistry* 131, 757-765.
- Peisach, E., Wang, J. Y., de los Santos, T., Reich, E., and Ringe, D. (1999). Crystal structure of the proenzyme domain of plasminogen. *Biochemistry* 38, 11180-11188.
- Philippsen, A. (1999). Visualizing structural biology (<http://www.Bioz.Unibas.Ch/~xray/dino>).
- Pierotti, A., Dong, K. W., Glucksman, M. J., Orłowski, M., and Roberts, J. L. (1990). Molecular-cloning and primary structure of rat testes metalloendopeptidase ec 3.4.24.15. *Biochemistry* 29, 10323-10329.
- Qiu, X. Y., Culp, J. S., DiLella, A. G., Hellmig, B., Hoog, S. S., Janson, C. A., Smith, W. W., and AbdelMeguid, S. S. (1996). Unique fold and active site in cytomegalovirus protease. *Nature* 383, 275-279.
- Ramachandran, G. N., and Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Advances in Protein Chemistry* 23, 283-437.
- Rawlings, N. D., and Barrett, A. J. (1993). Evolutionary families of peptidases. *Biochemical Journal* 290, 205-218.
- Rawlings, N. D., and Barrett, A. J. (1995). Evolutionary families of metallopeptidases. In *Proteolytic enzymes: Aspartic and metallo peptidases*, pp. 183-228.
- Rawlings, N. D., and Barrett, A. J. (1999a). Merops: The peptidase database. *Nucleic Acids Research* 27, 325-331.
- Rawlings, N. D., and Barrett, A. J. (1999b). Tripeptidyl-peptidase i is apparently the *cln2* protein absent in classical late-infantile neuronal ceroid lipofuscinosis. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* 1429, 496-500.

- Rieger, K. J., Saezservent, N., Papet, M. P., Wdzieczakbakala, J., Morgat, J. L., Thierry, J., Voelter, W., and Lenfant, M. (1993). Involvement of human plasma angiotensin i-converting enzyme in the degradation of the haemoregulatory peptide n-acetyl-seryl-aspartyl-lysyl-proline. *Biochemical Journal* 296, 373-378.
- Riordan, J. F. (2003). Angiotensin-i-converting enzyme and its relatives. *Genome Biology* 4.
- Roussel, S., and Cambillau, C. (1991). The turbo-frodo graphics package. In *Silicon graphics geometry partners directory* (Mountain View, California, USA, Silicon Graphics Corp.), pp. 77-79.
- Rudenko, G., Bonten, E., Dazzo, A., and Hol, W. G. J. (1995). 3-dimensional structure of the human protective protein - structure of the precursor form suggests a complex activation mechanism. *Structure* 3, 1249-1259.
- Sambrook, J., Fritsh, E.F., Maniatis, T. (1989). In *molecular cloning: A laboratory manual* (Cold Spring Harbor, New York, 2nd ed. Cold Spring Harbor Laboratory).
- Sauter, N. K., Mau, T., Rader, S. D., and Agard, D. A. (1998). Structure of alpha-lytic protease complexed with its pro region. *Nature Structural Biology* 5, 945-950.
- Schechter, I., and Berger, A. (1967). On size of active site in proteases .I. Papain. *Biochemical and Biophysical Research Communications* 27, 157-&.
- Serizawa, A., Dando, P. M., and Barrett, A. J. (1995). Characterization of a mitochondrial metallopeptidase reveals neurolysin as a homolog of thimet oligopeptidase. *Journal of Biological Chemistry* 270, 2092-2098.
- Serizawa, A., Dando, P. M., and Barrett, A. J. (1997). Oligopeptidase m (neurolysin). Targeting to mitochondria and cytosol in rat tissues. In *Proteolysis in cell functions*, V. K. Hopsu-Havu, M. Järvinen, and H. Kirschke, eds. (Amsterdam, IOS Press), pp. 248-255.
- Shibata, M., Dunn, B. M., and Oda, K. (1998). Substrate specificity of pepstatin-insensitive carboxyl proteinase from bacillus coagulans j-4. *Journal of Biochemistry* 124, 642-647.
- Shieh, H. S., Kurumbail, R. G., Stevens, A. M., Stegeman, R. A., Sturman, E. J., Pak, J. Y., Wittwer, A. J., Palmier, M. O., Wiegand, R. C., Holwerda, B. C., and Stallings, W. C. (1996). Three-dimensional structure of human cytomegalovirus protease (vol 383, pg 279, 1996). *Nature* 384, 288-288.
- Shinde, U. P., Liu, J. J., and Inouye, M. (1997). Protein memory through altered folding mediated by intramolecular chaperones. *Nature* 389, 520-522.
- Siezen, R. J., Devos, W. M., Leunissen, J. A. M., and Dijkstra, B. W. (1991). Homology modeling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases. *Protein Engineering* 4, 719-737.

- Siezen, R. J., and Leunissen, J. A. M. (1997). Subtilases: The superfamily of subtilisin-like serine proteases. *Protein Science* 6, 501-523.
- Skeggs, L. T., Kahn, J. R., and Shumway, N. P. (1956). The preparation and function of the hypertensin-converting enzyme. *Journal of Experimental Medicine* 103, 295-299.
- Skidgel, R. A., Engelbrecht, S., Johnson, A. R., and Erdos, E. G. (1984). Hydrolysis of substance-p and neurotensin by converting enzyme and neutral endopeptidase. *Peptides* 5, 769-776.
- Skidgel, R. A., and Erdos, E. G. (1987). The broad substrate-specificity of human angiotensin-i converting enzyme. *Clinical and Experimental Hypertension Part a-Theory and Practice* 9, 243-259.
- Sleat, D. E., Donnelly, R. J., Lackland, H., Liu, C. G., Sohar, I., Pullarkat, R. K., and Lobel, P. (1997). Association of mutations in a lysosomal protein with classical late-infantile neuronal ceroid lipofuscinosis. *Science* 277, 1802-1805.
- Strausberg, S. L., Alexander, P. A., Gallagher, D. T., Gilliland, G. L., Barnett, B. L., and Bryan, P. N. (1995). Directed evolution of a subtilisin with calcium-independent stability. *Bio-Technology* 13, 669.
- Suopanki, J., Partanen, S., Ezaki, J., Baumann, M., Kominami, E., and Tyynela, J. (2000). Developmental changes in the expression of neuronal ceroid lipofuscinoses-linked proteins. *Molecular Genetics and Metabolism* 71, 190-194.
- Tangrea, M. A., Alexander, P., Bryan, P. N., Eisenstein, E., Toedt, J., and Orban, J. (2001). Stability and global fold of the mouse prohormone convertase 1 pro-domain. *Biochemistry* 40, 5488-5495.
- Tangrea, M. A., Bryan, P. N., Sari, N., and Orban, J. (2002). Solution structure of the pro-hormone convertase 1 pro-domain from mus musculus. *Journal of Molecular Biology* 320, 801-812.
- Tanner, J. J., Hecht, R. M., and Krause, K. L. (1996). Determinants of enzyme thermostability observed in the molecular structure of thermus aquaticus d-glyceraldehyde-3- phosphate dehydrogenase at 2.5 angstrom resolution. *Biochemistry* 35, 2597-2609.
- Teague, S. J. (2003). Implications of protein flexibility for drug discovery. *Nature Reviews Drug Discovery* 2, 527-541.
- Tong, L., Qian, C. G., Massariol, M. J., Bonneau, P. R., Cordingley, M. G., and Lagace, L. (1996). A new serine-protease fold revealed by the crystal structure of human cytomegalovirus protease. *Nature* 383, 272-275.

- Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets - procedure and some applications. *Proceedings of the National Academy of Sciences of the United States of America* 76, 4350-4354.
- Towler, P., Staker, B., Prasad, S. G., Menon, S., Tang, J., Parsons, T., Ryan, D., Fisher, M., Williams, D., Dales, N. A., *et al.* (2004). Ace2 x-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. *Journal of Biological Chemistry* 279, 17996-18007.
- Turk, D. (1992) Weiterentwicklung eines programms für molekülgraphik und elektronendichte-manipulation und seine anwendung auf verschiedene protein-strukturaufklärungen (<http://www-bmb.ljs.si/doc>), Technische Universität, München.
- Turk, D., Podobnik, M., Kuhelj, R., Dolinar, M., and Turk, V. (1996). Crystal structures of human procathepsin b at 3.2 and 3.3 angstrom resolution reveal an interaction motif between a papain-like cysteine protease and its propeptide. *Febs Letters* 384, 211-214.
- Vickers, C., Hales, P., Kaushik, V., Dick, L., Gavin, J., Tang, J., Godbout, K., Parsons, T., Baronas, E., Hsieh, F., *et al.* (2002). Hydrolysis of biological peptides by human angiotensin-converting enzyme-related carboxypeptidase. *Journal of Biological Chemistry* 277, 14838-14843.
- Vincent, B., Dive, V., Yiotakis, A., Smadja, C., Maldonado, R., Vincent, J. P., and Checler, F. (1995). Phosphorus-containing peptides as mixed inhibitors of endopeptidase-3.4.24.15 and endopeptidase-3.4.24.16 - effect on neurotensin degradation in-vitro and in-vivo. *British Journal of Pharmacology* 115, 1053-1063.
- Vincent, B., Jiracek, J., Noble, F., Loog, M., Roques, B., Dive, V., Vincent, J. P., and Checler, F. (1997). Effect of a novel selective and potent phosphinic peptide inhibitor of endopeptidase 3.4.24.16 on neurotensin-induced analgesia and neuronal inactivation. *British Journal of Pharmacology* 121, 705-710.
- Voges, D., Zwickl, P., and Baumeister, W. (1999). The 26s proteasome: A molecular machine designed for controlled proteolysis. *Annual Review of Biochemistry* 68, 1015-1068.
- Waeber, B., Nussberger, J., and Brunner, H. R. (1990). Angiotensin converting enzyme-inhibition in arterial-hypertension. *Wiener Medizinische Wochenschrift* 140, 22-30.
- Wang, J. M., Hartling, J. A., and Flanagan, J. M. (1997). The structure of clpp at 2.3 angstrom resolution suggests a model for atp-dependent proteolysis. *Cell* 91, 447-456.
- Wlodawer, A., Li, M., Dauter, Z., Gustchina, A., Uchida, K., Oyama, H., Dunn, B. M., and Oda, K. (2001a). Carboxyl proteinase from pseudomonas defines a novel family of subtilisin-like enzymes. *Nature Structural Biology* 8, 442-446.

References

- Wlodawer, A., Li, M., Gustchina, A., Dauter, Z., Uchida, K., Oyama, H., Goldfarb, N. E., Dunn, B. M., and Oda, K. (2001b). Inhibitor complexes of the pseudomonas serine-carboxyl proteinase. *Biochemistry* *40*, 15602-15611.
- Wlodawer, A., Li, M., Gustchina, A., Oyama, H., Dunn, B. M., and Oda, K. (2003). Structural and enzymatic properties of the sedolisin family of serine-carboxyl peptidases. *Acta Biochimica Polonica* *50*, 81-102.
- Xiao, X. D., Chakraborti, S., Dimitrov, A. S., Gramatikoff, K., and Dimitrov, D. S. (2003). The sars-cov s glycoprotein: Expression and functional characterization. *Biochemical and Biophysical Research Communications* *312*, 1159-1164.
- Yang, H. Y. T., Erdos, E. G., and Levin, Y. (1970). A dipeptidyl carboxypeptidase that converts angiotensin i and inactivates bradykinin. *Biochimica Et Biophysica Acta* *214*, 374-&.
- Yanischperron, C., Vieira, J., and Messing, J. (1985). Improved m13 phage cloning vectors and host strains - nucleotide-sequences of the m13mp18 and puc19 vectors. *Gene* *33*, 103-119.
- Yaron, A., Berger, A., and Mlynar, D. (1972). Dipeptidocarboxypeptidase from e-coli. *Biochemical and Biophysical Research Communications* *47*, 897-&.
- Zhu, X. L., Ohta, Y., Jordan, F., and Inouye, M. (1989). Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process. *Nature* *339*, 483-484.
- Zwickl, P., Kleinz, J., and Baumeister, W. (1994). Separate expression of the 2 subunits of the thermoplasma proteasome. *Journal of Cellular Biochemistry*, 177-177.

8 Acknowledgements

The present work was carried out in the Abteilung für Strukturforshung at the Max-Planck-Institut für Biochemie in Martinsried (Germany), under the supervision of Prof. Dr. Robert Huber.

Mein besonderer Dank gilt Herrn Prof. Dr. Robert Huber für die überaus freundliche Aufnahme in seine Arbeitsgruppe sowohl für die große Freiheit bei der Anfertigung dieser Doktorarbeit.

Bei Herrn Prof. Dr. Wolfram Bode möchte ich mich für die Überlassung der interessanten Themen bedanken sowie für stimulierende Diskussionen, ständige Förderung und großzügige Unterstützung.

Herrn Dr. Klaus Maskos danke ich herzlich für nützlicher Ratschläge, viele wertvolle Anregungen und für die mühsam Korrektur dieser Arbeit. Ebenso möchte ich mich bei meinen Laborkollegen für die angenehme und produktive Arbeitsatmosphäre bedanken. Besonders danke ich Marianne Braun die mir am Anfang meiner Doktorarbeit so viel im Labor geholfen hat, und Charlotte Ungewickell mit ihrer magischen Pipette. Meinen Kollegen Tobias Krojer und Reiner Friedrich danke ich für die angenehme und freundschaftliche Atmosphäre und für seine unglaubliche Hilfsbereitschaft.

Un mega-gracias a Pablo Fuentes por tu ayuda incondicional y por tus sesiones (absolutamente voluntarias) de autofustigamiento corrigiendo esta tesis y unos cuantos embriones de artículos. Gracias también, por largas conversaciones en las que ninguno de los dos cambiamos nunca de opinión -o era que sí?- pero en las que poníamos todo nuestro afán de contradicción.

Mil obligadas a Sofia Macieira por ser tan hospitalaria y jovial y, naturalmente, por tus famosas opíparas cenas. Mil gracias también a Marta Garrido, sin la cual seguramente no hubiera venido a parar a Martinsried, por tu optimismo y tu buen humor. A David Reverter li dec llargues, llarguíssimes nits de Bob Dylan en tots els formats (video, CD, guitarra en directe...) i molt bons records per les muntanyes bavareses.

Bei Frau Renate Rüller und Herrn Werner Dersch bedanke ich mich sehr herzlich für ihre stete Hilfsbereitschaf und die vielfältige Unterstützung bei Problemen mit Verwaltung und Technik.

Acknowledgements

Mein dickes Dankeschön an Iris, alias die Klugscheißerin, Du bist unsere MEISTERIN für Dekoration, Basteln, Blumen und Farben geworden -I am not worthy. Danke für soooo viele Klatschtante- Morgen, Mittage, und Abende.

Mein fettes Hvala an die Dadi, Deine Großzügigkeit und Hilfsbereitschaft sind einfach unerschöpflich. Schön zu wissen dass wir immer B&B Rezepte zusammen kochen können und dabei unzählige ekelige Geschichten erfinden und unsere Cartucheras durchmessen. Respekt!!!

Eva Este Perpi, que t'he de dir que no t'hagi dit ja en les nostres famoses conferències telefòniques, comunicacions electròniques i trobades místico-virtuals en el món dels somnis? Pot ser que gràcies per compartir les meves neures, pors i histèries i convertir-les en unes "risas"? Doncs això, que gràcies per ser fantàstica.

Gràcies amb totes les lletres -i d'altres que no puc escriure perquè resultaria massa íntim- a la Diana i la Marta. Des de la llunyania, m'heu fet l'exili laboral a Munx agradable en els moments difícils -i genial en els bons moments- enviant-me postalletes, calendaris d'Advent, fotos, enviant-vos a vosaltres mateixes... espero que sempre sigueu tan Öpet!

Unes gràcies molt especials als meus pares, Doris i Josep, que sempre m'heu recolzat i donat suport, encara que fos per marxar a aquestes terres llunyanes del nord. No em puc imaginar uns pares millors -i això que de fantasia no me'n falta, eh?. També gràcies a tu Dani, per aguantar una germana tan repel·lentita.

Roland, sin ti, no solamente no hubiera escrito ni la primera letra de esta... cosa (todavía no he decidido que nombre ponerle...), sino que aún estaría sumergida en un mundo de tinieblas, completamente ignorante de lo que es ser un verdadero nerd. Be one with the force, Ro.