Thesis

Tobias Krojer

Crystal Structure of DegP from *Escherichia coli*

2004

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Crystal Structure of DegP from *Escherichia coli*

by Tobias Krojer

submitted for the degree of Ph.D. to
School of Biosciences
Cardiff University
Cardiff CF10 3US
United Kingdom

The research was performed at:
Max-Planck-Institut für Biochemie
Abteilung Strukturforschung
82152 Martinsried
Germany

2004

Part of the work presented here has been previously published in:

Krojer T, Garrido-Franco M, Huber R, Ehrmann M, Clausen T. (2002). Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine. *Nature* **416**(6879):455-9.

Summary

DegP is a heat-shock protein which is localized in the periplasm of *Escherichia coli*. It is a common protein quality control factor and represents the only protein that can alternate between the two antagonistic activities of a protease and a chaperone in a temperature-dependent manner. In this study the crystal structure of the hexameric form of $DegP_{S210A}$ from *Escherichia coli* was solved at 2.8Å resolution by multiple anomalous dispersion phasing. As the protein was crystallized at room temperature, this structure represents the chaperone conformation. Each DegP monomer is built up by a trypsin-like serine protease domain and two consecutive PDZ domains. The hexamer is a dimer of trimers with crystallographic D3 symmetry. Oligomerization is mediated by the protease domains and results in the formation of an internal cavity where the active sites are located. Access towards the internal cavity is controlled by the flexible PDZ domains. The protease activity is absent because access towards the active sites is blocked by the interaction of several surface loops. Furthermore, the active site geometry is distorted. The crystal structure of the wild-type DegP confirmed that the inactive conformation is not due to the artificial serine to alanine mutation of the $DegP_{S210A}$ structure but represents an inherent feature of the chaperone state to avoid unwanted proteolysis at room temperature. The crystal structure of DegP in complex with the covalent serine protease inhibitor diisopropyl fluorophosphate could not preserve the protease conformation. Analysis of degradation products by mass spectrometry revealed that the product length varies between 6 and 25 amino acid residues with a clear preference for small hydrophobic residues in the P1 position. Furthermore, time-dependent analyses of degradation products by high-performance liquid chromatography showed that DegP degrades its substrates in a processive fashion, similar to other cage-forming proteases.

Abbreviations

amp	ampicillin
APS	ammonium persulphate
BisANS	4,4'-dianilino-1,1'-binaphtyl-5,5'-disulfonic acid
BSA	bovine serum albumine
DFP	diisopropyl fluorophosphate
DIC	3,4-dichloroisocoumarin
DMSO	dimethyl sulfoxide
DLS	dynamic light scattering
Fig.	figure
GSSG	L-Glutathione oxidized
HAP	hydroxyapatite
HEPES	N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulphonic acid]
HPLC	high-performance liquid chromatography
IPTG	isopropyl-β-D-thiogalactopyranoside
kDa	kilo dalton
Ms-AAPV-CMK	N-(Methoxysuccinyl)-Ala-Ala-Pro-Val-chloromethyl ketone
MG-115	carbobenzoxy-L-leucyl-L-leucyl-L-norvalinal
MG-132	carbobenzoxy-L-leucyl-L-leucyl-leucinal
MME	monomethylether
MPD	2-Methyl-2,4-pentanediol
Ni-NTA	nickel-Nitrilotriacetic acid
NMM	new minimal medium
LB	Luria Bertani
MPI	Max-Plank Institute for Biochemistry (Martinsried, GERMANY)
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PEG	polyethylene glycol
PMSF	phenylmethanesulfonyl fluoride
PSI	carbobenzoxy-L-isoleucyl-γ-t-butyl-L-glutamyl-L-alanyl-L-leucinal
rpm	revolutions per minute
RT	room temperature
SDS	sodium dodecyl sulphate

SEC	size exclusion chromatography
SeMet	selenomethionine
Tab.	table
TCA	trichloroacetic acid
TEMED	N,N,N',N'-Tetramethylethylenediamine
TRIS	tris[hydroxymethyl]aminomethane
wt	wild-type
v/v	volume per volume
w/v	weight per volume

Contents

Acknowledgements vi

1. Abstract 1

2. Introduction 3

2.1 The periplasm of *Escherichia coli* 3

2.2 Protein quality control in *Escherichia coli* 6

2.2.1 Heat-shock proteins 6

2.2.2 The molecular chaperones 7

2.2.3 The proteases 8

2.2.4 Cage-forming proteases 11

2.2.5 Summary of protein quality control 14

2.3 PDZ domains: modules for protein-protein interactions 14

2.4 The periplasmic heat-shock protein DegP 17

2.4.1 Domain structure of DegP 17

2.4.2 Structure of DegP 17

2.4.3 Physiological functions of DegP 18

2.4.4 DegP homologues in *Escherichia coli* 19

2.4.5 DegP homologues in other organisms 19

2.4.6 DegP and virulence 20

2.4.7 Regulation of DegP 20

2.5 Protein X-ray crystallography 23
2.5.1 Overview: protein structure determination
by X-ray crystallography 23
2.5.2 The phase problem 24
2.5.2.1 Starting point 24
2.5.2.2 Two-dimensional representations of structure factors 26
2.5.2.3 Friedel's law 26
2.5.3 Solution of the phase problem
by multiple anomalous dispersion (MAD) 27
2.5.3.1 Anomalous diffraction 27
2.5.3.2 Determination of the heavy atom substructure 31
2.5.3.2.1 Patterson methods 31
2.5.3.2.2 Direct methods 32
2.5.3.3 Phase determination 33
2.5.3.4 Single anomalous dispersion (SAD) 35
2.6 Aim of the project 36

3. Materials and Methods 37

3.1 Chemicals and their suppliers 37
3.2 Protease substrates, peptides and inhibitors 37
3.3 Crystallization equipment 38
3.4 Bacterial strains 38
3.5 Plasmids 38
3.6 Media 38
3.7 Microbiological methods 41
3.7.1 Sterilisation 41
3.7.2 Growth conditions 41
3.7.3 Determination of the cell density in liquid cultures 41
3.8 Biochemical methods 41
3.8.1 SDS polyacrylamide gel electrophoresis (SDS-PAGE) 41
3.8.2 Coomassie blue staining 43

3.8.3 Determination of protein concentration 43
3.8.3.1 Bradford protein assay 44
3.8.3.2 Absorbance at 280nm 44
3.8.4 Electroblotting of proteins to PVDF membranes 44
3.8.5 Protein purification 45
3.8.5.1 Purification of DegP and $DegP_{S210A}$ 45
3.8.5.2 Purification of DegP in complex with DFP 46
3.8.6 N-terminal sequencing 47
3.8.7 Mass spectrometry 48
3.8.8 Size exclusion chromatography (SEC) 48
3.8.9 Dynamic light scattering (DLS) 49
3.8.10 Protease assay 49
3.8.10.1 Synthetic substrates 49
3.8.10.2 Resorufin-labelled casein 49
3.8.11 Sensitivity of DegP to various inhibitors 50
3.8.12 Protease assays in the presence of various peptides 50
3.8.13 Effect of DFP on DegP at various temperatures 51
3.8.14 Analyses of degradation products by mass spectrometry 51
3.8.15 Time-dependent degradation of substrate proteins 52
3.9 Crystallographic methods 52
3.9.1 Protein crystallization 52
3.9.2 Preparation of heavy atom derivatives 53
3.9.3 Improvement of diffraction properties 54
3.9.4 Data collection and data processing 54
3.9.5 Structure solution 55
3.9.6 Model building and refinement 56
3.9.7 Graphical representations and sequence alignments 56

4. Results **57**

4.1 Purification of $DegP_{S210A}$, wild-type DegP (DegP/$DegP_{(+DFP)}$) and DegP in complex with DFP 57
4.1.1 Purification of $DegP_{S210A}$ 57
4.1.2 Purification of DegP 60

4.1.3 Purification of DegP in complex with DFP 61
4.1.4 Purification of selenomethionine substituted DegP proteins 61
4.2 Identification of the purified proteins 62
4.2.1 Mass spectrometry 62
4.2.2 N-terminal sequencing 63
4.3 Oligomeric state of $DegP_{S210A}$ in solution 64
4.3.1 Size-exclusion chromatography (SEC) 64
4.3.2 Dynamic light scattering (DLS) 65
4.3.3 Summary 67
4.4 Inhibitors and activators of DegP 67
4.4.1 Protease assay: synthetic versus protein substrates 67
4.4.2 Sensitivity of DegP to different inhibitors 68
4.4.3 Effect of different peptides on the protease activity of DegP 69
4.5 Temperature-dependent reaction of DegP with DFP 70
4.6 DegP degrades its substrates in a processive fashion 71
4.7 The length distribution of DegP's degradation products 73
4.8 Protein crystallization 76
4.8.1 Screening for initial crystallization conditions 76
4.8.2 Optimization of crystal quality 77
4.8.3 Crystallization of different DegP constructs 78
4.8.3.1 Crystallization of $DegP_{S210A,12\text{-}B}$ 78
4.8.3.2 Crystallization of DegP+DFP and $DegP_{(+DFP)}$ 79
4.8.3.3 Crystallization of selenomethione-substituted DegP's 79
4.9 Improvement of diffraction properties 80
4.10 Crystal structure of $DegP_{S210A}$ 82
4.10.1 Data collection and structure determination 82
4.10.2 Model building and refinement 84
4.10.3 Description of the structure 85
4.10.3.1 Tertiary structure 85
4.10.3.2 Assembly of the hexamer 86
4.10.3.3 The protease domain 89
4.10.3.4 Characteristics of the inner cavity 90

4.10.3.5 Substrate binding properties of the PDZ domains 92
4.11 Crystal structure of DegP+DFP 95
4.11.1 Data collection and structure determination 95
4.11.2 Model building and refinement 97
4.11.3 Overall structure and comparison with the $DegP_{S210A}$ structure 98
4.11.4 The active site 100
4.12 Crystal structure of wild-type DegP ($DegP_{(+DFP)}$) 101
4.12.1 Data collection and structure determination 101
4.12.2 Model building and refinement 102
4.12.3 Overall structure and comparison with the $DegP_{S210A}$ structure 103
4.12.4 The active site 103

5. Discussion **105**

5.1 Quality of the structure 105
5.1.1 $DegP_{S210A}$ 105
5.1.2 DegP+DFP 105
5.1.3 $DegP_{(+DFP)}$ 106
5.2 Quaternary structure of DegP in solution 106
5.3 DegP, a novel type of cage-forming protease 107
5.4 A janus-headed molecular machine 108
5.5 Possible modes of regulation 110
5.6 PDZ domains of DegP: molecular gate keepers 113
5.7 DegP is particular in the choice of its inhibitors 114
5.8 The inhibition of DegP by DFP is temperature-dependent 115
5.9 DFP supports the crystallization of the active protease 115
5.10 DFP could not preserve the protease state of DegP 117
5.11 DegP is a processive protease 118

6. References **121**

Acknowledgements

The work presented here was completed in the department for structural research at the Max Planck-Institute for Biochemistry (Martinsried, Germany), under the supervision of Prof. Dr. Michael Ehrmann (Cardiff University).

I would like to thank Prof. Dr. Robert Huber for giving me the opportunity to perform this work in his department. Thank you for the unlimited scientific support and your continuous interest in the progress of this work.

I am grateful to Prof. Dr. Michael Ehrmann for supervising this work. Thanks for your continuous interest , for the inspiring discussions and your friendship.

Furthermore I would like to thank PD Dr. Tim Clausen for his help throughout this work and beyond. Thanks a lot for your advice, encouragement and friendship.

A big thank you to my office mates Mireia Comellas, Dr. Eva Estébanez Perpiñá, Dr. Marta Garrido Franco, Dr. Sofia Macieira and Emina Savarese for the friendly and helpful atmosphere.

Thank you very much to all the people who shared the laboratory with be for the relaxed atmosphere: Irena Bonin, Constanze Breithaupt, Dr. Uwe Jacob, Cora Keil, Snezan Marinkovic, Dr. Berta Martins.

Many thanks to people from the secretary and the maintenance staff: Gina Beckmann, Monika Bumann, Marion Heinze, Renate Rüller, Werner Dersch and Ottmar Paul.

I further want to thank people from synchrotron beamlines in Hamburg and Grenoble for their support as there are Dr. Gleb Bourenkov, Dr. Steffi Arzt, Dr. Stefanie Monaco and Dr. Joanne McCarthy.

Thanks to Dr. Cyril Boulenque and Markus Schütt for HPLC analyses and to Dr. Frank Siedler for mass spectrometry analyses. Christoph Spiess for providing cells for SeMet expression.

Further I want to acknowledge all members of Michael Ehrmann's group in Cardiff for the nice welcome and the great time I spent there: Sarah Amir, Dr. Alexandra Beil, Dr. Markus Eser, Dr. Sandra Grau, Dr. Natasha Mikhaleva

Finally thanks a lot to all the present and the former members of the Huber department for the great time: Dr. Werner Atzenhofer , Dr. Martin Augustin, Dr. Jae-Hyun Bae , Prof. Dr. Wolfram Bode, Stefanie Bauer, Petra Birle , Dr. Hans Brandstetter, Dr. Nediljko Budisa, Li-Chi Chang, Lorenz Chatwell, Dr. Richard Engh, Michael Engel, Dr. Rainer Friedrich, Dr. Iris Fritze, Dr. Pablo Fuentes-Prior, Dr. Peter Goettig, Dr. Stefan Henrich, Dr. Daniela Jozic, C.-P. Kaerlein, Dr. Norman Kairies, Dr. Reiner Kiefersauer, Dr. Michael Koch, Tatjana Krywcun, Otto Kyrieleis, Martin Locher, Dr. Klaus Maskos, PD Dr. Albrecht Messerschmidt, Andrea Papendorf, Arne Ramsperger, Marina Rubini, Monika Schneider, Dr. Peter Sondermann, Susanne Stanitzek, Thomas Steiner, Dr. Manuel Than, Anna Tochowicz, Charlotte Ungewickell, Milko Velarde, Waltraud Wenger, Katja Wenig, Rasso Willkomm and of course all those people I forgot to mention in this list.

Finally I would like to thank my parents; without their continuous support this work would have been impossible.

Chapter 1

Abstract

Molecular chaperones and proteases monitor the folded state of other proteins. In addition to recognizing non-native conformations, these quality control factors distinguish substrates that can be refolded from those that need to be degraded because they are severely damaged. DegP is one of the rare proteins that can combine both functions on a single polypeptide chain. The switch between the two antagonistic activities is controlled in a temperature-dependent manner, a unique feature that has never been observed before. At temperatures below 30°C, DegP serves exclusively as a chaperone, but with increasing temperature the protease activity starts to dominate.

In the first part of the study, the crystal structure of the hexameric form of $DegP_{S210A}$ from *Escherichia coli* was solved at 2.8Å resolution by multiple anomalous dispersion (MAD) phasing. Each monomer is built up by a trypsin-like serine protease domain and two consecutive PDZ domains. The whole particle is a dimer of trimers with crystallographic D3 symmetry. The protease domains interact with each other and constitute the framework of the particle while the PDZ domains act as flexible side walls. The protease interaction generates an inner cavity with a height of about 15Å and a diameter of about 50Å. The six active sites are located in this inner cavity. The crystals were grown at room temperature, thus the structure represents the chaperone conformation which is devoid of protease activity due to two reasons. First, the active site geometry is distorted and second, the active site loops L1 and L2 and the elongated loop LA, originating from an opposing monomer interact with each other and thereby block substrate access towards the active site. In the crystal structure, the particle could be observed in two distinct states, an open and a closed form with different orientations and positions of the respective PDZ domains. The PDZ domains represent the flexible sidewalls of the particle. They may therefore be considered as the mobile sidewalls of the structure which may on the one hand capture substrates by their inherent protein binding capability and on the other hand serve as gate keepers, thereby controlling access of substrates

towards the inner cavity. The walls of the inner cavity are lined by a total of 24 highly flexible phenylalanine residues which may represent a suitable binding platform for unfolded protein substrates.

The structure of DegP in complex with the covalent serine protease inhibitor diisopropyl fluorophosphates (DFP) was solved in the second part of this work by MAD phasing at 2.9Å resolution. Complete inhibition of DegP was achieved by incubating the sample at 40°C, but the structure of DegP in complex with DFP revealed that the small molecule inhibitor was not able to fix DegP in the protease conformation. Several conformational changes throughout the active site region could be observed but a great part of the relevant active site loops became very flexible and could not be modelled. The structure looks rather like a transition state between protease and chaperone.
During the first purification of the DegP+DFP complex, complete inhibition of DegP failed, but led to well diffracting crystals of the wild-type protein. The absence of DFP moieties in the crystals was checked by mass spectrometry. The crystal structure of wild-type DegP was refined to 2.5Å resolution with the $DegP_{S210A}$ structure serving as a starting model. The structure was very similar to the $DegP_{S210A}$ structure. The active site loops showed the same inactive conformation as in the initial $DegP_{S210A}$ structure, thus the observed unusual active site geometry was not an artefact due to the serine to alanine mutation but represents an inherent feature of the chaperone conformation in order to avoid unwanted proteolysis.

In the final part of this work the protease activity of DegP was studied by biochemical methods. The time-dependent degradation of substrate proteins was monitored by HPLC and revealed that proteins were degraded at multiple sites without the release of high molecular weight intermediates. Thus, DegP degrades its substrates in a processive fashion like other cage-forming proteases. Further analysis of degradation products by mass spectrometry revealed that the product length varies between 6 and 25 amino acid residues. Additionally, DegP has a clear preference for small hydrophobic residues in the P_1 position but is rather indiscriminate in the choice of the P_1' residue.

Chapter 2

Introduction

2.1 The periplasm of *Escherichia coli*

The periplasm is the space between the inner and the outer membrane in gram-negative bacteria (Fig. 2.1). The exact dimensions of this compartment remain unclear, although a number of electron microscopy studies dealing with this topic have been carried out. The estimated width is about 13 to 25nm and its volume includes between 8-16% of the total cell volume. The obvious variation in dimension is probably due to different sample preparation methods. Another structural feature located in the intermembrane space is the peptidoglycan layer also termed murein sacculus. The sacculus is in close association with the outer membrane and has a thickness of 5-8nm (Oliver, 1996) (Fig. 2.1).

Measurements of lateral diffusion rates of proteins within the periplasm have revealed a 1000-fold lower number than comparable measurements yielded in aqueous solution and a 100-fold lower number than expected for cytoplasmic diffusion rates (Bass *et al.*, 1996). This implies that the periplasm has a gel-like consistency. Reasons for the extremely reduced diffusion rates are the high content of protein and non-polymerised peptidoglycan. Moreover this may be due to the poor mobility within the murein sacculus because of a sieving effect and furthermore due to the severely reduced aqueous space within the periplasm.

Proteins residing in the periplasmic space fulfil a number of important functions. They are responsible for the detection and processing of essential nutrients and their transport into the cell. They promote the biogenesis of proteins entering this compartment along with compounds destined for incorporation into the peptidoglycan, outer membrane or capsular layers (Oliver, 1996). Based on the different functions the proteins can be divided into several categories. First the solute or ion binding proteins that function in conjunction with ABC-transporters (Oliver, 1996) or chemotaxis receptors. Second and third the catabolic and the detoxifying enzymes. Finally, the enzymes that promote the biogenesis of major envelope proteins as the fourth category (Oliver, 1996). A closer look at the last category reveals some quite interesting features among the proteins of the periplasm:

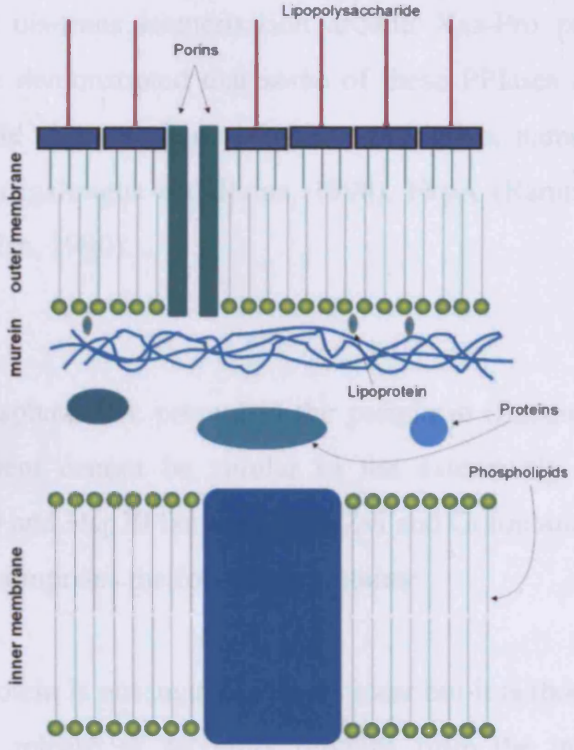

Fig. 2.1. The periplasm of *Escherichia coli*. The cell envelope of gram-negative bacteria like *Escherichia coli* is typically built up by an inner and an outer membrane. Both membranes consist of lipid bilayers and are separated by the periplasmic space. Transmembrane proteins transverse the outer membrane, so called porins, that act as hydrophilic channels and are responsible for the uptake of low-molecular, hydrophilic substances. The murein sacculus is located at the periplasmic side of the outer membrane.

a) Disulphide bond formation:

It can be clearly demonstrated, that the periplasmic compartment is an oxidizing environment whereas the cytoplasm is a reducing environment. Thus the formation of disulphide bonds after the passage across the inner membrane becomes possible. At the moment two separate pathways are known which are responsible for this task. The first system consists of the two proteins DsbA and DsbB (Bardwell *et al.*, 1993) while the second system consists of the recently described proteins DsbC and DsbD (Rietsch and Beckwith, 1998). DsbE is required for cytochrome c biogenesis (Sambongi and Ferguson, 1994) and the most recently described DsbG has chaperone activity in addition to its ability to catalyse disulphide bond formation (Andersen *et al.*, 1997).

b) Peptidyl-Prolyl-Isomerases:

These proteins catalyse cis-trans isomerisation around Xaa-Pro peptide bonds. However, recent experiments have demonstrated that some of these PPIases also facilitate folding of envelope proteins. At the moment four activities are known, namely SurA (Rouviere and Gross, 1996), PpiD (Dartigalongue and Raina, 1998), FkpA (Ramm and Pluckthun, 2000), PpiA/RotA (Liu and Walsh, 1990).

c) Chaperones:

As no nucleoside triphosphates are present in the periplasm (Rosen, 1987), chaperones that work in this compartment cannot be similar to the extensively studied ATP-dependent chaperones of the Hsp60 and Hsp70 families (Ben-Zvi and Goloubinoff, 2001). The family of periplasmic chaperones comprises the following proteins:

α) SecD protein:

The exact task of this protein is not unambiguously clear but it is thought that SecD promotes the proper folding and release of secretory proteins from the inner membrane into the periplasm (Pogliano and Beckwith, 1994).

β) Pilus-specific chaperones:

This kind of protein represents a large family of specialised periplasmic chaperones that are required for the assembly of pili and fimbriae (Holmgren *et al.*, 1992).

γ) LolA/LolB:

The incorporation of lipoproteins into the outer membrane is catalysed by LolA/p20, a periplasmic shuttle protein and the outer membrane lipoprotein LolB (Matsuyama *et al.*, 1995; Matsuyama *et al.*, 1997).

δ) HlpA protein (also named Skp or OmpH):

Skp has been proposed to be a general chaperone for outer membrane proteins (OMPs) (Chen and Henning, 1996).

ε) DegP (HtrA):

Spiess *et al.* (Spiess *et al.*, 1999) could demonstrate that this heat-shock protein exhibits general molecular chaperone activity in addition to its protease activity. A process which is controlled in a temperature-dependent manner.

2.2 Protein quality control in *Escherichia coli*

The appearance and maintenance of functional proteins within cells depends on more than the fidelity of transcription and translation. After initial folding and assembly, proteins may suffer damage in response to various stresses or insults (Wickner *et al.*, 1999). In this situation both proteases and chaperones serve to maintain quality control of cellular proteins. Any of the two protein families has to recognize regions that are commonly found on misfolded or unfolded proteins but not on native proteins (Wickner *et al.*, 1999). It can be demonstrated, that both systems recognize hydrophobic regions exposed on unfolded proteins (Wickner *et al.*, 1999). But how is the decision made if a protein gets refolded or degraded? It has been proposed that in procaryotic systems, these pathways function stochastically, i.e. the fate of the protein depends on the kinetics of interaction (binding and release) of the protein with molecular chaperones or the ATPase components of proteases (Wickner *et al.*, 1999). For several of these ATPase components, chaperone activity can be demonstrated (Wickner *et al.*, 1999). In *Escherichia coli*, there are at least five energy-dependent proteases known to date (Lupas *et al.*, 1997): ClpAP, ClpXP, HslUV (ClpYQ), Lon and FtsH.

2.2.1 Heat-shock proteins

When organisms are confronted with elevated temperatures in their environment, they react with the increased synthesis of a group of proteins commonly termed heat-shock proteins (Morimoto *et al.*, 1994). The majority of the heat-shock proteins play a fundamental role as molecular chaperones or components of proteolytic systems (Morimoto *et al.*, 1994). In different organisms, the response is induced at very different temperatures. In each case, the organism would be expected to cope with such temperatures in its natural environment (Lindquist and Craig, 1988). The heat-shock proteins are one of the most highly conserved systems known. However, heat is not the only folding stress which leads to an overexpression of this class of proteins. Other examples are anoxia, ethanol and certain heavy metal ions. It is important to note, that the term heat-shock protein is potentially misleading as the proteins are more or less essential for cell growth at all temperatures (Morimoto *et al.*, 1994).

2.2.2 The molecular chaperones

Anfinsen was the first one who could demonstrate, that all the information necessary for a polypeptide chain to fold correctly into a three-dimensional structure is encoded in its primary sequence (Anfinsen, 1973). However in living cells, a polypeptide chain is confronted with a situation which is dramatically different to the artificial environment of a test tube. Under cellular conditions of high protein concentration, cellular crowding and high temperatures, many proteins tend to form stable, insoluble aggregates devoid of biological activity (Ben-Zvi and Goloubinoff, 2001). Moreover the cell may have to master various stresses such as heat and cold shock, dehydration, oxidation, salt and osmotic shock and even mutations which all increase the possibility, that native proteins unfold or seek alternative stable, but in the end misfolded states (Ben-Zvi and Goloubinoff, 2001). To overcome the various folding problems, the cell is in possession of a variety of helper proteins that are collectively termed molecular chaperones. They are required for successful folding, assembly, transport and even degradation of proteins within the cell (Saibil and Wood, 1993). The chaperones are an abundant and ubiquitous protein family, many of which are heat-shock proteins (Saibil and Wood, 1993). They interact with non-native protein subunits, stabilise protein folding intermediates, prevent aggregation of misfolded proteins or are even able to disaggregate and reactivate various kinds of protein aggregates (Ben-Zvi and Goloubinoff, 2001).

There are several major classes of chaperones, namely the Hsp100 (ClpA/B/X, HslU), Hsp90 (HtpG), Hsp70 (DnaK), Hsp60 (GroEL) and small Hsps (IbpA/B) (Ben-Zvi and Goloubinoff, 2001). The corresponding *E.coli* chaperones are given in parentheses. Under physiological conditions, the chaperones fulfil important house-keeping functions, whereas during and following stress, their main task is a damage-control function (Morimoto *et al.*, 1994).

Among the best studied representatives of this family are the hsp60s and their partner proteins the hsp10s, which were named chaperonins (Tilly *et al.*, 1981). There are actually two subfamilies, namely the GroE chaperonins found in eubacteria, mitochondria and chloroplasts (Group I) and the TCP1 chaperonins found in archeaebacteria and in the eucaryotic cytosol (Group II) (Ditzel *et al.*, 1998). The most striking feature about the chaperonins is their three-dimensional structure. GroEL, the well known paradigm for a chaperonin, is an assembly of 14 identical subunits with a molecular weight of about 800 kDa. The proteins build a cylindrical oligomer, formed by two heptameric rings that are stacked back to back. Each ring encloses a central cavity with small side windows (Braig *et al.*, 1994). The misfolded substrate protein is refolded by binding to GroEL and subsequent release. The binding and

release of substrate is controlled by ATP and ADP (Saibil and Wood, 1993). There is still some debate going on about the mechanism, by which refolding is achieved. Basically, there are two main models under consideration (Ellis and Hartl, 1996). The first is the Anfinsen cage model that is based on the view, that protein folding is limited by intermolecular reactions that produce aggregation. In that model, the internal cavity of GroEL provides a sequestered microenvironment, where folding to the native state can occur, while the substrate protein is protected from aggregation (Ellis and Hartl, 1996). The iterative annealing model is based on the view, that the rate-limiting step in slow protein folding is the intramolecular reorganization of misfolded and trapped protein segments, dependent on some degree of protein unfolding (Todd *et al.*, 1996). In this model, ATP hydrolysis is coupled to forceful unfolding of the misfolded protein und its release either to the shielded internal cavity or to the exterior, so that misfolding is relieved and forward folding can resume (Shtilerman *et al.*, 1999).

2.2.3 The proteases

Proteases are an important class of enzymes that hydrolyse peptide bonds. They can either work as exo- or as endopeptidases. Exopeptidases cleave one, two or three residues from the N-or C-terminus of a polypeptide-chain. Therefore they are called aminopeptidases or carboxypeptidases, respectively. Endopeptidases cleave somewhere in the middle of a polypeptide chain, depending on their substrate specificity. According to the nomenclature of Rawlings and Barrett, endoproteases can be classified according to their activities and their functional groups as the cysteine proteases, the serine proteases, the aspartate proteases, the metalloproteases and the proteases with unknown function (Barrett, 1994).
Proteases perform a variety of fundamental biological activities. Examples include the precise processing of proteins in the blood clotting cascade (Fersht, 1999) or the indiscriminate degradation of abnormal proteins (Wickner *et al.*, 1999). Proteases are, furthermore, associated with a variety of diseases and therefore they are potential targets for new drugs (Fersht, 1999).

The serine proteases are probably the most extensively studied family of proteases. Based on sequence alignments and crystal structures, the serine proteases can be further separated into a number of subfamilies (Rawlings and Barrett, 1994). A comparison of chymotrypsin with

subtilisin, two of the best studied representatives of the serine protease family, points out the common features as well as the differences between the subfamilies. Although chymotrypsin and subtilisin show a radically different overall fold, their catalytic triad (His-Asp-Ser), oxyanion binding pocket and S_1 specificity site exhibit an astonishingly similar spatial organization (Fersht, 1999). At the moment, there are five distinct types of folds known among the serine proteases: the chymotrypsin family, the subtilisin family, the wheat serine carboxypeptidase II family, the human cytomegalovirus serine protease family with a His/His/Ser catalytic triad and the proteolytic component ClpP of the ATP-dependent protease family (Czapinska and Otlewski, 1999).

The following paragraph gives a short description of the catalytic mechanism and the main determinants of substrate specificity of serine proteases considering chymotrypsin as example. It should be noted, that this overview is mainly adopted from the books of Alan Fersht (Fersht, 1999) and Arthur Lesk (Lesk, 2001). The nomenclature corresponds to the respective chymotrypsin residues:

The catalytic mechanism

The hydrolysis of peptide and synthetic ester substrates by serine proteases is accomplished by the acylenzyme mechanism. The sidechain of the substrate at the residue N-terminal to the scissile bond binds in the specificity pocket (Fig. 2.2 and Fig. 2.3). The catalytic triad Asp102-His57-Ser210 positions and polarizes the sidechain of Ser195 for nucleophilic attack on the C atom of the scissile bond. An intermediate in which this carbon atom is tetrahedral is stabilized by the 'oxyanion hole' – in which there are hydrogen bonds from the negatively charged oxygen of the substrate (formerly the carbonyl oxygen of the peptide group of the scissile bond) to the NH groups of residues 193 and 195. A proton is transferred from Ser195 to His57, and releasing the C-terminal moiety of the substrate. The acyl-enzyme complex is then hydolysed by a similar mechanism: nucleophilic attack by the hydroxyl group of a water molecule which is hydrogen bonded to His57 releases the carboxylic acid product, restoring the enzyme to its initial state.

S_1 specificity pocket

The binding site for a polypeptide substrate consists of a series of subsites across the surface of the enzyme. However, very often only the primary binding pocket, also called the S_1 specificity pocket is well defined. The specificity of the pocket depends on its depth as well as on the electrostatic properties of the residues forming the surface of the pocket (Fig. 2.3). In case of the example depicted in figure 2.3, the specificity of thrombin is mainly caused by ASP189 at the bottom of the pocket, which makes hydrogen bonds to the guanidino group of the arginine residues of the choromethylketone inhibitor.

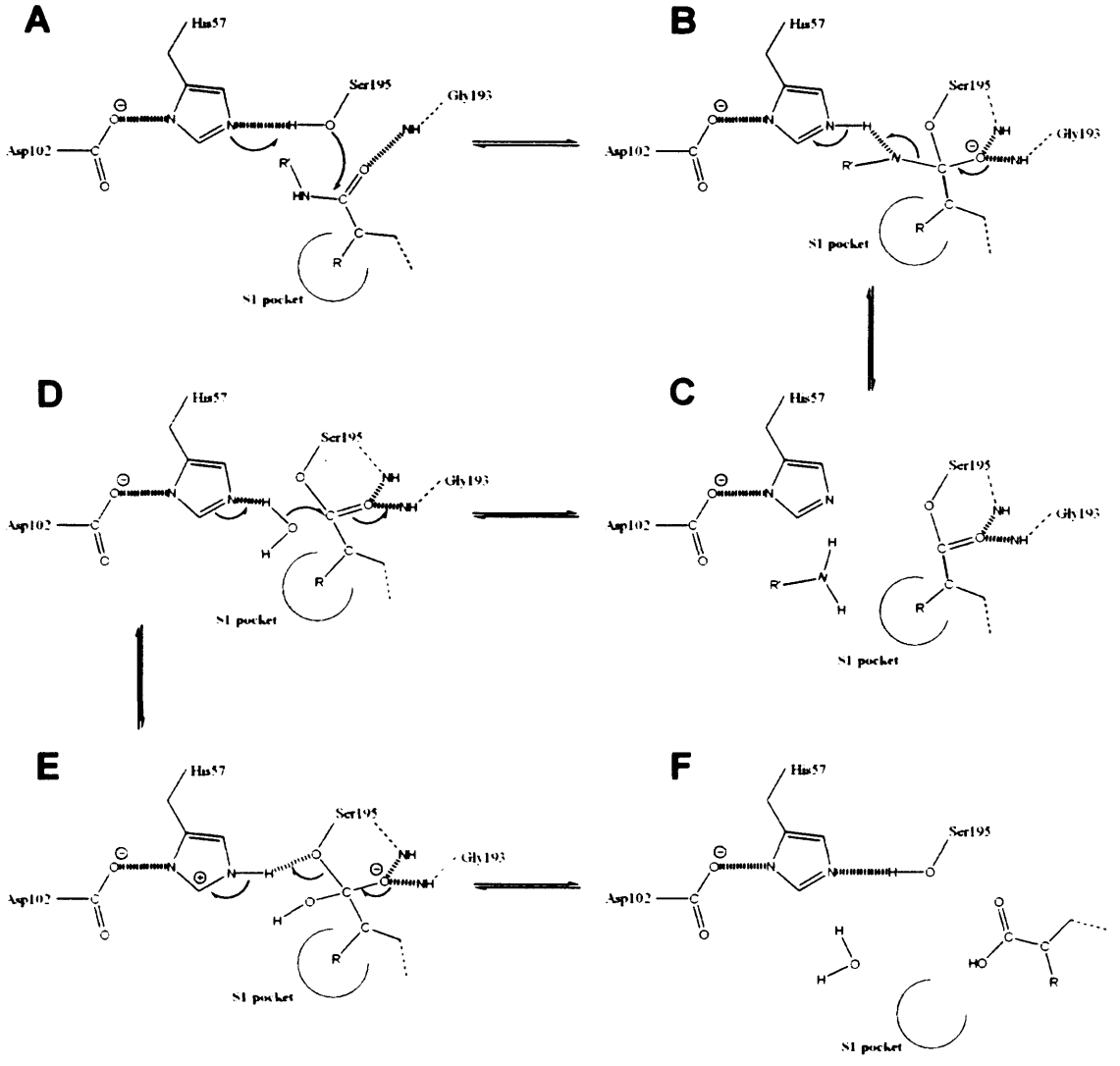

Fig. 2.2. The mechanism of catalysis of chymotrypsin-like serine proteases. A: Formation of Michaelis complex, B: Transition state, C: Acyl-enzyme intermediate, D + E: Transition state for deacylation, F: Product release and return to initial state. The scheme has been adopted with slight variations from Fersht (Fersht, 1999).

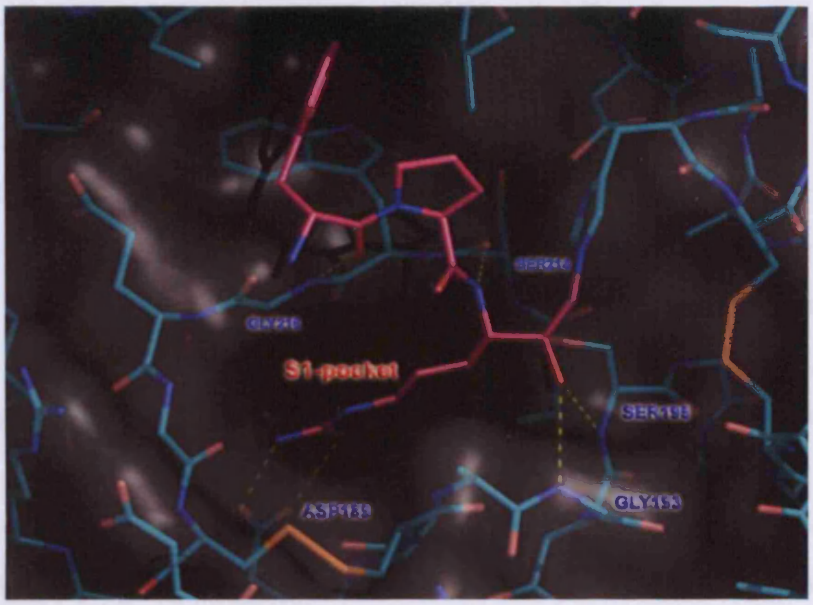

Fig. 2.3. The active site of chymotrypsin-like serine proteases. Human thrombin in complex with D-Phe-Pro-Arg-Chloromethylketone (Banner and Hadvary, 1991); residues of the protease are coloured in cyan, whereas residues of the inhibitor are coloured in purple. The protease domain is coated by a transparent van-der Waals surface. The Protein Data Bank entry code for the structure is 1DWE.

2.2.4 Cage-forming proteases

Beside other important biological functions, a special class of multi-subunit proteases is almost exclusively responsible for the removal of damaged or denatured proteins and the recycling of their amino acids (Schneider and Hartl, 1996). During the last years, extensive structural studies have revealed the molecular construction of these multi-subunit proteases, which all share the common feature of a relatively large central cavity, burying the proteolytic active sites (Yao and Cohen, 1999).

Three prominent examples, namely the proteasome (Groll *et al.*, 1997), ClpP (Wang *et al.*, 1997) and HslV (Bochtler *et al.*, 1997) have been analysed by X-ray crystallography. Although the subunit structure of ClpP on the one hand and of HslV and the proteasome on the other hand is strikingly different, they all show the same kind of quaternary structure (Lupas *et al.*, 1997). In all cases, the proteolytic subunits associate into multimeric rings that stack upon each other to form a barrel-shaped complex. Thus, the subunits enclose a central cavity, giving access to the active sites (Lupas *et al.*, 1997). This strategy prevents the potentially destructive activities of these proteases from harming the cell (Lupas *et al.*, 1997). Due to the limited dimensions of the axial pores, restricting access to unfolded substrate molecules, the protease provides selectivity and processivity (Lupas *et al.*, 1997). Interestingly

this kind of arrangement has also been found in the energy-independent proteases tricorn (Brandstetter *et al.*, 2001), bleomycin hydrolase (Gal6) (Joshua-Tor *et al.*, 1995) and very recently in the DppA aminopeptidase (Remaut *et al.*, 2001). In this context it is interesting to see, that the afore-mentioned family of chaperonins has converged to a similar quaternary structure despite a completely different evolutionary origin, subunit structure and mechanism. However, the protease complexes alone are not sufficient for the removal of abnormal proteins, as the protease components cannot take up a misfolded protein through their narrow axial pores (Fig. 2.4). The delivery of substrates to the internal proteolytic chamber with the active sites is accomplished by their corresponding ATPase components. For HslV, the corresponding ATPase is HslU and for ClpP it is ClpA or ClpX. The ATPase components also consist of oligomeric rings, that stack vis-à-vis on the protease complex (Ramachandran *et al.*, 2002). When the ATPase components where studied isolated *in vitro*, it was found out, that they have molecular chaperone activity (Wickner *et al.*, 1999). Recent studies have demonstrated that a substrate protein first gets unfolded in the substrate binding chamber of the ATPase subcomplex. Then the unfolded protein is translocated through a narrow axial channel to the protease complex where it gets degraded (Fig. 2.4) (Ishikawa *et al.*, 2001). In the absence of the corresponding protease complex, it is believed, that due to the complete unfolding of the substrate, it has now the possibility to overcome its previously kinetically trapped folding-incompetent conformation and to fold to its native, stable state (Wickner *et al.*, 1999). However, it is not clear if this process is relevant *in vivo* (Wickner *et al.*, 1999).

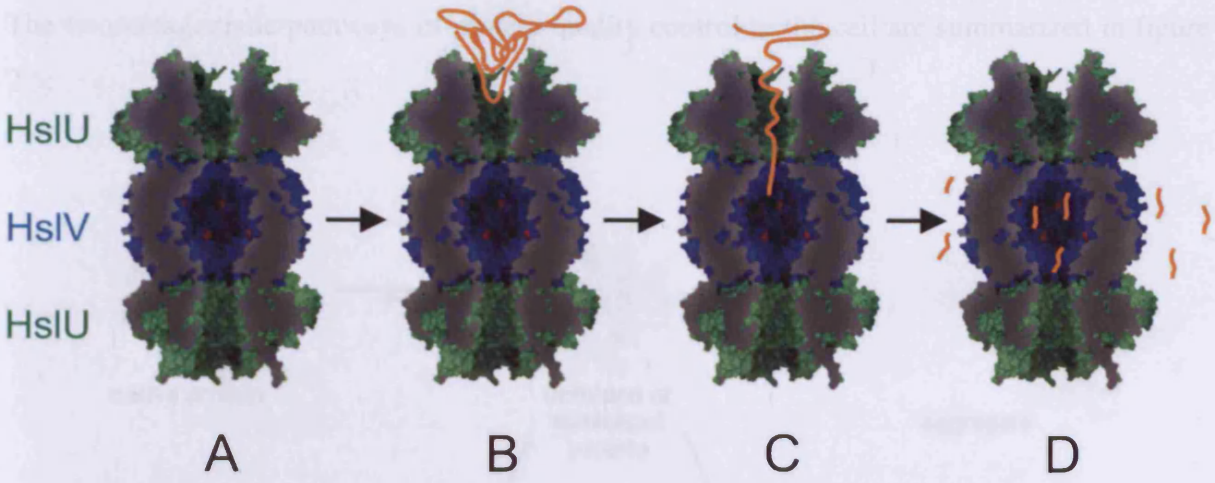

Fig. 2.4. Degradation of abnormal proteins by the HslVU protease. Interior view into the HslUV protease with the following colour code: ATPase subunits (HslU): green; protease subunit (HslV): blue; active site: red. In the scheme, the non-functional target protein (orange drawing) is first bound to the ATPase subunit (B). There the protein is unfolded by using ATP and subsequently translocated through the narrow internal channel to the protease subunit (C). In the proteolytic chamber, the protein is degraded to oligopeptides in an ATP-independent fashion (D). The Protein Data Bank entry code for HslUV is 1G3I.

2.2.5 Summary of protein quality control

The two antagonistic pathways of protein quality control in the cell are summarized in figure 2.5:

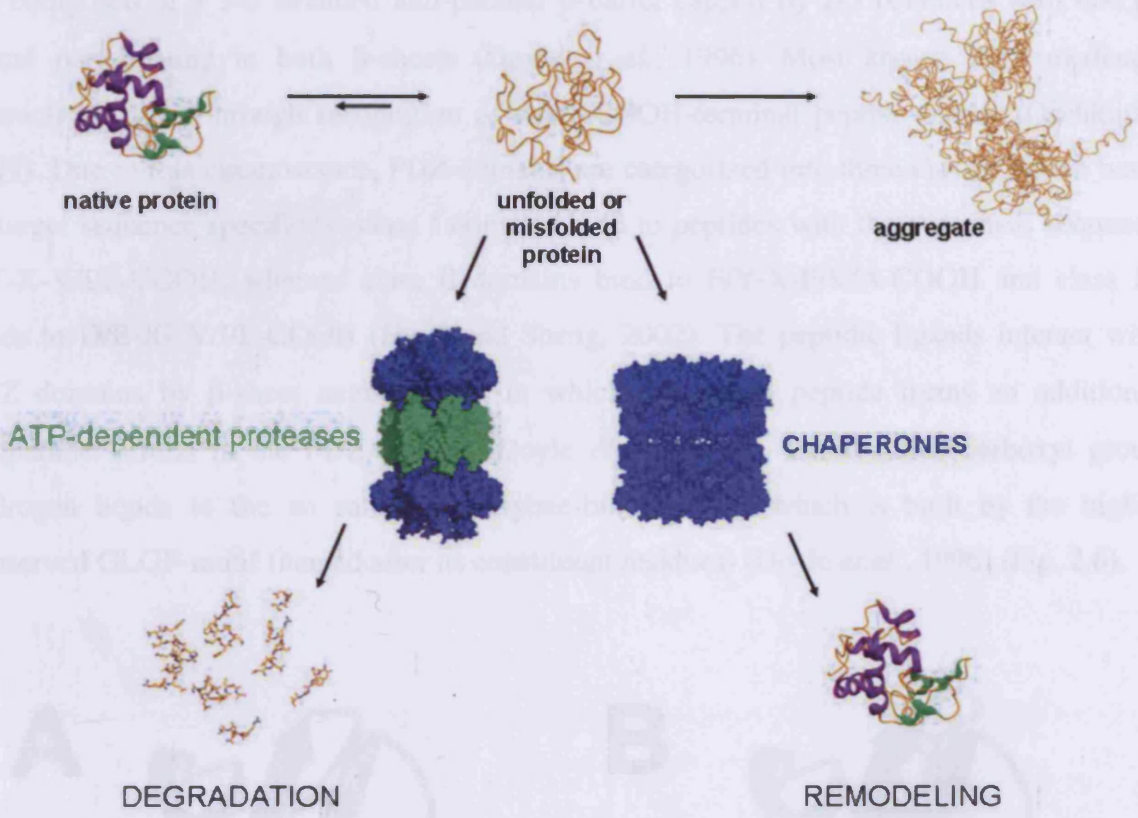

Fig. 2.5. Resurrection or destruction? Unfolded or misfolded proteins can either be refolded by molecular chaperones or degraded by cage-forming proteases. Both principles serve the same goal, namely the avoidance of protein aggregates, which are possibly lethal for the cell. Lysozyme (PDB entry code 2LYZ) represents the native enzyme, HslUV (PDB entry code 1G3I) the ATP-dependent protease and GroEL (PDB entry code 1GR5) the chaperone.

2.3 PDZ domains: modules for protein-protein interactions

PDZ-domains are conserved protein modules that mediate specific protein-protein interactions. Their name derives from the three proteins in which they were first discovered: **P**SD-95, **D**lg-1, and **Z**O-1 (Karthikeyan *et al.*, 2001). PDZ-domains play a central role in organizing diverse signalling pathways by organizing such protein networks and therefore PDZ scaffolding proteins increase the efficiency and specificity of signal transduction

(Fanning and Anderson, 1996). Further examples for domains that work in a similar way are the Src homology domains (SH2) and SH3 and the phosphotyrosine binding PTB domains (Songyang *et al.*, 1997).

PDZ-domains have a length of approximately 100 amino acids and are highly conserved. They are composed of a 5-6 stranded anti-parallel β-barrel capped by 2-3 α-helices with one β-strand participating in both β-sheets (Doyle *et al.*, 1996). Most known PDZ mediated interactions occur through recognition of short COOH-terminal peptide motifs (Oschkinat, 1999). Due to this circumstance, PDZ-domains are categorized into three classes on the basis of target sequence specificity: class I domains bind to peptides with the consensus sequence S/T-X-V/I/L-COOH, whereas class II domains bind to F/Y-X-F/V/A-COOH and class III binds to D/E-X- V/I/L-COOH (Hung and Sheng, 2002). The peptidic ligands interact with PDZ domains by β-sheet augmentation in which the bound peptide forms an additional antiparallel strand in the PDZ β-sheet (Doyle *et al.*, 1996). The terminal carboxyl group hydrogen bonds to the so called carboxylate-binding loop which is built by the highly conserved GLGF-motif (named after its constituent residues) (Doyle *et al.*, 1996) (Fig. 2.6).

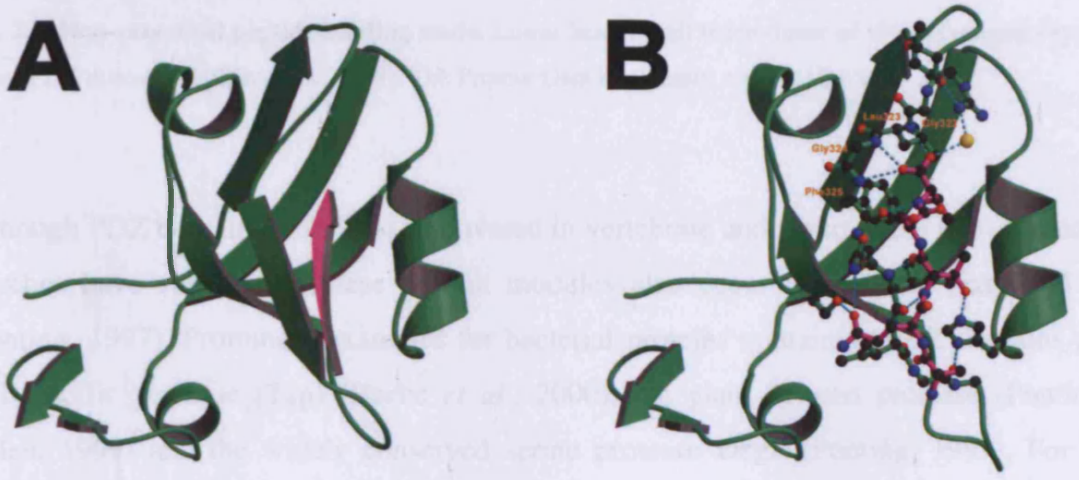

Fig. 2.6. Canonical peptide binding mode. A: Ribbon diagram of the third PDZ domain from the synaptic protein PSD-95 in complex with a peptide ligand (Doyle *et al.*, 1996). B: Chemical interactions involved in peptide binding. The PDZ domain is coloured in green and the ligand in purple. Hydrogen bonds are indicated by dotted lines coloured in cyan. The Protein Data Bank entry code is 1BE9.

Beside this canonical mechanism of target recognition, PDZ domains can also interact with internal sequences. The recently published crystal structure of a nNOS-syntrophin complex has revealed that the domains interact in an unusual head-to-tail arrangement (Hillier *et al.*,

1999). Linear association occurs because the nNOS PDZ domain has an unusual polarized structure with two distinct faces: a receptor face comprising the canonical PDZ domain with its peptide binding groove and a ligand comprising a β-hairpin "finger". This β-finger acts as a PDZ-ligand, docking into the PDZ binding groove of the syntrophin PDZ domain (Hillier *et al.*, 1999).

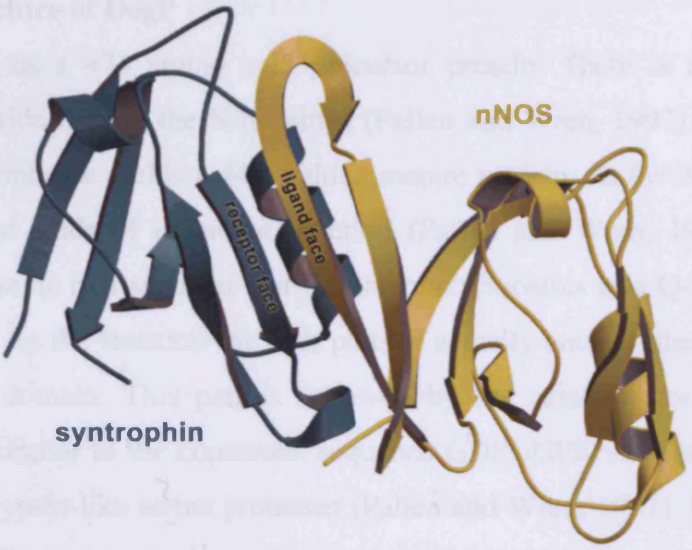

Fig. 2.7. Non-canonical peptide binding mode. Linear head-to-tail heterodimer of nNOS (yellow) - synthropin (blue) PDZ domains (Hillier *et al.*, 1999). The Protein Data Bank entry code is 1QAV.

Although PDZ domains were first discovered in vertebrate and invertebrate proteins, database searches have shown that these protein modules also occur in bacteria, yeast and plants (Ponting, 1997). Prominent examples for bacterial proteins containing PDZ domains are the Tail-specific protease (Tsp) (Beebe *et al.*, 2000), the giant Tricorn protease (Ponting and Pallen, 1999) and the widely conserved serine protease DegP (Ponting, 1997). For Tsp it could be shown, that a PDZ domain located approximately in the middle of the polypeptide chain is responsible for substrate recognition.

Very recently, it could be demonstrated, that peptide binding to the PDZ domain of the DegP homologue DegS leads to an activation of this protease. The main aim in that case is not substrate binding, but the mediation of a stress signal and the subsequent trigger of the protease activity (Walsh *et al.*, 2003).

2.4 The periplasmic heat-shock protein DegP

DegP was discovered nearly 20 years ago, when Swamy *et al.* (1983) undertook a systematic study of proteases in *Escherichia coli* (Swamy *et al.*, 1983). Synonyms for DegP are protease Do and HtrA. DegP is localized in the periplasm of *Escherichia coli* (Skórko-Glonek *et al.*, 1997).

2.4.1 Domain structure of DegP

DegP is translated as a 474 amino acid precursor protein. There is a signal peptidase I cleavage site 26 residues from the N-terminus (Pallen and Wren, 1997). Thus, translocation across the inner membrane yields a 448 residue mature protein. At the N-terminus there is a stretch of 50 amino acids of unknown function (Pallen and Wren, 1997), followed by a proline/serine/glutamine rich segment that has the characteristics of a Q-linker (Wootton and Drummond, 1989). As the function for both parts is actually unclear, they will be referred to as the N-terminal domain. This part is followed by the catalytic domain. The sequence G208NSGGAL is similar to the consensus sequence GDSGGPK surrounding the active site residues of many trypsin-like serine proteases (Pallen and Wren, 1997). Further evidence for the assertion that DegP is a serine protease comes from the fact that it is inhibited by diisopropyl fluorophosphate (DFP), a specific inhibitor for serine-proteases. Furthermore, mutational studies have demonstrated that a mutation concerning serine 210 or histidine 105 leads to a complete loss of protease activity (Skórko-Glonek *et al.*, 1995). Aspartate 135 has been postulated to be the third member of the catalytic triad (Pallen and Wren, 1997). Finally the C-terminal half of the protein consists of two PDZ domains (Ponting, 1997) (Fig. 2.8).

2.4.2 Structure of DegP

It is at least clear that DegP forms oligomeric complexes but investigations from different groups yielded different results: DegP has been proposed to exist in an equilibrium of hexameric and dodecameric species in solution (Kolmar *et al.*, 1996). Other groups showed that DegP is exclusively a hexamer (Sassoon *et al.*, 1999) or a dodecamer (Kim *et al.*, 1999), respectively. Oligomerization may allow allosteric interactions between the subunits or the presence of multiple cutting sites clustered within a single enzyme might increase processivity (Pallen and Wren, 1997). A proposed model of DegP contains two features: first the oligomerization results in the formation of a ring-like structure with an inner chamber containing the active sites. Therefore only unfolded substrates would be able to enter the

channel to gain access to the active sites where they can be degraded (Kolmar *et al.*, 1996). Furthermore Pallen and Wren (Pallen and Wren, 1997) extended the model to a so called 'anemone model'. They postulated that the PDZ domains of DegP might act as tentacular arms, binding to unfolded target proteins and then delivering them to the inner chamber.

In 1999 Kim *et al.* (Kim *et al.*, 1999) used electron microscopy to specify the three-dimensional architecture of DegP. The electron micrographs showed a dodecameric protein with two stacks of hexameric rings building a central pore with a diameter of 30Å. The overall dimensions of the particle are 120Å in diameter and 80Å along the sixfold axis. This arrangement is reminiscent of the HslV protease (Bochtler *et al.*, 1997). However, the crystal structure of DegP presented in this study will indicate a different quaternary state.

2.4.3 Physiological functions of DegP

The *htrA* gene has been identified by two phenotypes of *htrA* null mutants. These mutants were thermosensitive (Lipinska *et al.*, 1989) and showed a decreased degradation of abnormal periplasmic proteins (Strauch *et al.*, 1989). This suggests that the main physiological role of the protein is to degrade abnormal periplasmic proteins. Furthermore it could be demonstrated that DegP is a heat-shock protein which is under the exclusive control of the alternative σ^E factor (Lipinska *et al.*, 1988) and that DegP is indispensable for bacterial growth at temperatures above 42°C (Skórko-Glonek *et al.*, 1995). The only natural substrate identified so far is the colicin A lysis protein (Cavard *et al.*, 1989). Studies using artificial substrates indicated that there is a requirement for valine at the P_1 position (Kolmar *et al.*, 1996). There are severe conformational constraints for the substrate as it has to be at least partially unfolded so that cleavage can occur. Furthermore, the presence of intramolecular disulphide bonds in the denatured form of the oxidized protein must prevent degradation (Kolmar *et al.*, 1996).

In addition to the protease function Ehrmann and co-workers (Spiess *et al.*, 1999) could demonstrate that DegP also possesses general molecular chaperone activity. *In vitro* studies have shown that the chaperone function is present at low temperatures, whereas the proteolytic activity dominates at elevated temperatures. The implications for protein folding in the periplasm of *Escherichia coli* are not yet clear but a protease deficient mutant protein where the active site serine was replaced by alanine could suppress the temperature sensitive phenotype by overexpression in *degP* null mutants (Spiess *et al.*, 1999). Bass *et al.* (1996) initially proposed a chaperone activity for the DegP family. They argued that there is a HtrA

homolog in *Rickettsia tsutsugamushi* which completely lacks the residues building the catalytic triad and therefore appears to act in a protease independent manner (Bass *et al.*, 1996).

2.4.4 DegP homologues in *Escherichia coli*

There are two homologues of DegP in *Escherichia coli*, namely DegQ (HhoA) and DegS (HhoB) (Waller and Sauer, 1996). N.B. Hho stands for HtrA homolog. DegQ is a 455 amino acid protein with approximately 60% sequence identity to DegP. The overproduction of the protein suppresses the temperature sensitive defect of a *degP* null mutant strain. Furthermore DegQ also appears as an oligomer with a domain structure similar to DegP and it has obviously the same substrate profile as DegP (Kolmar *et al.*, 1996).

DegS is a 355 amino acid protein which is more different to DegP than DegQ. It has been proposed that the N-terminal part of the protein is an inner membrane anchor which is essential for the wild-type regulation of the σ^E pathway (Alba *et al.*, 2001). DegS is essential in *E.coli* as it provides the cell with active σ^E by degrading its σ^E specific anti-sigma factor RseA. In the absence of DegS, the σ^E pathway is not inducible and RseA remains stable (Alba *et al.*, 2001). A schematic overview of the domain structure of DegQ and DegS is given in Fig. 2.8.

2.4.5 DegP homologues in other organisms

There are DegP homologues ascribed in many other bacteria but also in plants and humans (Pallen and Wren, 1997). In nearly all cases, the three active site residues of the protease domain have been identified. Homologues in Gram-positive bacteria, in archaea and in cyanobacteria contain only one PDZ-domain. However, *degP* null mutations revealed overlapping phenotypes like thermosensitivity or an elevated sensitivity against oxidative and osmotic stress (Lipinska *et al.*, 1989; Strauch *et al.*, 1989).

At the moment four human homologues have been identified. L56 (HtrA1) is implicated in cell growth regulation. The C-terminus of the protein is homologous to the protease domain and the PDZ1 domain of the *E.coli* DegP. The N-terminus shows a different domain structure with an additional mac 25 domain and a kazal-type inhibitor motif. The mac 25 domain exhibits high homology to IGF binding proteins and contains a serine protease inhibitor motif (Hu *et al.*, 1998; Zumbrunn *et al.*, 1996) (Fig. 2.8). Omi (HtrA2) is localized in the endoplasmatic reticulum. The protein has several novel putative protein-protein interaction

motifs, a PDZ domain, a putative transmembrane region and a Src homology 3-binding domain repeated three times (Faccio *et al.*, 2000) (Fig. 2.8). Experiments have shown that Omi is associated with apoptosis (Suzuki *et al.*, 2001).

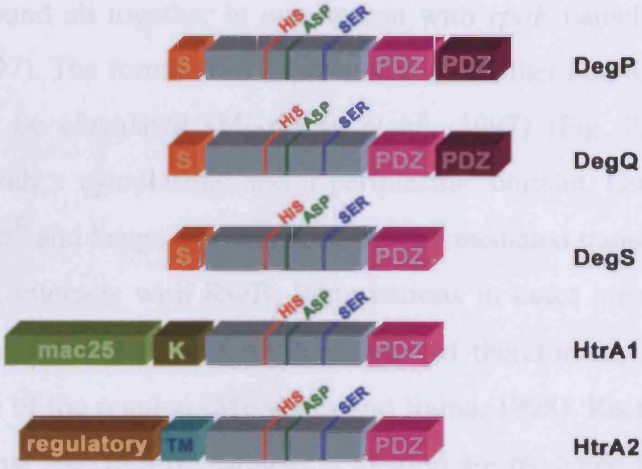

Fig. 2.8. Domain structure of HtrA family of proteins. Schematic representation of the *E.coli* proteins DegP, DegQ and DegS and of human HtrA homologous HtrA1 and HtrA2. Abbreviations: S: signal peptide, K: kazal-type inhibitor motif, TM: transmembrane region.

2.4.6 DegP and virulence

It has been demonstrated that a functional *degP*-gene is indispensable for the virulence of many pathogenic bacteria (Pallen and Wren, 1997). The reason for this is not clear. Probably it is their increased sensitivity to oxidative stress that makes them more susceptible to damage from the host's oxygen-dependent killing mechanisms (Pallen and Wren, 1997). However, it is also possible that other bactericidial mechanisms cause an accumulation of denatured proteins in the periplasm, which might lead to a lethal derangement in *degP* null mutants (Pallen and Wren, 1997).

2.4.7 Regulation of DegP

In *Escherichia coli* there are two types of pathways that respond to unfolding or misfolding of periplasmic proteins: the σ^E and the Cpx envelope stress response (Raivio and Silhavy, 1999). Both pathways exhibit a partial overlap in their downstream targets that are upregulated upon envelope stress. Interestingly, expression of DegP is upregulated by both mechanisms (Lipinska *et al.*, 1988; Danese *et al.*, 1995).

σ^E envelope stress response

The alternative sigma-factor σ^E is a member of the subclass of ECF sigmas, because they regulate extracytoplasmic functions (Missiakas and Raina, 1998). σ^E appears to be specifically induced by misfolded periplasmic proteins and the regulation of σ^E activity is ensured by three anti-sigma factors found all together in one operon with *rpoE* namely *rseA*, *rseB* and *rseC* (Missiakas *et al.*, 1997). The former two interact with each other but, for the latter, the precise function remains to be elucidated (Missiakas *et al.*, 1997) (Fig. 2.9). RseA is an inner membrane protein with a cytoplasmic and a periplasmic domain. On the cytoplasmic side, RseA interacts with σ^E and keeps it bound such that σ^E mediated transcription is inhibited. In the periplasm, RseA interacts with RseB. Perturbations in outer membrane protein folding lead to a destabilization of the RseA:σ^E complex and therefore to a release of σ^E and a subsequent induction of the regulon (Missiakas and Raina, 1998). Recent studies have shown that the conditions that lead to a σ^E pathway activation are mediated by proteolysis of RseA accomplished by the DegP homologue DegS (Ades *et al.*, 1999) (Fig. 2.9).

Cpx envelope stress response

Like the σ^E pathway, the Cpx envelope stress response senses misfolding of envelope proteins and upregulates the expression of corrective protein folding and degradation factors. In *Escherichia coli* the response consists of a two component regulatory system: the sensor histidine kinase CpxA and the cytoplasmic response regulator CpxR (Raivio and Silhavy, 1999). Subsequent studies have furthermore revealed a small periplasmic inhibitory factor with no informative homologues, namely CpxP (Danese and Silhavy, 1998). It has been anticipated that transduction of activating signals would occur through the conserved phosphotransfer reactions typical for this family of regulatory proteins: therefore inducing cues like misfolded proteins should lead to a conformational change in the sensor kinase resulting in the autophosphorylation of a conserved histidine residue. This is followed by a phosphotransfer to a conserved aspartate on the cognate response regulator (Raivio and Silhavy, 2001) (Fig. 2.9).

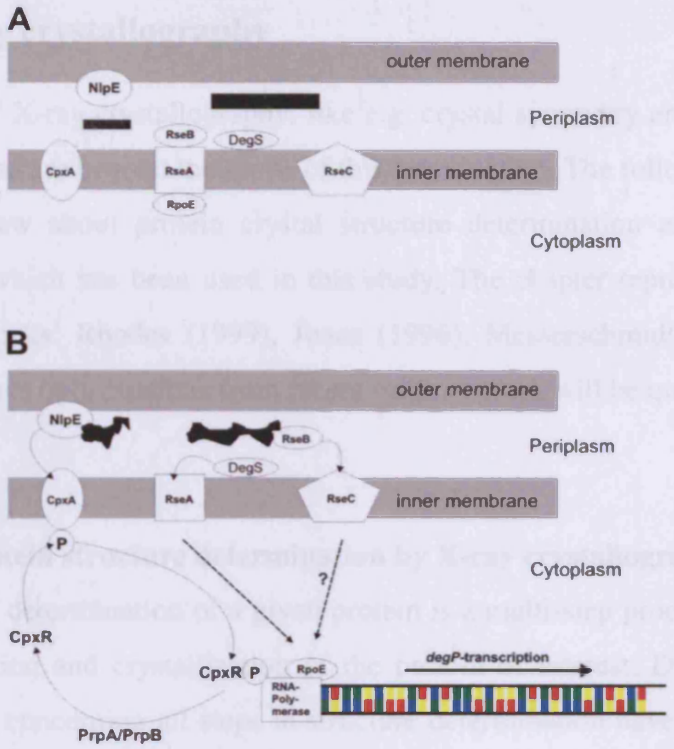

Fig. 2.9. Regulation of *degP*-transcription. In A, the cell envelope of *Escherichia coli* is shown under non-stress conditions. Native, functional proteins present in the periplasmic space are represented as black rectangles. The scenario under stress conditions is shown in part B. Proteins may unfold (here represented as shapeless, black objects) and therefore the two regulatory systems, namely the σ^E and the Cpx pathway are activated.

2.5 Protein X-ray crystallography

The fundamentals of X-ray crystallography, like e.g. crystal symmetry are described in detail in many textbooks and are beyond the scope of this introduction. The following paragraph will give a short overview about protein crystal structure determination and about the MAD phasing technique, which has been used in this study. The chapter represents a compilation from several text books: Rhodes (1999), Jones (1996), Messerschmidt and Huber (2000), Matthews (2001). Thus only citations from recent original work will be quoted.

2.5.1 Overview: protein structure determination by X-ray crystallography

The crystal structure determination of a given protein is a multi-step process starting with the expression, purification and crystallization of the protein of interest. During the last years, enormous advances, concerning all steps in structure determination have been made, namely the availability of high-energy synchrotron sources, modern CCD-detectors, the ever increasing computing power of PC's, better and more automated software packages and so on. However, despite of the sophisticated hard- and software available today, the main-obstacle still remains the technically simple process of crystallization. Though structural genomics initiatives try to circumvent the problem by the use of robots that are able to screen thousands of different crystallization conditions while consuming very little protein, it still remains a trial-and-error process, because it is impossible to predict the outcome of this multi-parameter process. As illustrated in figure 2.10, the process of a protein crystal structure determination can roughly be divided into three consecutive parts: crystallization/ data collection (step 1), phasing (step 2), model building/ refinement (step 3).

The first part has already been shortly outlined. The middle part, the phase determination, is probably the most demanding step of the whole process from a crystallographic point of view. This part will be discussed in more detail in the subsequent chapters. In the last part, the model building and refinement, a molecular model is first built into the electron density, which is obtained after successful phasing. The subsequent refinement process is then an iterative process whose goal is on the one hand to maximize the agreement between experimentally obtained raw data (the intensity of the individual reflections) to the model and on the other hand to build a model that is chemically realistic, i.e. it must possess bond length, bond angles, conformational angles and distances between neighbouring groups that are all in keeping with established principles of molecular structure and stereochemistry.

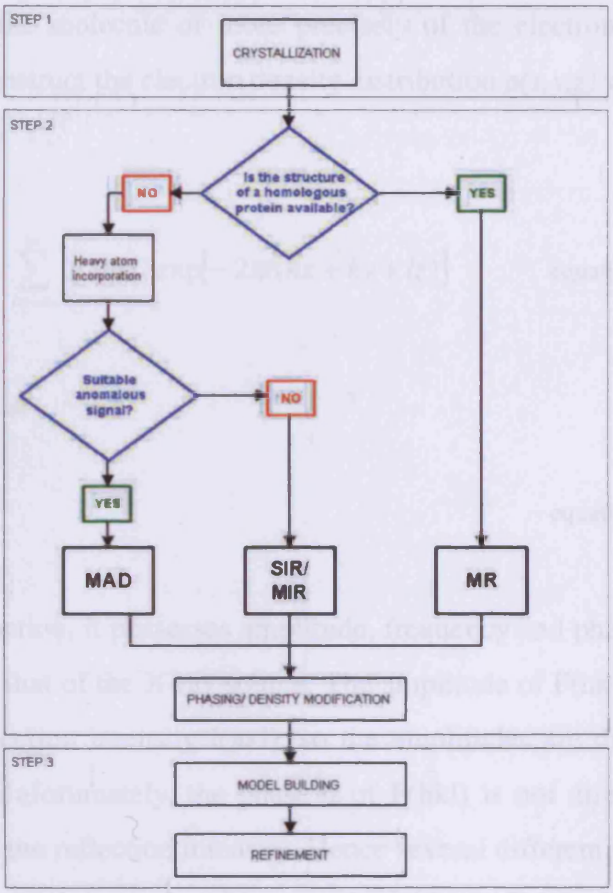

Fig. 2.10. Overview: protein crystallography. Abbreviations: molecular replacement (MR), single isomorphous replacement (SIR), multiple isomorphous replacement (MIR) multiple anomalous dispersion (MAD).

2.5.2 The phase problem

2.5.2.1 Starting point

As already mentioned, the aim of a crystallographer is to reveal the electron density distribution within the unit cell of a given crystal and use this as a template to reconstruct the corresponding molecular model of the crystallized molecule. Even though individual atoms diffract X-rays, it is still not possible to produce a focused image of a molecule, because X-rays cannot be focused by lenses. The electron density distribution in the crystal is a complicated periodic function and it can be described as a Fourier series. Each reflection observed in the diffraction pattern is the sum of the diffractive contributions from all atoms in the unit cell and all the reflections together compose the transform of the electron density distribution. The Fourier transform describes the mathematical relationship between an object and its diffraction pattern. Thus this is the lens-simulating operation the computer performs to

produce an image of the molecule or more precisely of the electron clouds in the crystal. Therefore, we can reconstruct the electron density distribution $\rho(x,y,z)$ within the unit cell with the following formula:

$$\rho(x,y,z) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F(hkl) \exp[-2\pi i(hx + ky + lz)] \qquad \text{equation (1)}$$

F(hkl) can be rewritten as:

$$F(hkl) = |F(hkl)| e^{i\alpha(hkl)} \qquad \text{equation (2)}$$

F(hkl) is a periodic function, it possesses amplitude, frequency and phase. It is a diffracted X-ray, so its frequency is that of the X-ray source. The amplitude of F(hkl) is proportional to the square root of the reflection intensity I(hkl), so the amplitudes are directly obtainable from measured intensities. Unfortunately, the phase α of F(hkl) is not directly obtainable from a single measurement of the reflection intensity. Hence several different methods were invented throughout the years in order to solve this central problem of crystallography, the so called *"phase problem"*. In the legend to figure 2.10, the three main methods are already stated, namely MIR, MR and MAD. The method called multiple isomorphous replacement is thereby the oldest method and was used by Kendrew and Perutz in the 1950s to solve the crystal structures of myolobin and haemoglobin, respectively. The molecular replacement (MR) technique gets more and more attractive as this method depends on the existence of a closely homologous structure, whose existence may become more and more likely with the increasing number of entries in the protein data bank (PDB). The last method, the multiple anomalous dispersion (MAD) method experienced its breakthrough during the last decade, although its theoretical foundation was known for many years. Today this method represents the routine approach for the solution of novel protein structures, but its broad applicability was dependent on the availability of specialized synchrotron beamlines. Just for completeness, a fourth method shall be mentioned whose application is restricted to the very rare case where crystals are available that diffract to ultrahigh resolution (≤ 1.2Å). Here the phase problem can be solved by *"direct methods"*, a technique routinely used in small-molecule crystallography.

2.5.2.2 Two-dimensional representations of structure factors

Each structure factor F(hkl) can be displayed as a vector on a two-dimensional plane of complex numbers of the form a + ib, where i is the imaginary number $(-1)^{1/2}$ (Fig. 2.11). The length of F is proportional to the square root of the measured intensity I, and the angle between F and the positive real axis is the phase α.

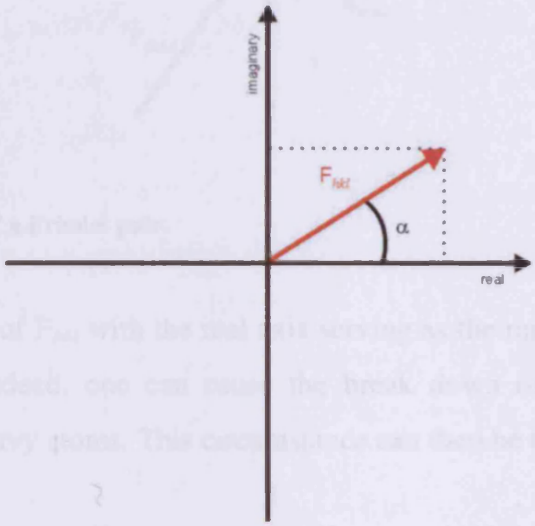

Fig. 2.11. The structure factor F, represented as a vector on the plane of complex numbers.

2.5.2.3 Friedel's law

All reciprocal lattices and therefore also all diffraction data possess a symmetry element called a centre of symmetry or a point of inversion, irrespective of the fact, that chiral molecules as proteins can only crystallize in non-centrosymmetric space groups. The reason for this is, that every reflection on a certain *hkl* plane causes a reflection on the corresponding *–h-k-l* plane and the intensity of both reflections is equal. The fact, that $I_{hkl} = I_{-h-k-l}$ is called Friedel's law. The pair of structure factors F_{hkl} and F_{-h-k-l} are called Friedel mates and even though their intensities are equal, F_{hkl} and F_{-h-k-l} are not (Fig. 2.12).

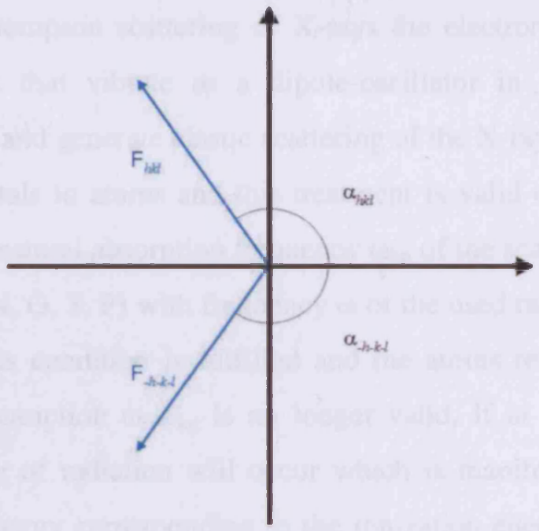

Fig. 2.12. Structure factors of a Friedel pair.

F_{-h-k-l} is the mirror image of F_{hkl} with the real axis serving as the mirror, so that Friedel mates have opposite phases. Indeed, one can cause the break down of the law by the specific introduction of certain heavy atoms. This circumstance can then be used for the solution of the phase problem.

2.5.3 Solution of the phase problem by multiple anomalous dispersion (MAD)

2.5.3.1 Anomalous diffraction

X-rays are diffracted by the electron clouds of the molecules in the crystal. The intensity of diffraction of X-rays from each type of atom is determined by the quantity f, called the scattering factor, which is proportional to the number of electrons in the atom and inversely proportional to the angle of diffraction. The structure factor for a certain reflection *hkl* is therefore a Fourier series in which each term is the contribution of one atom, treated as a simple sphere of electron density. So the contribution of each atom j to F_{hkl} depends on the kind of element, which determines f_j, the amplitude of the contribution, and its position in the unit cell (x_j,y_j,z_j), which establishes the phase of its contribution:

$$F_{hkl} = \sum_{j=1}^{n} f_j e^{2\pi i(hx_j+ky_j+lz_j)} \qquad \text{equation (3)}$$

So far, in the normal Thompson scattering of X-rays the electrons in the atom have been treated as free electrons that vibrate as a dipole-oscillator in response to the incident electromagnetic radiation and generate elastic scattering of the X-rays. However, the electrons are bound to atomic orbitals in atoms and this treatment is valid only if the frequency ω is large compared with any natural absorption frequency ω_{kn} of the scattering atom. For the light atoms in proteins (H, C, N, O, S, P) with frequency ω of the used radiation (in the wavelength range of 0.4 to 3.5Å) this condition is fulfilled and the atoms really scatter normally. For heavier elements, the assumption $\omega \gg \omega_{kn}$ is no longer valid. If ω is equal to an absorption frequency ω_{kn}, absorption of radiation will occur which is manifested by the ejection of a photo electron with an energy corresponding to the ionization energy for this electron. This transition goes to a state in the continuous region because the discrete energy states are all occupied in the atom. The absorption frequencies for the K, L or M shells are connected with the corresponding absorption edges which are characterized by a sharp drop in the absorption curve or a sharp increase in the fluorescence spectrum at the edge position, respectively (Fig. 2.13).

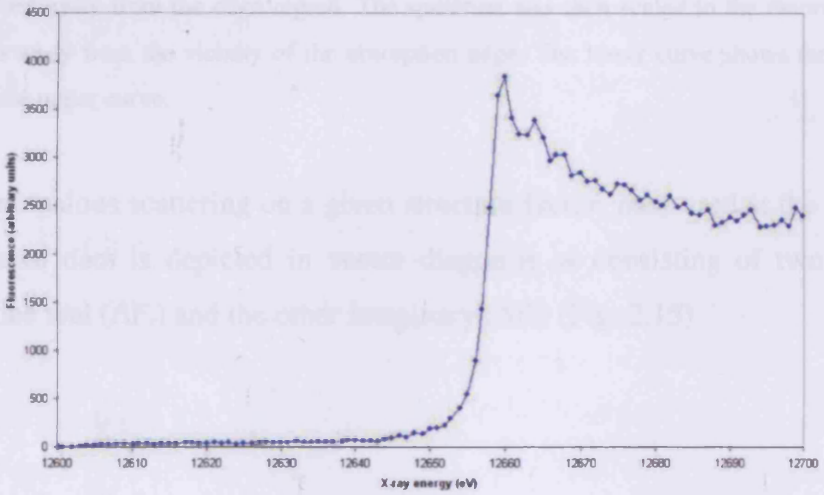

Fig. 2.13. Raw fluorescence spectrum from selenomethionine substituted DegP+DFP crystals measured at the Se K edge at the ESRF (Grenoble, France), beamline ID14-4.

The effects of anomalous scattering are described mathematically by two correction terms which are applied to the normal atomic form factor or Thompson scattering factor f. The modified scattering factor is described by $f=f_0+f'+f''$, where f' is the real part and f'' is the imaginary part of the anomalous scattering correction term. These atomic scattering factors vary most rapidly near the characteristic absorption edges of atoms, when the energy of the

incident X-rays is similar to the binding energy of the absorbing electrons (Fig. 2.14) (Evans and Pettifer, 2001).

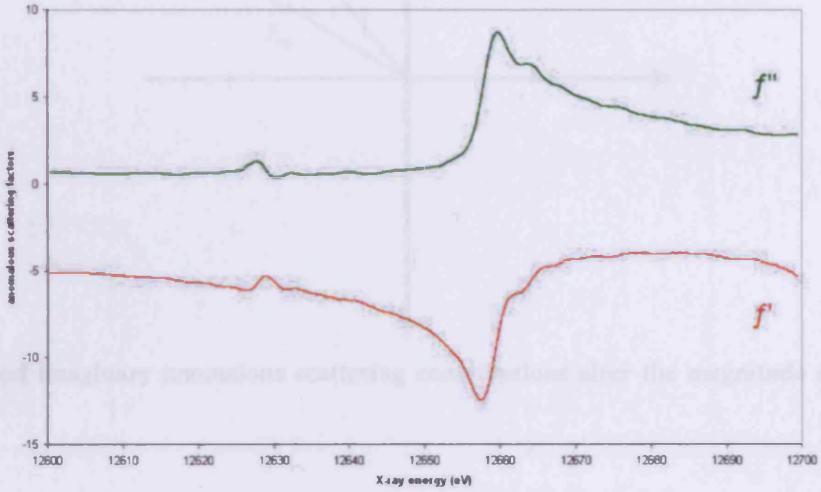

Fig. 2.14. Experimental values for the anomalous scattering factors from crystals of seleno-methionine substituted DegP+DFP. The values were calculated from the raw fluorescence spectrum shown in figure 2.13 with the program CHOOCH (Evans and Pettifer, 2001). The top curve shows the scaled and fitted absorption spectrum. Background correction was applied to the measured fluorescence spectrum by fitting a suitable polynomial to values away from the edge region. The spectrum was then scaled to the theoretically computed spectrum at points away from the vicinity of the absorption edge. The lower curve shows the Kramers-Kronig transformation of the upper curve.

The effect of anomalous scattering on a given structure factor, measured at the wavelength λ1 in the heavy-atom data is depicted in vector diagrams as consisting of two perpendicular contributions, one real (ΔF_r) and the other imaginary (ΔF_i) (Fig. 2.15)

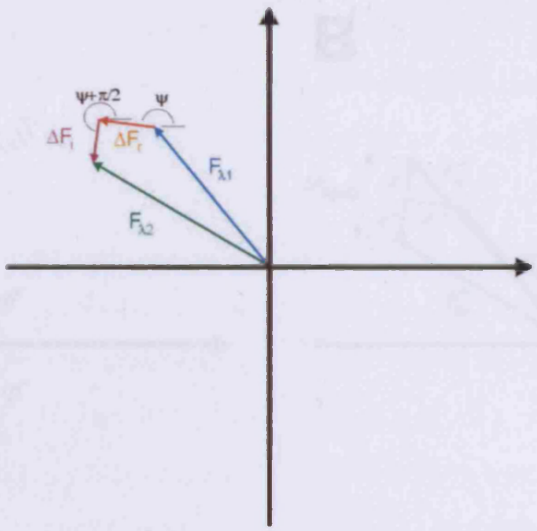

Fig. 2.15. Real and imaginary anomalous scattering contributions alter the magnitude and phase of the structure factor.

The quantity f', in phase with f_0, is usually negative, and f'', the imaginary part is always $\pi/2$ ahead of the phase of the real part (f_0+f'). It may be noted that by using different wavelength, the term f' is equivalent to a change in scattering power of the heavy atom and produces intensity differences similar to a normal isomorphous replacement, except that in this case the isomorphism is exact. This is the basis for the multiple-anomalous dispersion (MAD) method.

A consequence of the anomalous scattering is the already mentioned break down of Friedel's law (Fig. 2.16). The real contributions ΔF_r^+ and ΔF_r^- to the reflections of a Friedel pair are, like the scattering factors $F_{\lambda 1}^+$ and $F_{\lambda 1}^-$ themselves, reflections of each other in the real axis. The imaginary contribution to $F_{\lambda 2}^-$ is the inverted reflection of that for $F_{\lambda 2}^+$. That is ΔF_i^- is obtained by reflecting ΔF_i^+ in the real axis and then reversing its sign. Because of this difference between the imaginary contributions to each Friedel mate the two structure factors $F_{\lambda 2}^+$ and $F_{\lambda 2}^-$ are no longer precisely equal in intensity, nor are they precisely opposite in phase.

The relation between the structure factors of the reflection *hkl* and its Friedel mate *–h-k-l* is illustrated in figure 2.16 (A). The situation can be conveniently represented by reflecting the *–h-k-l* diagram through the real axis onto the *hkl* diagram. The difference $\Delta F_{ANO}=(F_{\lambda 2}^+-F_{\lambda 2}^-)$ is called the Bijvoet difference or simply the anomalous difference (Fig. 2.16 B).

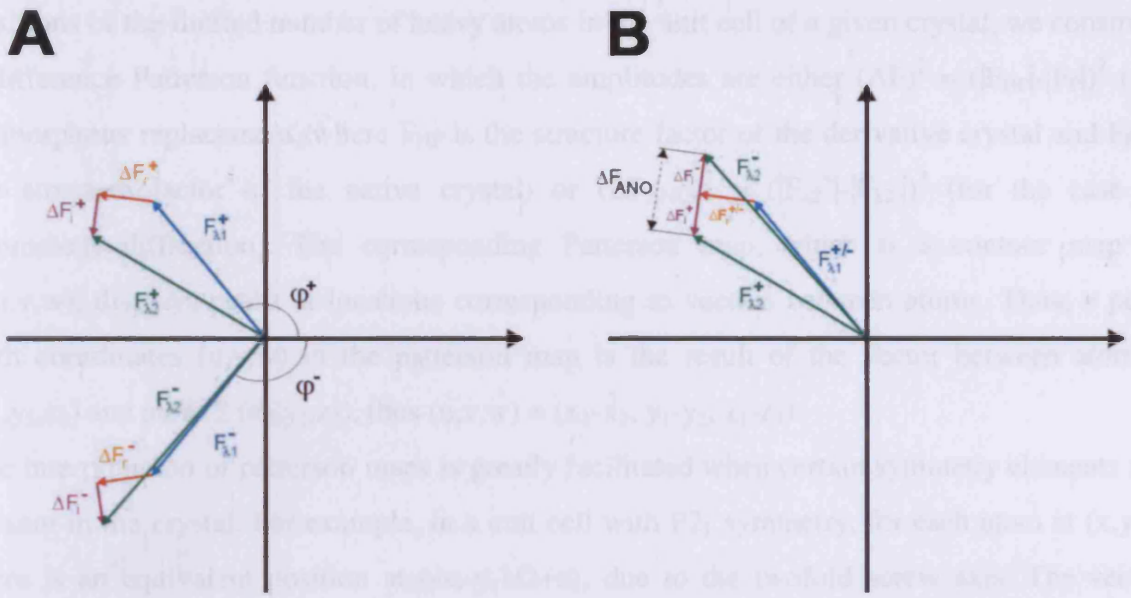

Fig. 2.16.Breakdown of Friedels law. A: Vector diagrams illustrating the effect of anomalous scattering on reflection *hkl* and *–h-k-l*. **B:** combined vector diagram for reflections *hkl* and *–h-k-l*.

2.5.3.2 Determination of the heavy atom substructure

2.5.3.2.1 Patterson methods

In 1935 A.L. Patterson published a classic paper where he presented the Patterson function P(u,v,w), a variation of the Fourier series used to compute ρ(x,y,z), where only the phaseless quantities $|F|^2$ are used as coefficients:

$$P(u,v,w) = \frac{1}{V}\sum_h\sum_k\sum_l \left|F_{hkl}^2\right| e^{-2\pi(hu+kv+lw)} \qquad \text{equation (4)}$$

In the usual synthesis with F's as coefficients (equation 1) we get the distribution of atoms in the cell. But in the phaseless series calculated with $|F|^2$ we get peaks corresponding to all the interatomic vectors. In contrast to small molecule crystallography, where the unit cell is relatively small and therefore only a limited number of interatomic vectors are present, this approach is not directly applicable in a protein crystal structure determination because of the enormous number of peaks representing vectors between light atoms in the protein.

To overcome this problem, heavy atoms are introduced into the protein crystal. This can be achieved either by soaking, co-crystallization or by incorporating them during protein expression using modified amino acids (e.g. selenomethionine). In order to determine the

positions of the limited number of heavy atoms in the unit cell of a given crystal, we construct a difference Patterson function, in which the amplitudes are either $(\Delta F)^2 = (|F_{HP}|-|F_P|)^2$ (for isomorphous replacement, where F_{HP} is the structure factor of the derivative crystal and F_P is the structure factor of the native crystal) or $(\Delta F_{ANO})^2 = (|F_{\lambda 2}^+|-|F_{\lambda 2}^-|)^2$ (for the case of anomalous diffraction). The corresponding Patterson map, which is a contour map of P(u,v,w), displays peaks at locations corresponding to vectors between atoms. Thus, a peak with coordinates (u,v,w) in the patterson map is the result of the vector between atom 1 (x_1,y_1,z_1) and atom 2 (x_2,y_2,z_2), thus $(u,v,w) = (x_1-x_2, y_1-y_2, z_1-z_2)$.

The interpretation of patterson maps is greatly facilitated when certain symmetry elements are present in the crystal. For example, in a unit cell with $P2_1$ symmetry, for each atom at (x,y,z) there is an equivalent position at (-x,-y,1/2+z), due to the twofold screw axis. The vector connecting such symmetry-related atoms will therefore lie at (u,v,w) = (2x,2y,1/2) in the Patterson map That means we can find the peaks on a plane that cuts the Patterson unit cell at w = ½. Such planes, which contain the Patterson vectors for symmetry-related atoms, are called Harker planes and the corresponding vectors are called Harker vectors. It is important to note that every Harker vector has to lie on a Harker plane, but not every vector on a Harker plane has necessarily to be a Harker vector. If more than one heavy atom is present in the asymmetric unit, the crystallographer, after having solved the position of the first heavy atom by the inspection of the harker planes, looks for additional heavy atom sites by searching for cross vectors between the first site and additional sites. Thus a model which describes the correct relationship between the heavy atom sites and the corresponding patterson map is successively established. In practice, the patterson maps may be quite noisy and therefore hard to interpret. Their interpretation is, for example greatly facilitated by trial-and-error software like RSPS (Knight, 2000) which is based on vector-search methods.

2.5.3.2.2 Direct methods

During the last years, with the advent of cheap and fast computers and with the ever increasing application of selenomethionine substituted crystals (meaning sometimes enormous numbers of heavy atom sites within the asymmetric unit of the crystal and therefore yielding almost not interpretable patterson maps), the direct methods approach known from small-molecule crystallography and represented by programs like Shake and Bake (Weeks and Miller, 1999) and SHELXD (Schneider and Sheldrick, 2002), became more and more popular.

Shake and Bake tries out random arrangements of atoms, simulates the diffraction pattern and compares the simulated patterns with those obtained from the crystal. In the case of a crystal that is diffracting to ultra-high resolution and harbouring only a small protein, this approach can be used for *ab initio* structure solution by restraining the arrangement of trial atoms to physically reasonable positions. However when using heavy atom difference data the requirements are not so high. Usually data with a resolution of 3.5Å is sufficient to solve the comparatively low number of heavy atom sites.

2.5.3.3 Phase determination

Once located, the heavy-atom parameters (xyz positions, occupancies and Debye-Waller thermal factors B) are refined and used to calculate a more accurate $|F_H|$ and its corresponding phase α_H (Taylor, 2003). The further procedure can most suitably be illustrated with the Harker construction (Fig. 2.17). First we can now draw the vector F_H of the heavy atom into the Argand diagram, because we know its phase and length. As we treated $F_{\lambda 2}^+$ and $F_{\lambda 2}^-$ as separated reflections, we know the absolute value of the anomalous difference ΔF_{ANO} and we know that the corresponding vector is always $\pi/2$ ahead of the vector F_H (see also figure 2.16 B). The endpoints of the two vectors are now the corresponding centre of a Harker phase circle with a radius corresponding to the magnitude of the respective Friedel mate. Note that this construction corresponds to the construction shown in figure 2.16 (B), with the starting point of vector F_H shifted to the origin. However, phase determination at this point is ambiguous, the phase circles have two intersections.

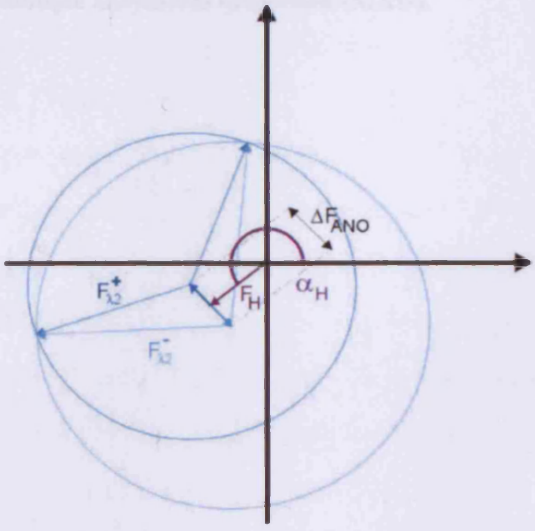

Fig. 2.17. Harker construction for single anomalous dispersion (SAD).

Thus in order to resolve this problem, a typical MAD experiment is performed at different wavelength, where the differences in f' and f'' are exploited to resolve the phase ambiguity (Fig. 2.18). The phase α_H of the heavy atom stays the same for each wavelength because it is only dependent on the position of the heavy atom in the unit cell. The difference in length is due to the different f' values. In the absence of errors, the phase circles now intersect at one point and allow the unique assignment of the protein phase angle which is necessary for the calculation of the electron density distribution within the crystals unit cell (Fig. 2.19). Typically a complete MAD experiment consists of data from three different wavelength. The first dataset is collected were f'' is maximal (peak), the second were f' is minimal (inflection) and the third at the high-energy side of the peak wavelength (remote). Up to now, only specialized synchrotron beamlines provide the opportunity to tune the wavelength of the incident beam.

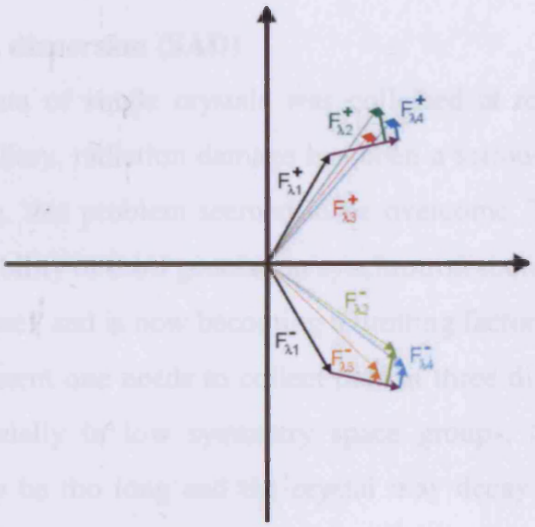

Fig. 2.18. Phase diagram for multiple anomalous dispersion (MAD).

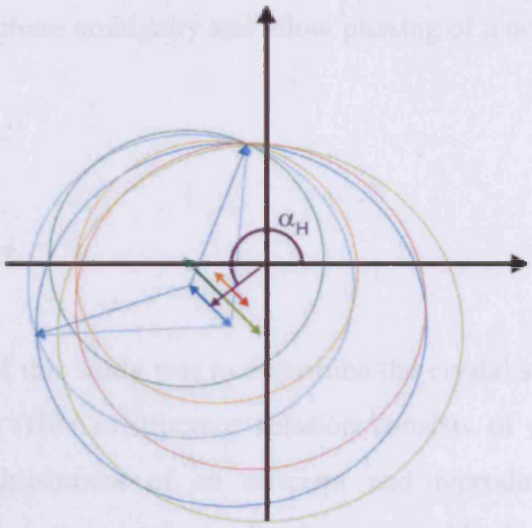

Fig. 2.19. Harker construction for multiple anomalous dispersion (MAD). Only the Friedel mates of one wavelength are depicted for the sake of clarity.

2.5.3.4 Single anomalous dispersion (SAD)

As long as diffraction data of single crystals was collected at room temperature with the crystal enclosed in a capillary, radiation damage has been a serious problem. However, with the advent of cryocooling, this problem seemed to be overcome. The situation has changed once again with the availability of third generation synchrotron sources with a high X-ray flux. Radiation damage came back and is now becoming a limiting factor again (Burmeister, 2000). For a usual MAD-experiment one needs to collect data at three different wavelength from a single crystal, thus especially in low symmetry space groups, the time required for the complete experiment may be too long and the crystal may decay just after data at the first wavelength has been collected. For that purpose, it is now becoming common practice to collect data only at the peak wavelength and use this to solve the structure by the single anomalous dispersion (SAD) method.

It is not formally possible to evaluate a protein phase exactly if there are only two experimental measurements, e.g. when the data are restricted to one wavelength with only a single anomalous difference available. Even assuming that the measured protein amplitudes, F^+ and F^-, and the calculated amplitude and phase contributions of the anomalous partial structure are error-free, there is a twofold ambiguity in the estimation of the protein phase as seen before. Nevertheless, probabilistic considerations led to the conclusion, that one of the two possible solutions is more likely than the other (Dauter *et al.*, 2002). Modern maximum likelihood-based phasing programs like SHARP (de la Fortelle and Bricogne, 1997) use this

fact in order to break the phase ambiguity and allow phasing of a novel structure from a single experiment.

2.5 Aim of the project

In the first part, the aim of this study was to determine the crystal structure of the periplasmic heat-shock protein DegP (HtrA). Structure solution consists of several consecutive steps, beginning with the establishment of an efficient and reproducible protein purification procedure. After purification, the protein can be characterized by N-terminal sequencing, mass spectrometry, size-exclusion chromatography and dynamic light scattering in order to detect any unwanted post-translational modifications, to elucidate the molecular weight of the oligomer in solution and the homogeneity of the sample. The crystallization of the protein can be performed by employing the widely used sparse-matrix screening methods. After production of diffraction quality crystals, the crystal structure of DegP can be solved by employing either multiple isomorphous replacement (MIR) or multiple anomalous diffraction (MAD) methods.

With the crystal structure of DegP, it should be possible to elucidate the organization of the oligomer, the structural determinants of the active site and the exact role of the PDZ domains. Moreover one could assign the areas important for chaperone activity and propose a model for the temperature-dependent switch between chaperone and protease activity.

In the second part of the study, the main focus is to look for inhibitors and substrates of DegP. It shall be tried to characterize their interaction with DegP and to co-crystallize them with DegP in order to specify the mode of action of this complex protease-chaperone system.

Chapter 3

Materials and Methods

3.1 Chemicals and their suppliers

antibiotics	Sigma
crystallization kits	Hampton Research
Ni-NTA-agarose	Qiagen
hydroxyapatite	BioRad
Superdex-200	PHARMACIA
further chemicals	Merck, Fluka, Serva, Sigma

3.2 Protease substrates, peptides and inhibitors

Substrates for protease assays:

Alpha-amylase (MalS), alkaline phosphatase (PhoA) and trehalase (TreA) were a kind gift from Dr. Alexandra Beil (Cardiff University). Resorufin-labeled casein was purchased from ROCHE.

Peptides:

All the peptides mentioned in this thesis with the exception of SPCJ-1 were synthezized by the bioorganic chemistry group at the MPI, purified by HPLC and their molecular weights were confirmed by electrospray-mass spectrometry. SPCJ-1 was a kind gift of C. Hal Jones (Wyeth Research, USA).

Inhibitors:

The inhibitors, DCI, DFP and Ms-AAPV-CMK were purchased from SIGMA. MG-115, MG-132 and PSI were purchased from SERVA and BisANS was purchased from Molecular Probes.

3.3 Crystallization equipment

All kinds of crystallization equipment like screens, plates, loops and so on were purchased from Hampton Research (USA).

3.4 Bacterial strains

Name	CLC198	CS83
Relevant genotype	MC4100 *degP::Tn10*	MC4100 *degP::Tn10, met*$^-$
Reference/source	Christine Cosma/ Ann Flower	Christoph Spiess

Tab. 3.1. bacterial strains

3.5 Plasmids

The plasmids used in this study have the following properties:

Name	*pCS20*	*pCS21*
Genotype	pQE60 derivative with *lac IQ*, *degP* (C-terminal His_6-tag)	pQE60 derivative with *lac IQ*, $degP_{S210A}$ (C-terminal His_6-tag)
Resistance	amp	amp
Reference/ source	(Spiess *et al.*, 1999)	(Spiess *et al.*, 1999)

Tab. 3.2. plasmids

3.6 Media

LB (Luria Bertani) medium:

10g Bacto-tryptone
5g Yeast extract
10g NaCl
made up to 1l with H_2O and autoclaved

NM-medium:

740ml H_2O
100ml 10x salts
100ml 10x amino acids
60ml 1mg/ml seleno-DL-methionine
autoclave H_2O, then add the other sterile filtered components

10x salts:

1M $(NH_4)_2SO_4$	75ml
5M NaCl	17ml
1M $MgSO_4$	10ml
1M Glucose	200ml
1mg/ml $CaSO_4*2H_2O$	10ml
Trace elements	10ml
10mg/ml Thiamine	10ml
10mg/ml Biotine	10ml
$FeNH_4(SO_4)_2*6H_2O$	1ml (add before use)
H_2O	65ml

Trace elements:

10mg/ml $MnCl_2$	5µl
10mg/ml $CuSO_4*5H_2O$	5µl
10mg/ml Na_2MoO_4	5µl
10mg/ml $ZnSO_4$	5µl

made up to 50ml with H_2O

10x amino acids:

H_2O	274ml
1M K_2HPO_4	476ml
1M KH_2PO_4	220ml
Alanine	0.5g
Arginine	0.5g
Aspartate	0.5g

Aspartic acid	0.5g
Cysteine	0.5g
Glutamic acid	0.5g
Glutamate	0.5g
Glycine	0.5g
Histidine	0.5g
Isoleucine	0.5g
Leucine	0.5g
Lysine	0.5g
Proline	0.5g
Serine	0.5g
Threonine	0.5g
Valine	0.5g
Amino acids II	30ml

Amino acids II:

H_2O	28.5ml
Phenylalanine	0.5g
Tryptophan	0.5g
Tyrosine	0.5g
32% (v/v) HCl	1.5ml

3.7 Microbiological methods

3.7.1 Sterilisation

Media and solutions were sterilised by autoclaving. In the course of this, they were heated to 120°C with a pressure of 1bar for 20min. Heat sensitive solutions were sterile filtered with a 0.22µm filter (Millipore).

3.7.2 Growth conditions

Liquid cultures were grown in Erlenmeyer flasks of appropriate size. In order to improve the oxygen supply, the flasks contained chicanes. The flasks were shaken at 140rpm and incubated at a temperature of approximately 30°C.

Recombinant *degP* and $degP_{S210A}$ were expressed in *E.coli* CLC198 cells. Both proteins contain a C-terminal His_6-tag protein. The cells were grown in 6l LB-medium at 30°C until they reached an OD_{600} of 1.0 followed by induction with 1mM IPTG for 12h. *E.coli* CS83 cells were used for the expression of DegP(SeMet) and $DegP_{S210A}$(SeMet). The rest of the expression was carried out exactly the same way as described previously, only NM-medium was used instead of LB-medium.

3.7.3 Determination of the cell density in liquid cultures

The cell density of liquid cultures was determined by measurement of the optical density in a photometer at a wavelength of 600nm. Distilled H_2O was used as a blank.

3.8 Biochemical methods

3.8.1 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used for the electrophoretic separation of proteins according to the method of Laemmli (Laemmli, 1970). The stacking and the separating gel were prepared as described below (Tab. 3.3). Gels were poured in an apparatus (MPI Martinsried, Germany) for 8 minigels. The protein samples were mixed with 5µl of sample buffer, boiled for 2min. and

loaded onto the gel. The electrophoresis was performed at 150V for about 60min. Afterwards, the gel could either be blotted (see 4.7.4) or stained with Coomassie brilliant blue (see 4.7.2).

component	separating gel (12.5%)	stacking gel (5%)
Buffer A	37.5ml	-
Buffer B	-	2.5ml
acrylamide stock	40ml	8.3ml
H_2O	21ml	38.4ml
10% (w/v) SDS	1ml	0.5ml
TEMED	80µl	40µl
10% (w/v) APS	500µl	250µl
Pyronine G	-	0.5ml

Tab. 3.3. Composition of polyacrylamide gels.

Buffer A: 1M Tris/HCl
pH 8.8

Buffer B: 2.5M Tris/HCl
pH 6.8

Acrylamide stock: 30% (w/v) Acrylamide
0.8%(w/v) Bisacrylamide

Sample buffer: 1.25ml 2M Tris/HCl, pH 6.8
0.8g SDS
3.2ml Glycerol (87%)
180µl 2-Mercaptoethanol
+ a few grains of bromophenolblue
filled with H_2O to 10ml

Running buffer: 25mM Tris/HCl
200mM Glycine
0.1% (w/v) SDS
pH 8.3

3.8.2 Coomassie blue staining

The polyacrylamide gels were stained by soaking in a staining solution and boiling in a microwave oven. Afterwards, the gels were gently shaken for at least one hour at room temperature. For destaining, the gels were transferred to a destaining solution and again boiled for several seconds and subsequently shaken for approximately half an hour at room temperature. The destaining procedure was repeated until the gel was completely free of background stain.

Staining solution:
2.5g Coomassie brilliant blue R-250
250ml ethanol
80ml acetic acid
filled with H_2O to 1l

Destaining solution:
1.25l ethanol
0.4l acetic acid
filled with H_2O to 5l

3.8.3 Determination of protein concentration

Two distinct methods have been used in order to measure the protein concentration of a given sample. Note that concentrations measured by the Bradford method are given in mg/ml, whereas the concentrations determined by absorbance at 280nm are given in molar concentrations.

3.8.3.1 Bradford protein assay

For the determination of the protein concentration according to the method of Bradford (Stoscheck, 1990) 795µl H_2O were mixed with 5µl of protein solution and 200µl of Bradford-solution (BioRad, Germany) in a plastic-cuvette. The Absorption of this mixture was measured in a photometer at a wavelength of 595nm against a blank solution containing 800µl H_2O and 200µl of Bradford-solution. The corresponding protein concentration was calculated from a calibration curve using BSA as a standard.

3.8.3.2 Absorbance at 280nm

The Absorbance of a protein sample at 280nm has been measured on a Beckman UV7500 in a quartz cuvette. A solution containing the respective buffer was used as a blank. The concentration was calculated according to the Beer-Lambert relation:

$$A = \varepsilon * c * d$$

A is the absorption at 280nm, ε is the molar absorbance coefficient, c the molar concentration of the protein solution and d is the cell length of the cuvette. The molar absorbance coefficient of DegP has been calculated as described by Gill and von Hippel (Gill and Vonhippel, 1989). According to that, the molar absorbance coefficient of DegP is $6400 M^{-1} cm^{-1}$.

3.8.4 Electroblotting of proteins to PVDF membranes

After gel electrophoresis the proteins could be transferred to a PVDF (polyvinylidene diflouride)-membrane (FluoroTrans, Pall, Ireland) for subsequent N-terminal sequencing. This was achieved by semi-dry electroblotting in a blotting apparatus (BioRad). Six layers of Whatman paper (Schleicher & Schuell, Germany) were soaked in Anode solution I and laid on the anode, followed by three layers of paper soaked in Anode solution II. The next layer consisted of the membrane, which was activated before use in 100% methanol. Then the SDS-gel was laid on the membrane, followed by six layers of paper, soaked in Cathode solution. The blot was performed at 50mA for 90min.

After the blot, the membrane was stained in staining solution (see 4.7.2) for one minute and afterwards destained in destaining solution for 15min. Then the membrane was air-dried so that the blotted bands became visible.

Anode solution I: 300mM Tris
20% (v/v) methanol

Anode solution II: 25mM Tris
20% (v/v) methanol

Cathode solution: 25mM Tris
40mM 6-Aminohexanoic acid
20% (v/v) methanol

3.8.5 Protein purification

3.8.5.1 Purification of DegP and $DegP_{S210A}$

Cells were harvested by centrifugation, resuspended in 100ml of buffer A and disrupted by sonication. N.B., from this point on, all steps of protein purification were performed at 4°C. The lysate was centrifuged (70000xg, 30min.) and the supernatant was loaded on a Ni-NTA resin (Qiagen) equilibrated with buffer A. Afterwards, the column was washed with buffer B to remove unspecific bound proteins. $DegP_{S210A}$ was eluted using buffer C. The eluted protein was dialyzed against 5l of buffer D for 3h. The dialysate was applied on a hydroxyapatite column (Biorad) and the protein was eluted with a linear gradient of 0 to 500mM potassium phosphate in buffer D. Eluted fractions were analysed by SDS-PAGE and separated into two species named as $DegP_{S210A}$ and $DegP_{S210A,12}$, respectively (see chapter 5). The samples were concentrated using an AMICON-cell (cut-off 30kDa) and individually applied to a Superdex-200 prep grade column (Pharmacia) equilibrated with buffer E. After gelfiltration fractions were combined according to their elution volumes: $DegP_{S210A}$ (151ml), $DegP_{S210A,12\text{-}B}$(132ml), $DegP_{S210A,12\text{-}A}$(128ml). The purified proteins were concentrated with a centriprep concentrator (Amicon, cut-off 30kDa). Then the purified samples were applied to a NAP-10 column (Pharmacia) and the buffer was exchanged to buffer F. Finally $DegP_{S210A}$ was concentrated using a centricon concentrator (Amicon, cut-off 30kDa) up to about 10-14mg/ml. The purification of selenomethionine derivatized $DegP_{S210A}$ was carried out in exactly the same way. $DegP_{S210A}$ prepared in the latter way will be referred to as $DegP_{S210A}$(SeMet).

3.8.5.2 Purification of DegP in complex with DFP

The purification for DegP inhibited by DFP was carried out the same way as for $DegP_{S210A}$ with the following modifications: after the Ni-NTA column, the protein was immediately applied to the HAP-column without previous dialysis and eluted as described in 4.7.5.1. The respective $DegP_{12}$ fractions were removed and only the hexameric form of DegP was used for complex formation.

Incomplete inhibition by DFP:

Approximately 20ml of 19µM DegP were warmed to 43°C, 625µl 10mM DFP in isopropanol were added and the sample was incubated for 2 hours. DegP+DFP was concentrated with a centriprep concentrator (Amicon, cut-off 30kDa), then applied to a NAP-10 column (Pharmacia) and the buffer was exchanged to buffer E. The next day, the protein was applied to a Superdex-200 column (Pharmacia) equilibrated with buffer E. The purified protein was concentrated with a centriprep concentrator (Amicon, cut-off 30kDa). Then the purified samples were applied to a NAP-10 column (Pharmacia) and the buffer was exchanged to buffer F. Finally DegP was concentrated using a centricon concentrator (Amicon, cut-off 30kDa) up to 12mg/ml. As it turned out, DegP was only slightly inhibited by this procedure, but showed a remarkable different behaviour during crystallization than the wild-type DegP. Therefore this protein will be referred to as $DegP_{(+DFP)}$.

Complete Inhibition by DFP:

DegP was dialyzed overnight against 5l of modified buffer D with a pH of 7.0. Approximately 20ml of DegP with a concentration of 23µM were warmed to 39°C, then four times 200µl of 10mM DFP in isopropanol were added with an interval of one hour between subsequent additions. After the last addition, the sample was incubated for another one and a half hour at 39°C, then DegP+DFP was stored at 4°C overnight. DegP+DFP was concentrated using an AMICON-cell (cut-off 30kDa) and applied to a Superdex-200 column (Pharmacia) equilibrated with buffer E. The purified protein was concentrated with a centriprep concentrator (Amicon, cut-off 30kDa). Then the purified samples were applied to a NAP-10 column (Pharmacia) and the buffer was exchanged to buffer F. Finally DegP+DFP was concentrated using a centricon concentrator (Amicon, cut-off 30kDa) up to 12mg/ml. The purification of the selenomethionine derivatized DegP+DFP was carried out exactly the same way. This protein will be referred to as DegP+DFP (SeMet).

buffer A: 200mM NaCl
100mM HEPES/NaOH
pH 8.0

buffer B: 40mM Imidazole
200mM NaCl
100mM HEPES/NaOH
pH 8.0

buffer C: 150mM Imidazole
200mM NaCl
100mM HEPES/NaOH
pH 8.0

buffer D: 50mM NaCl
50mM HEPES/NaOH
pH 8.0

buffer E: 300mM NaCl
50mM HEPES/NaOH
pH 8.0

buffer F: 10mM HEPES/NaOH
pH 8.0

3.8.6 N-terminal sequencing

N-terminal sequencing was performed by Edman-degradation. The analyses were performed at the Max-Planck Institute for Biochemistry (Martinsried, Germany) in the department Proteinanalytik by Dr. K.-H. Mann.

3.8.7 Mass spectrometry

Electrospray-ionization mass spectrometry (ESI-MS) was carried out on purified samples of DegP, DegP+DFP, DegP+DFP(SeMet), $DegP_{S210A}$ and $DegP_{S210A}$(SeMet). Furthermore several crystals of $DegP_{(+DFP)}$ were washed with water and subsequently dissolved in 5mM HEPES/NaOH, pH 8.0 by crushing them with a pipette tip. Remaining crystal debris was removed by centrifugation, afterwards mass spectrometric analysis was carried out. The analyses were performed at the Max-Planck Institute for Biochemistry (Martinsried, Germany) in the group of Dr. F. Siedler.

3.8.8 Size Exclusion Chromatography (SEC)

The molecular weight of $DegP_{S210A}$, $DegP_{S210A,12\text{-}A}$ and $DegP_{S210A,12\text{-}B}$ in solution was determined by analytical gel filtration chromatography on a Superose-6 3.2/30 column (Pharmacia) and a Superose-12 3.2/30 column (Pharmacia), respectively. The columns were equilibrated with 300mM NaCl, 50mM HEPES/NaOH, pH 8.0 at room temperature. Protein samples with a concentration of about 1mg/ml were applied to the columns at a flow rate of 50µl/min.

A calibration curve for molecular weight determination was set up as described. Blue Dextran was used to determine the void volume (V_0) of the respective column. Then the elution volumes (V_E) of the following molecular weight standards were measured: ferritin (450kDa), aldolase (158kDa) and chymotrypsinogen (25kDa). The total volume of the column (V_T) was calculated according to its length and diameter. The K_{av} value for every molecular weight standard was calculated:

$$K_{av} = \frac{V_E - V_0}{V_T - V_0}$$

The K_{av} value for the molecular weight standards was then plotted against the logarithm of their molecular weight and the points were fit to a linear equation. For proteins with unknown molecular weight, the K_{av} value was determined and the corresponding molecular weight was calculated from the calibration curve. The K_{av} value has the advantage that it is pressure independent. The analyses were carried out on a SMART-system (Pharmacia).

3.8.9 Dynamic Light Scattering (DLS)

DLS was carried out using a DynaPro-801(Protein-Solutions Inc.) molecular sizing instrument. A 50µl sample of $DegP_{S210A}$ with a concentration of about 2mg/ml in 10mM HEPES/NaOH, pH8.0 was filtered through a 0.22µm pore-sized membrane (Whatman) into a 12µl chamber quartz cuvette. The measurements were performed at 20°C. The data were analysed using the Dynamics 4.0 software (Protein-Solutions Inc.) as described by Moradian-Oldak *et al.* (1998) (Moradian-Oldak *et al.*, 1998).

3.8.10 Protease assay

3.8.10.1 Synthetic substrates

The protease substrate N-(Methoxysuccinyl)-Ala-Ala-Pro-Val 4-nitroanilide (SIGMA) was used in order to assess the activity of DegP against synthetic protease substrates. For that purpose 485µl of a solution containing 25mM Tris/HCl pH7.2, 50mM NaCl, 5mM $MgCl_2$ and 360µM DegP was mixed with 15µl of 20mM substrate (dissolved in DMSO) in a plastic cuvette. A blank was prepared the same way, but without DegP

The cuvettes were placed in an UVIKON 948 double beam spectrophotometer, with a cell holder connected to a thermostat, equilibrated at 37°C. The wavelength was set to 405nm and the increase of the absorbance of the sample cell against the reference cell was measured.

3.8.10.2 Resorufin-labelled casein

The proteolytic activity of DegP and its derivatives was determined with resorufin-labeled casein (Roche, Germany). Fifteen microliters of 0.4% (w/v) resorufin-labelled casein was added to 100µl incubation buffer containing approximately 1µM DegP and incubated at 37°C for several hours. The reaction was stopped by precipitation of casein with 480µl 10% (w/v) TCA. Samples were again incubated for 10 min. at 37°C and subsequently centrifuged (10 min., 10000 x g, RT). 400µl of the supernatant was mixed with 600µl 1M Tris/HCl, pH 8.8 to determine the absorbance at 574nm. A sample without DegP was used as a blank.

incubation buffer: 100mM Tris/HCl
100mM NaCl
pH 7.5

3.8.11 Sensitivity of DegP to various inhibitors

The inhibitory effects of DFP, BisANS, DIC, MG-115, MG-132, PSI and Ms-AAPV-CMK, on DegP have been analysed. Stock solutions with a concentration of 10µM were made by dissolving each inhibitor in DMSO. The stock solutions were then diluted in incubation buffer (see 3.8.10.2) containing 1µM DegP to the specified concentration. A reference sample was prepared with the same amount of solvent and DegP but without inhibitor. Every sample, the probe with the corresponding inhibitor as well as the reference was prepared three times. The samples were preincubated for 30 min. at 37°C, then 15µl of 0.4% (w/v) resorufin-labeled casein were added and the proteolytic activity was determined as described under 3.8.10.2 Average absorption values for the probe and the reference samples were calculated and the remaining activity of the inhibited probe was related to the reference.

3.8.12 Protease assays in the presence of various peptides

Cleavage assays concerning the influence of certain peptides on the protease activity were performed by preparing a solution in incubation buffer (see 3.8.10.2) with a final concentration of 1µM DegP and 100µM of the respective peptide. Again, reference as well as blank samples were prepared containing no peptides or no DegP, respectively. All the samples were preincubated for 30 min. at 37°C, then 15µl of 0.4% (w/v) resorufin-labelled casein was added and the proteolytic activity was determined as described under 3.8.10.2. Every sample, the probe with the corresponding inhibitor as well as the reference was prepared three times. Average absorption values for the probe and the reference samples were calculated and the remaining activity of the inhibited probe was related to the reference. The following peptides were analysed:

NH_2-DNRDGNVYQF-COOH (OmpC[1]), NH_2-NTDNIVALGLVYQF-COOH (OmpC[2]), NH_2-KSMCMKLSFS-COOH (SPCJ-1)

3.8.13 Effect of DFP on DegP at various temperatures

100µl of 8µM DegP in incubation buffer and 10mM DFP were incubated at 4°C, RT (≈25°C), 30°C, 37°C, 42°C and 48°C. A set of reference samples were prepared and treated the same way but without DFP. All probe and reference samples at every temperature were made as triplets. After 3h, non-reacted DFP was removed by precipitating DegP with 800µl of icecold acetone and the samples were put for 3h at -20°C. Then the samples were centrifuged for 15min. at 5000 x g, the supernatant was discarded and the pellet dissolved in 100µl incubation buffer. Fifteen µl 0.4% (w/v) resorufin-labeled casein was added and all the samples were incubated overnight at 37°C. The next day, the protease assay was developed as described under 3.8.10.2.

3.8.14 Analyses of degradation products by mass spectrometry

a) Degradation of substrate proteins

The substrates MalS, PhoA and TreA were present in a buffer containing 8M Urea with concentrations of 6mg/ml, 29mg/ml and 23mg/ml, respectively. These proteins were diluted in an assay buffer composed of 20mM HEPES/NaOH, pH8.0, 5mM $MgCL_2$, 10mM DTT and 1µM DegP to a final Urea concentration of 400mM. The samples were incubated at 37°C overnight. Citrate synthase as a fourth substrate protein was digested in 20mM Tris/HCl, pH 7.5, 1mg/ml citrate synthase and 1µM DegP. This assay was incubated at 43°C overnight. After the assays were finished, the samples were passed on to the mass spectrometry department at the MPI.

b) Liquid chromatography and mass spectrometry

The analyses of the samples were carried out by Dr. Frank Siedler from the service department mass spectrometry at the MPI. Briefly, the degradation products were separated by reversed-phase chromatography and subsequently applied to a Reflex III, MALDI-TOF mass spectrometer (Bruker Daltonics). Afterwards, individual peptides could be identified according to their molecular weight. The analyses of the raw mass spectrometry data were performed with Xcalibur mass spectrometry data system software (Thermo Electron Corporation).

3.8.15 Time-dependent degradation of substrate proteins

Solutions of 2µM DegP in 50mM MES/NaOH, pH6.5, 50mM NaCl, 5mM $MgCl_2$ containing 1.2mg/ml resorufin-labeled casein, or 1mg/ml citrate synthase were incubated at 37 or 43°C, respectively. At certain time points a sample was taken from the respective experiment, mixed with 6.6M Guanidiniumhydrochchloride to obtain a final concentration of 1M and stored at 4° for further analyses by reversed-phase (RP) HPLC.

The analysis was made by Markus Schütt and Dr. Cyril Boulegue from the department of bioorganic chemistry at the MPI. HPLC was carried out on a X-Terra-MS C4 RP column. The column was equilibrated with 5% of acetonitrile and 95% of 2% H_3PO_4 and eluted with a linear gradient of 5-90% of acetonitrile and 95-10% of 2% H_3PO_4 in 45 minutes at a flow rate of 1 ml/min. Degradation products were detected by UV at 210nm,

3.9 Crystallographic methods

3.9.1 Protein crystallization

$DegP_{S210A}$:

For initial surveys of crystallization conditions, standard screening kits purchased from Hampton Research (USA) were used, i.e. Crystal Screen I and Crystal Screen II. Therefore $DegP_{S210A}$ at approximately 12mg/ml was used with typically 2µl of protein solution mixed with 1µl of reservoir solution set up as sitting-drop vapour-diffusion experiments using cryschem plates (Charles Supper Company, Massachusetts) with 500µl of reservoir solution at 4°C and 18°C, respectively. The initial condition consisting of 10% (v/v) isopropanol, 0.1M HEPES/NaOH, pH 7.5, 10% (w/v) PEG 4000 was subsequently refined by a systematic variation of precipitant, buffer, pH, drop ratio and by addition of certain additives to the drop mixture.

$DegP_{S210A}$(SeMet):

Crystals of $DegP_{S210A}$(SeMet) were grown using sitting drops equilibrated against a reservoir of 10% (v/v) isopropanol, 0.1M TRIS/HCl, pH 8.5, 10% (w/v) PEG 2000 MME. Drops were

set up with 2µl $DegP_{S210A}$(SeMet) (12mg/ml), 0.3µl 100mM $MgCl_2$ and 1µl reservoir solution.

DegP+DFP:

Crystals of DegP in complex with DFP were grown using sitting drops equilibrated against a reservoir of 10% (v/v) isopropanol, 0.1M TRIS/HCl, pH 8.5, 10% (w/v) PEG 2000 MME. Drops were set up with 2µl DegP+DFP (12mg/ml), 0.3µl 100mM $MgCl_2$ and 1µl reservoir solution.

DegP+DFP(SeMet):

Crystals of DegP in complex with DFP were grown using sitting drops equilibrated against a reservoir of 10% (v/v) isopropanol, 0.1M TRIS/HCl, pH 8.5, 10% (w/v) PEG 2000 MME. Drops were set up with 2µl DegP+DFP (12mg/ml), 0.3µl 100mM $MgCl_2$ and 1µl reservoir solution.

$DegP_{(+DFP)}$:

Crystals of DegP, incubated with the inhibitor DFP, but still active as a protease were grown using sitting drops equilibrated against a reservoir of 10% (v/v) isopropanol, 0.1M TRIS/HCl, pH 8.5, 10% (w/v) PEG 2000 MME. Drops were set up with 2µl $DegP_{(+DFP)}$ (10mg/ml), 0.3µl 100mM $MgCl_2$ and 1µl reservoir solution.

3.9.2 Preparation of heavy atom derivatives

The $DegP_{S210A}$ crystals were used for heavy atom derivatization. When the crystallization process had ceased, 20µl of the heavy atom solution were added in order to prepare potential heavy atom derivatives. For the purpose, the heavy atom compounds were dissolved in a solution identical to the respective reservoir solution. As the ideal condition for derivatization has to be found empirically, different heavy atom compounds with different concentrations were used for different periods. The following heavy atom compounds were used for soaking: Ta_6Br_{14}, K_2PtCl_4, $HgCl_2$, $AgNO_3$, AuCl, $ErCl_3$, Thiomersal, OsCl, IrCl, ReCl, OsO_4 in concentrations from 0.1 to 20 mM.

3.9.3 Improvement of diffraction properties

Crystals were handled the following way in order to improve their diffraction properties. After crystallization at room temperature, the crystals were shifted to 4°C. 24h later, 30µl of reservoir solution were added to the drop. The rest of the reservoir solution was discarded and replaced by a new solution containing 12% isopropanol, 50% (w/v) PEG 2000 MME, 0.1M Tris/HCl pH 8.5. An equilibrium between the drop and the new reservoir solution was reached approximately one week later. Crystals were then flash frozen by plunging into liquid nitrogen without using any additional cryoprotectant and maintained at ~100K in a nitrogen gas stream during data collection.

3.9.4 Data collection and data processing

Single crystals were mounted in Hampton Research Cryo loops of appropriate size, flash frozen to 100K in a cold nitrogen-gas stream and subjected to X-ray diffraction. Before the method described in section 3.9.3 was established, the crystals were transferred to cryobuffer for several seconds prior to flash-freezing in order to avoid ice rings and crystal damage due to cooling. Data were collected on beamline BW6 at the DESY (Hamburg, Germany) and on beamlines ID14-2, ID14-4 and ID29 at the ESRF (Grenoble, France).

For MAD experiments, prior to data collection an X-ray fluorescence spectrum of the respective crystal was recorded in order to determine the peak, inflection and remote wavelength corresponding to the anomalous scatterer present in the crystal. After recording and processing a test oscillation image, the angle range which is necessary to achieve an optimal completeness was determined by using the program STRATEGY (Leslie, 1992). The most precise anomalous differences are obtained by collecting the so-called Friedel mates at inverse beam geometry. For this purpose, for every wavelength a second data set was collected after rotating the crystal by 180° corresponding to the first starting value.

All diffraction data were processed and scaled using the programs DENZO and SCALEPACK (Otwinowski and Minor, 1997). The indexed intensities were converted and reduced to structure factors using the program TRUNCATE from the CCP4 program suite (Bailey, 1994).

cryobuffer: 15% (v/v) MPD
0.1M Tris/HCl, pH 8.5
15% (w/v) PEG 2000 MME

3.9.5 Structure solution

$DegP_{S210A}$

The heavy atom sites of the platinum derivative were identified with the automated program system SOLVE (Terwilliger and Berendzen, 1999). Afterwards the selenium sites could be localized by difference Fourier analyses, using programs from the CCP4 program suite (Bailey, 1994). Refinement of heavy atom parameters and phase calculation was done with SHARP (delaFortelle and Bricogne, 1997). The resulting electron density map was modified and improved by solvent flattening with SOLOMON (Abrahams and Leslie, 1996).

$DegP_{(+DFP)}$

As the crystals of $DegP_{(+DFP)}$ were nearly isomorphous to the $DegP_{S210A}$ crystals, the coordinates, constructing the asymmetric unit of the $DegP_{S210A}$ structure were used to calculate an initial rigid-body refinement, using data from 20 – 3.2Å. After this procedure, both R_{work} and R_{free} dropped well below 40%.

DegP+DFP

Strategy 1:

As the crystals of DegP+DFP were nearly isomorphous to the $DegP_{S210A}$ crystals, the coordinates, constructing the asymmetric unit of the $DegP_{S210A}$ structure were used to calculate an initial rigid-body refinement, using data from 20 – 3.2Å. After this procedure, both R_{work} and R_{free} dropped well below 40%.

Strategy 2:

The selenium sites from DegP+DFP(SeMet) crystals were identified with direct methods using Shake and Bake (Weeks and Miller, 1999). Refinement of heavy atom parameters, phase calculation and subsequent solvent flattening were done with CNS (Brunger *et al.*, 1998). Further phase improvement was achieved by phase combination using model phases

from the model obtained from strategy 1 (see above) and the experimental phases from strategy 2. All calculations were again performed with programs from CNS.

3.9.6 Model building and refinement

The program O (Jones *et al.*, 1991) was used for model building. For $DegP_{S210A}$ and $DegP_{(+DFP)}$ energy-restrained crystallographic refinement was carried out with maximum likelihood algorithms implemented in CNS (Brunger *et al.*, 1998), using the protein parameters of Engh and Huber (1991) (Engh and Huber, 1991). For DegP+DFP, refinement was performed with a maximum-likelihood target that incorporated experimental phase information *via* the inclusion of Hendrikson-Lattman coefficients from the MAD phasing, also as implemented in CNS (MLHL target). Bulk solvent, overall anisotropic B-factor corrections and non-crystallographic (NCS) restraints were introduced depending on the behaviour of the free R index. Refinement proceeded in several cycles, which were interrupted for manual rebuilding with the program O. The stereochemistry of the model was validated with PROCHECK (Laskowski *et al.*, 1993).

3.9.7 Graphical representations and sequence alignments

Graphical presentations were prepared with the programs MOLSCRIPT (Kraulis, 1991), RASTER3D (Merritt and Bacon, 1997), DINO (Philippsen, 2002), SETOR (Evans, 1993) and PYMOL (DeLano, 2004). Surface and electrostatics calculations were made with MSMS (Sanner *et al.*, 1996) and GRASP (Nicholls *et al.*, 1993). Initial sequence alignments were made with ClustalW (http://www.ch.embnet.org/software/ClustalW.html) and presentations of sequence alignments were made with ALSCRIPT (Barton, 1993).

Chapter 4

Results

4.1 Purification of $DegP_{S210A}$, wild-type DegP (DegP/$DegP_{(+DFP)}$) and DegP in complex with DFP

4.1.1 Purification of $DegP_{S210A}$

$DegP_{S210A}$ was purified by using a novel three-step purification procedure. After cell lysis by sonication, the clarified supernatant was applied to a Ni-NTA column. The proteins were eluted using a step-gradient of imidazole. After the unbound protein passed through the column, the unspecific bound proteins were eluted by using buffer B (40mM imidazole), followed by elution of $DegP_{S210A}$ with buffer C (150mM imidazole) (Fig. 4.1 & 4.2).

The protein was dialysed against buffer D and then $DegP_{S210A}$ was applied to a hydroxyapatite-column. The bound protein was subsequently eluted with a linear potassium phosphate-gradient from 0 to 500mM phosphate (Fig. 4.3). On the chromatogram, two peaks are distinguishable. To decide which one of them contains $DegP_{S210A}$, all fractions of the two peaks were analyzed by SDS-PAGE (Fig. 4.4). As seen from the gel, all fractions contain a band corresponding to $DegP_{S210A}$. The appearance of a band called 'crude' on the gel was used as a marker for the separation into two hydroxyapatite fractions: $DegP_{S210A}$ and $DegP_{S210A,12}$. After concentrating the two fractions by ultrafiltration with an AMICON-cell and centrifugal concentrators, the fractions were loaded on a Superdex-200 prep grade gel filtration column. The two gel filtration runs showed basically the same peak pattern but with different amplitudes.(Fig. 4.5 & 4.6). Peaks were separated and subsequently combined according to their elution volume. Figure 4.6 illustrates how the separation was made.

SDS-PAGE analysis of $DegP_{S210A}$ indicated that the protein preparation was virtually pure except for some minor contaminations (Fig. 4.7). SDS-PAGE analysis of $DegP_{S210A,12\text{-}A}$ and $DegP_{S210A,12\text{-}B}$ indicated that beside the band corresponding to $DegP_{S210A}$, the band previously termed "crude" is still present.

After gel filtration, the three fractions were further concentrated, the buffer exchanged by using a NAP-10 (Pharmacia) desalting column and finally the samples were stored at -70°C until further analysis. It should be noted that $DegP_{S210A}$ makes up the majority of purified protein (>80%), whereas the two other fractions namely $DegP_{S210A,12\text{-}A}$ and $DegP_{S210A,12\text{-}B}$ were only present in minor amounts. In total, a 12l expression yielded approximately 20mg of pure $DegP_{S210A}$.

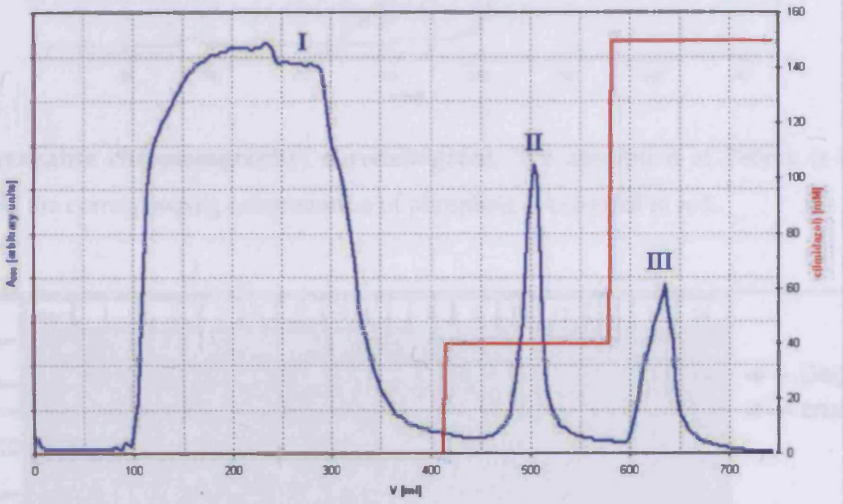

Fig. 4.1. Ni-NTA chromatography: chromatogram. The absorption at 280nm is coloured in blue (arbitrary units) and the corresponding concentration of imidazole is coloured in red. Peaks are labelled with roman numbers.

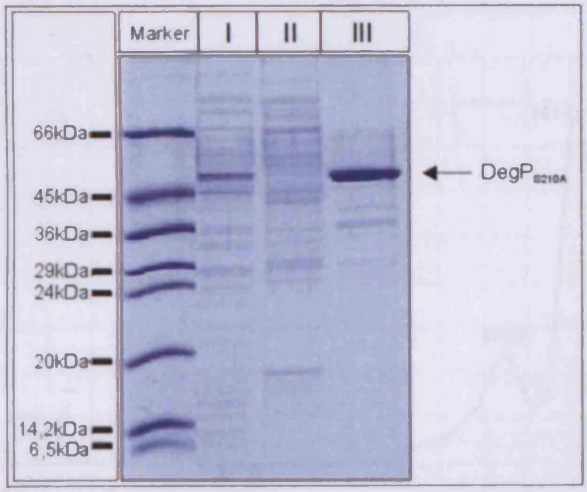

Fig. 4.2. Ni-NTA chromatography: 12.5% SDS-PAGE. Lane "Marker": molecular weight marker; Lane I: Peak I (see Fig. 4.1), unbound proteins; Lane II: Peak II, unspecific bound proteins; Lane III: Peak III, eluted protein.

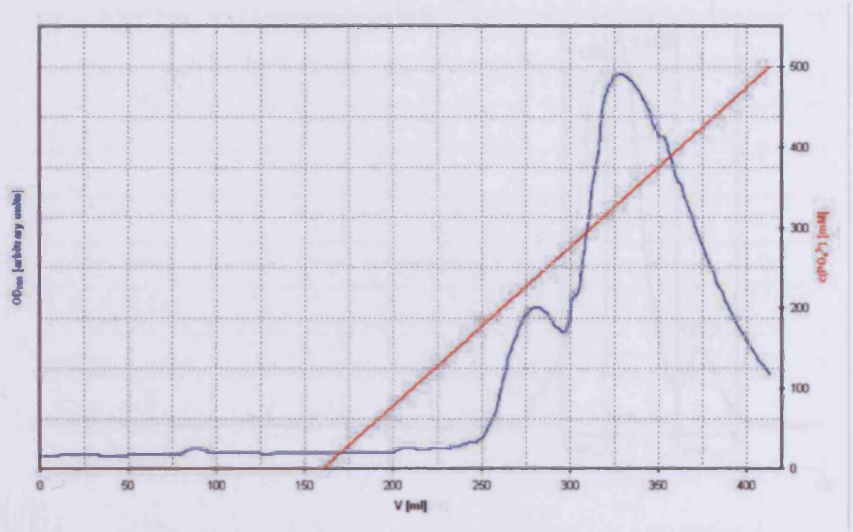

Fig. 4.3. Hydroxyapatite chromatography: chromatogram. The absorption at 280nm is coloured in blue (arbitrary units) and the corresponding concentration of phosphate is coloured in red.

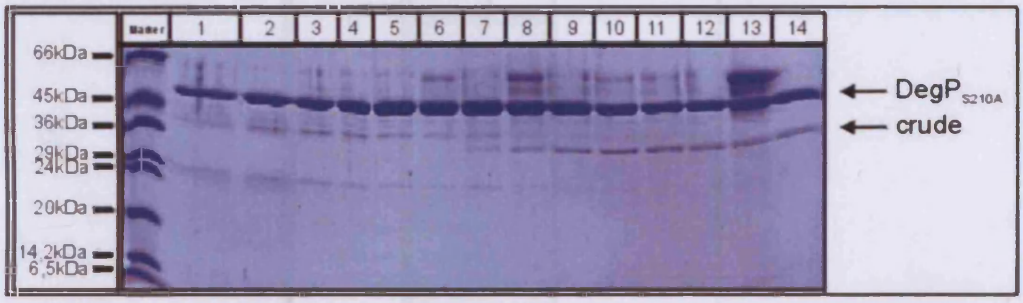

Fig. 4.4. Hydroxyapatite chromatography: 12.5% SDS-PAGE. Lane "Marker": molecular weight marker; Lane 1 – 14: consecutive peak fractions.

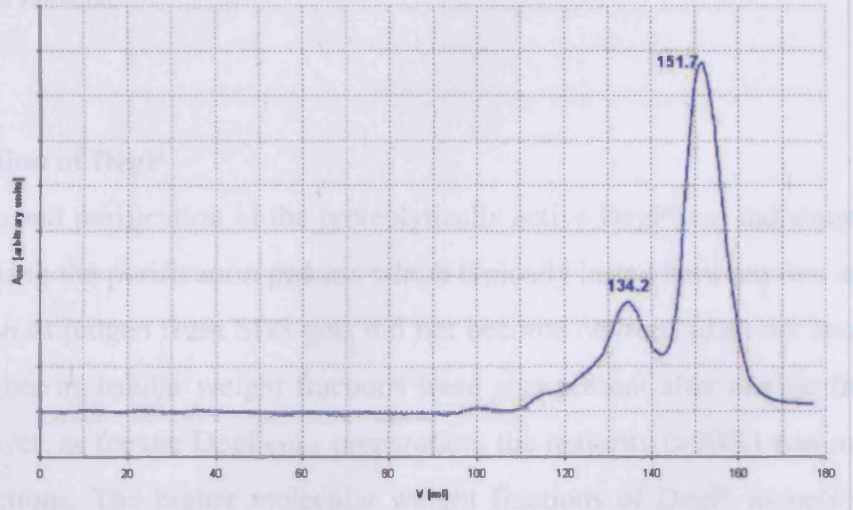

Fig. 4.5. Gel filtration (Superdex-200): $DegP_{S210A}$.

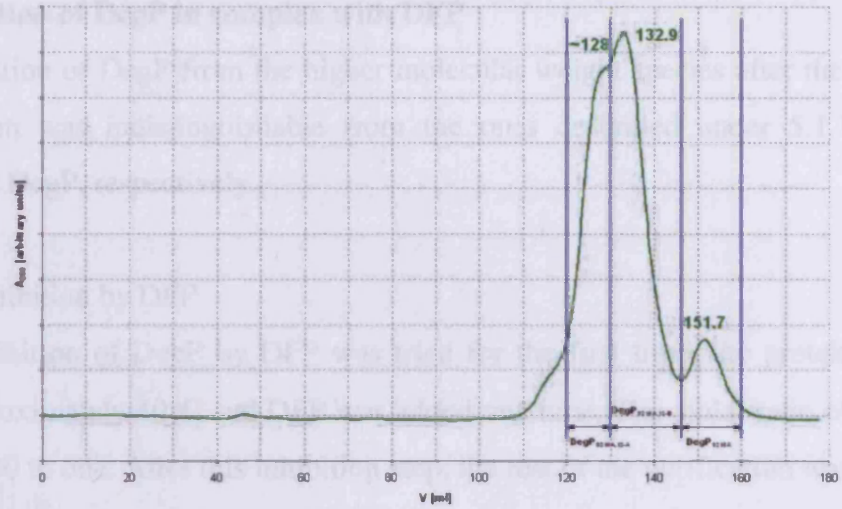

Fig. 4.6. Gel filtration (Superdex-200): $DegP_{S210A,12}$. Fraction were separated as indicated by the blue lines.

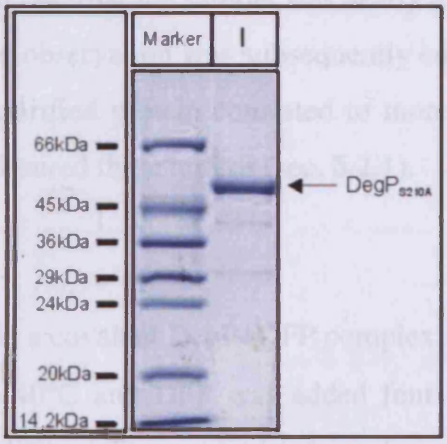

Fig. 4.7. 12.5% SDS-PAGE of purified $DegP_{S210A}$. Lane "Marker": molecular weight marker; Lane I: purified $DegP_{S210A}$ after gel filtration.

4.1.2 Purification of DegP

The expression and purification of the proteolytically active DegP was indistinguishable from $DegP_{S210A}$. During the purification process which typically lasted between two and three days, autodegradation as judged from SDS-gels did not become obvious (data not shown). Like for $DegP_{S210A}$, higher molecular weight fractions were also present after elution from the HAP-column. However, as for the $DegP_{S210A}$ preparation, the majority (>90%) was made up by the hexameric fractions. The higher molecular weight fractions of DegP, namely $DegP_{12\text{-}A}$ and $DegP_{12\text{-}B}$ were not considered for further purification.

4.1.3 Purification of DegP in complex with DFP

Till the separation of DegP from the higher molecular weight species after the HAP-column, the purification was indistinguishable from the ones described under 5.1.1 or 5.1.2 for $DegP_{S210A}$ and DegP, respectively.

Incomplete inhibition by DFP

When the inhibition of DegP by DFP was tried for the first time, the protein solution was heated to approximately 40°C and DFP was added one time. The molar ratio of DFP to DegP was at least 500 to one. After this inhibition step, the rest of the purification was carried out as described previously and the protein didn't show any kind of different behaviour. However, when the remaining protease activity of the protein was determined using resorufin-labelled casein as a substrate, it turned out, that the sample was nearly as active as an untreated control sample (data not shown). The observation was subsequently confirmed by mass spectroscopy analyses indicating that the purified protein consisted to more than 95% of wild-type DegP and only a minor fraction harboured the inhibitor (see. 5.2.1).

Complete inhibition by DFP

As the first attempt to produce a covalent DegP+DFP complex was unsuccessful, this time the sample was again heated to 40°C and DFP was added four times with a delay of 45min. between subsequent DFP additions. The approximate molar ratio of DFP to DegP for every addition was 500 to 1. The rest of the purification was carried out as previously described without any differences. In the end, mass spectrometric analyses confirmed, that the purified protein consisted to nearly 100% of DegP+DFP, although even this batch exhibited some residual protease activity (data not shown).

4.1.4 Purification of selenomethionine substituted DegP proteins

The proteins $DegP_{S210A}$(SeMet) and DegP(SeMet) were expressed in methionine auxotroph CS83 cells. Though expression in NM-medium and CS83 cells was somewhat slower and the apparent yield of overexpressed protein was lower than for the expression of the native proteins, neither $DegP_{S210A}$(SeMet) nor DegP+DFP displayed any apparent differences during the purification as well as the inhibition procedure, respectively. Higher molecular weight fractions of DegP, previously named as $DegP_{S210A,12\text{-}A,B}$ were discarded after they had been isolated from the HAP-column.

4.2 Identification of the purified proteins

4.2.1 Mass spectrometry

Purified samples of DegP, DegP+DFP, DegP+DFP(SeMet), $DegP_{S210A}$ and $DegP_{S210A}$(SeMet) were analyzed by electrospray ionization mass spectrometry (ESI-MS). The analysis was carried out at the Max-Planck Institute for Biochemistry in the group of Dr. Siedler. The results are summarized in the following table:

	expected M_W	*measured* M_W
DegP	47895Da	47896Da
crystals of $DegP_{(+DFP)}$	?	47897Da
DegP+DFP	48059Da	48059Da
DegP+DFP(SeMet)	48714Da	48711Da
$DegP_{S210A}$	47879Da	47882Da
$DegP_{S210A}$(SeMet)	48534Da	48537Da

Table 4.1. Mass spectrometry results.

The expected molecular weight for all DegP variants used throughout this study is 1066Da higher, than calculated from the amino acid composition of the native protein. This stems from the fact, that at its C-terminus, the recombinant protein contains a small linker consisting of a serine and an arginine residue followed by the polyhistidine tag which contributes six additional histidine residues to a DegP monomer (Spiess *et al.*, 1999).

The measured molecular weight of DegP and $DegP_{S210A}$ only shows minor differences to the expected molecular weight as calculated from the proposed primary structure. These differences are well within the error limits of the instrument. Although the covalent inhibitor DFP was used in the purification of the crystals of $DegP_{(+DFP)}$, the mass spectrum clearly demonstrated, that DFP was obviously not present in the crystals.

DFP has a molecular weight of 174Da. After reaction with the hydroxyl group of the active site serine a HF molecule is separated and the diisopropyl phosphate moiety is covalently attached to the serine, thus the molecular weight of the complex should increase by 164Da. When DFP was added four times during the purification of DegP, one could observe a

difference of 164Da to the molecular weight of uninhibited DegP, confirming the presence of DFP in the active site of the protein.

The difference between the measured molecular weight of $DegP_{S210A}$ and $DegP_{S210A}$(SeMet) is 655Da. If we know that the difference of the molecular weight between selenium and sulphur is about 46Da and there are 14 methionine residues per DegP monomer, then this difference appears to be the result of the correct incorporation of the selenomethionine into $DegP_{S210A}$. For the DegP+DFP(SeMet) protein, the mass spectrometry analyses, confirmed both, namely the quantitative substitution of DegP with DFP moieties and the complete incorporation of selenomethionine residues.

4.2.2 N-terminal sequencing

After SDS-PAGE, DegP was blotted to a PVDF membrane and subsequently analyzed by N-terminal sequencing via Edman degradation. The analysis has been carried out at the Max-Planck Institute for Biochemistry by Dr. K.-H. Mann in the Department Proteinanalytik. The sequence of the first N-terminal residues is A-E-T-S-S-A-T-T, corresponding to the N-terminal residues of the mature DegP. Therefore one can assume that the signal-peptide has been cleaved correctly and that the protein was correctly translocated into the periplasmic compartment.

Although the $DegP_{S210A,12\text{-}B}$ protein eluted as a single peak during size-exclusion chromatography (Fig. 4.6) and furthermore displayed a monomodal particle distribution in dynamic light scattering experiments (Tab 4.2), SDS-PAGE analysis of the purified sample clearly revealed two distinct bands with an estimated molecular weight of 35kDa and 48kDa, respectively (Fig. 4.8). The upper band clearly corresponds to the DegP protein whereas N-terminal sequencing of the lower band resulted in the amino acid sequence A-E-V-Y-N-K-D-G-N-K. This sequence corresponds to the *E.coli* porin OmpC, which has a molecular mass of 38307.5kDa. This finding is in good agreement with a recent publication from CastilloKeller and Misra who showed that the expression of an unfolded OmpC mutant protein led to a sequestering of that protein by DegP (CastilloKeller and Misra, 2003).

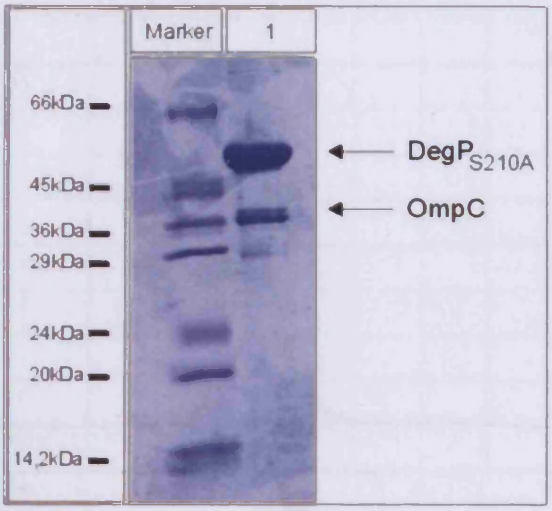

Fig. 4.8. 12.5% SDS-PAGE of $DegP_{S210A,12\text{-}B}$. Lane "Marker": molecular weight marker; Lane I: purified $DegP_{S210A,12\text{-}B}$ after gel filtration. The upper band has a molecular weight corresponding to $DegP_{S210A}$, whereas N-terminal sequencing of the lower band revealed that this band corresponds to the *E.coli* porin OmpC.

4.3 Oligomeric state of $DegP_{S210A}$ in solution

4.3.1 Size-exclusion chromatography (SEC)

The native molecular weight of both $DegP_{S210A}$ and $DegP_{S210A,12\text{-}B}$ have been analyzed on a Superose-6 column. The resulting chromatograms are shown in Fig. 4.9 and in Fig. 4.10 respectively.

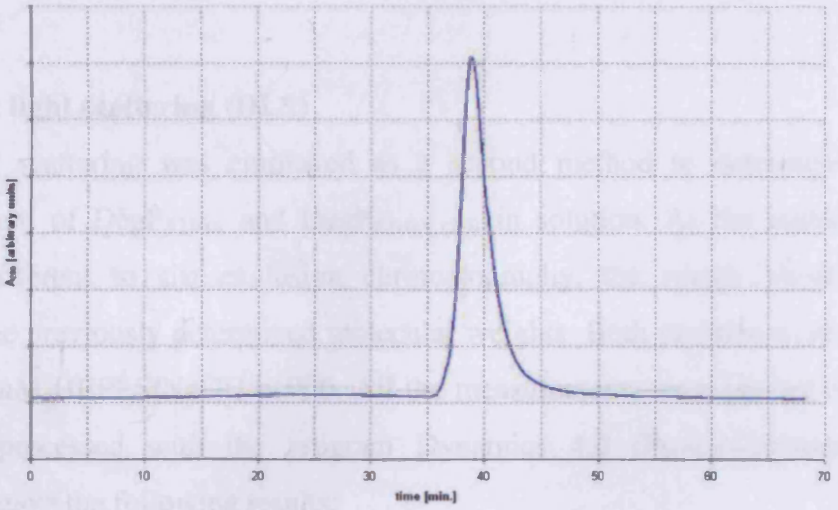

Fig. 4.9. Analytical gel filtration (Superose-6): $DegP_{S210A}$.

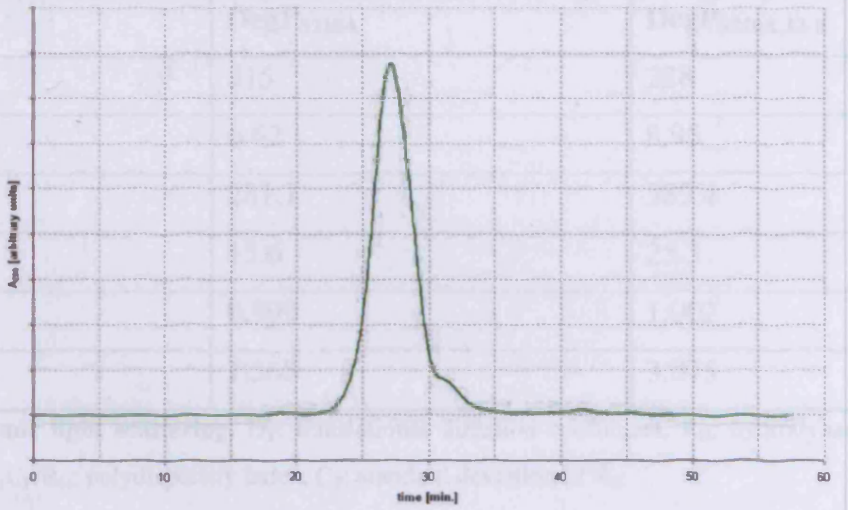

Fig. 4.10 Analytical gel filtration (Superose-6): $DegP_{S210A,12\text{-}B}$.

Both chromatograms show one clear peak. According to the marker proteins, the molecular weight of $DegP_{S210A}$ and $DegP_{S210A,12\text{-}B}$ were calculated as 310kDa and 480kDa respectively. Therefore $DegP_{S210A}$ seems to be a hexamer in solution whereas $DegP_{S210A,12\text{-}B}$ seems to be a nonamer. It should be noted, that analyses of $DegP_{S210A,12\text{-}B}$ and $DegP_{S210A,12\text{-}A}$ on a Superose-12 column resulted in an apparent molecular weight of 500kDa and 950kDa respectively (data not shown). It should furthermore be noted, that the calculated molecular weight is critically dependent on the measured retention times or elution volumes of the marker proteins. As there is a logarithmic correlation between the K_{av} value and the molecular weight, even small variations in the measured values may have dramatic effects on the apparent molecular weight.

4.3.2 Dynamic light scattering (DLS)

Dynamic light scattering was employed as a second method to determine the apparent molecular weight of $DegP_{S210A}$ and $DegP_{S210A,12\text{-}B}$ in solution. As the underlying physical principle is different to size-exclusion chromatography, the results should support the reliability of the previously determined molecular weights. Both proteins were measured in a solution of 10mM HEPES/NaOH pH8.0. All the measurements were carried out at 20°C and were further processed with the program Dynamics 4.0 (Protein-Solutions Inc.). The measurements gave the following results:

Parameter	$DegP_{S210A}$	$DegP_{S210A,12-B}$
D_T ($10^{-9}cm/s^2$)	315	228
R_H (nm)	6.62	8.95
M_W (kDa)	281.1	589.4
C_P/R_H (%)	15.6	25.3
Baseline	0.999	1.002
SOS	1.368	3.075

Table 4.2. Dynamic light scattering. D_T: translational diffusion coefficient, R_H: hydrodynamic radius, M_W: molecular weight, C_P/R_H: polydispersity index, C_P: standard deviation of R_H.

The interpretation and the use of the statistical parameters as calculated by Dynamics 4.0 are given below. The values have been adapted from the DynaPro-801 Operator Manual (Protein Solutions Inc.):

Parameter	Interpretation
Baseline	
0.977-1.002	Monomodal distribution
1.003-1.005	Bimodal distribution
>1.005	Bimodel/multimodal distribution
Sum of Squared (SOS)	
1.000-5.000	Low noise, negligible error
5.000-20.000	Background error owing to noise, low protein concentration or a small amount of polydispersity
>20.000	High noise/error owing to high polydispersity in size distribution (aggregation), irregular solvent
Polydispersity	Note: this parameter should be used for monomodal distribution only
$C_P/R_H<15\%$	Monodisperse solution, very likely to crystallize
$C_P/R_H<30\%$	A moderate amount of polydispersity, more likely to crystallize
$C_P/R_H>30\%$	A significant amount of polydispersity, less likely to crystallize

Table 4.3. Interpretation and use of the statistical parameters as calculated by Dynamics 4.0 (Protein Solutions Inc.)

4.3.3 Summary

The results of size exclusion chromatography and dynamic light scattering are summarized in the following table:

Method	$DegP_{S210A}$	$DegP_{S210A,12\text{-}B}$	$DegP_{S210A,12\text{-}A}$
SEC	310kDa	480kDa	950kDa
DLS	280kDa	590kDa	-

Table 4.4. Molecular weight of DegP in solution: summary

Three different oligomers of $DegP_{S210A}$ are existent side by side in solution, i.e. three different, stable oligomeric forms could be isolated. They are not interconvertible and are obviously present as a hexamer, a dodecamer and an octadecamer. However, one has to keep in mind that the measurements are not error-free and that the results differ significantly between SEC and DLS. Moreover, N-terminal sequencing revealed that the porin OmpC is bound to $DegP_{S210A,12\text{-}A}$ and $DegP_{S210A,12\text{-}B}$ but not to $DegP_{S210A}$. Though the stoichiometry of these DegP – OmpC complexes is not known, it is arguable whether speaking about a dodecamer or an octadecamer is correct.

4.4 Inhibitors and activators of DegP

4.4.1 Protease assay: synthetic versus protein substrates

Although very high DegP (360µM) concentrations were used for the degradation assay with the synthetic protease substrate N-(Methoxysuccinyl)-Ala-Ala-Pro-Val 4-nitroanilide, no increase in the absorbance at 405nm could be observed. Even incubation for several hours at various temperatures had no effect. Thus, as a colorimetric activity assay only the degradation of resorufin-labelled casein remained, with the disadvantage, that this substrate is not properly suited for the determination of kinetic constants, because the casein is not evenly substituted with resorufin moieties.

4.4.2 Sensitivity of DegP to different inhibitors

Up to date only a very limited number of serine protease inhibitors have been analyzed, concerning their ability to avoid the protease activity of DegP. Goldberg and co-workers (Swamy *et al.*, 1983) showed that the proteolytic activity of DegP is completely prevented by the serine protease inhibitor DFP. PMSF another typical serine protease inhibitor inhibited the protein only to a minor extent and it was not at all inhibited by the classical chloromethylketone inhibitors (Swamy *et al.*, 1983).

In this study, the inhibitory effect of DFP, DIC, BisANS, MG-115, MG-132, PSI and Ms-AAPV-CMK was elucidated. The DegP concentration used in the study was 1μM. The corresponding inhibitor concentration and the results of this investigation are given in table 4.5:

Inhibitor	Concentration	inhibition of DegP [%]
DFP	5mM	91
DIC	5mM	77
BisANS	2mM	97
MG-115	5mM	< 5
MG-132	5mM	< 5
PSI	5mM	< 5
Ms-AAPV-CMK	2mM	< 5

Table 4.5. Sensitivity of DegP to different inhibitors. DFP: diisopropyl fluorophosphate, **DIC:** 3,4-dichloroisocoumarin, **BisANS:** 4,4'-dianilino-1,1'-binaphtyl-5,5'-disulfonic acid, **MG-115:** carbobenzoxy-L-leucyl-L-leucyl-L-norvalinal, **MG-132:** carbobenzoxy-L-leucyl-L-leucyl-leucinal, **PSI:** carbobenzoxy-L-isoleucyl-γ-t-butyl-L-glutamyl-L-alanyl-L-leucinal, **Ms-AAPV-CMK**: N-(Methoxysuccinyl)-Ala-Ala-Pro-Val-chloromethyl ketone.

Beside DFP, DIC proved to be a potent protease inhibitor for DegP. However, when it was tried to use the inhibitor for protein purification, it turned out that its addition led to the precipitation of DegP, a feature of the inhibitor that was not recognized, during these measurements on an analytical scale. Furthermore it was quite remarkable that the tripeptidic inhibitors, which all possess a small hydrophobic amino acid residue in their P_1 position, did not have any apparent inhibitory effect. Typically, the relative error within sample triplets was in the range of ±8%. Thus it seems to be obvious that the observed small inhibitory effect of

the aldehyde peptides MG-115, MG-132 and PSI and the chloromethylketone inhibitor Ms-AAPV-CMK lies within the experimental error. Therefore, the small residual inhibitory effect may be considered to be insignificant. Even a systematic variation of the reaction conditions, namely temperature and pH did not influence the results (data not shown). Very interesting, especially in light of the crystal structure is the ability of BisANS to prevent the protease activity of DegP. BisANS (Fig. 4.11) is a hydrophobic probe that has been extensively used to analyse the hydrophobic sites on the surface of proteins (Lee *et al.*, 1997). A possible explanation for the inhibitory effect could be binding of BisANS to the internal phenylalanine cluster of the DegP active site cage (see Fig. 4.32 and Fig. 4.33), thereby clogging the internal cavity and preventing substrate access towards the active sites.

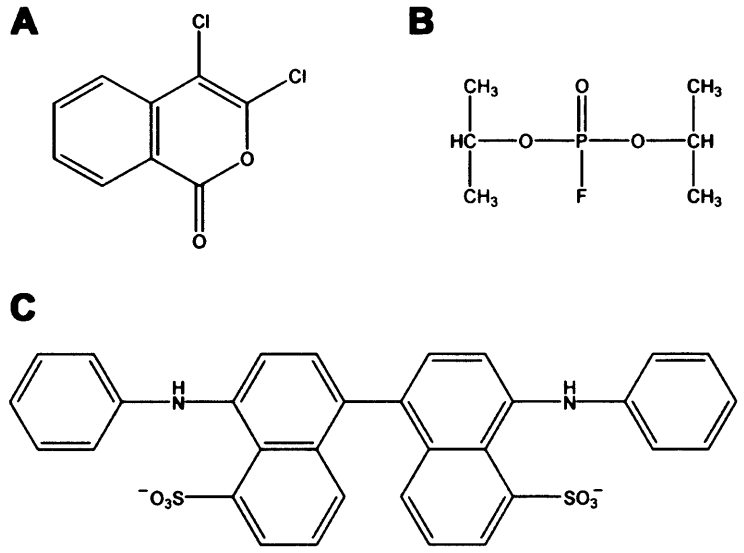

Fig. 4.11. Structure of inhibitors of DegP. A: 3,4-dichloroisocoumarin (DIC), **B:** diisopropyl fluorophosphate (DFP), **C:** 4,4'-dianilino-1,1'-binaphtyl-5,5'-disulphonic acid (BisANS).

4.4.3 Effect of different peptides on the protease activity of DegP

Recently it was demonstrated for some DegP homologues that peptides can stimulate their peptidase activity by binding to the respective PDZ domain (Martins *et al.*, 2003; Walsh *et al.*, 2003). A similar behaviour was shown for DegP, although in that case the authors could only speculate about the binding site (Jones *et al.*, 2002). Thus, C-terminal peptides derived from OmpC, which could turn on the protease activity of DegS against its natural substrate RseA, and the peptide SPCJ-1 were used in order to address this question. The results of the survey are given in table 4.6. The activity stated is referred to DegP without peptides (corresponds to 100%).

peptide	sequence	concentration	activity
SPCJ-1	DNRDGNVYQF	170µM	150%
OmpC[1]	NTDNIVALGLVYQF	100µM	100%
OmpC[2]	KSMCMKLSFS	25µM	96%

Tab. 4.6. Effect of different peptides on the protease activity of DegP.

Both peptides derived from the C-terminus of OmpC and described by Sauer and coworkers in a study about DegS (Walsh *et al.*, 2003) did not show any detectable effect on the protease activity of DegP. SPCJ-1 a peptide derived from the carboxyl-terminal pilin subunit peptide from *E.coli* was able to increase the protease activity of DegP, like it has been reported in a previous study (Jones *et al.*, 2002).

4.5 Temperature-dependent reaction of DegP with DFP

DegP was incubated with the covalent inhibitor DFP at temperatures ranging from 4°C to 48°C. After 3h, free, non-reacted DFP was removed by acetone precipitation, the protein pellet was dissolved in incubation buffer and its remaining proteolytic activity was determined. The results are illustrated in figure 4.12. While the temperature increases during the incubation of DegP with DFP, the inhibition of the protease is improved.

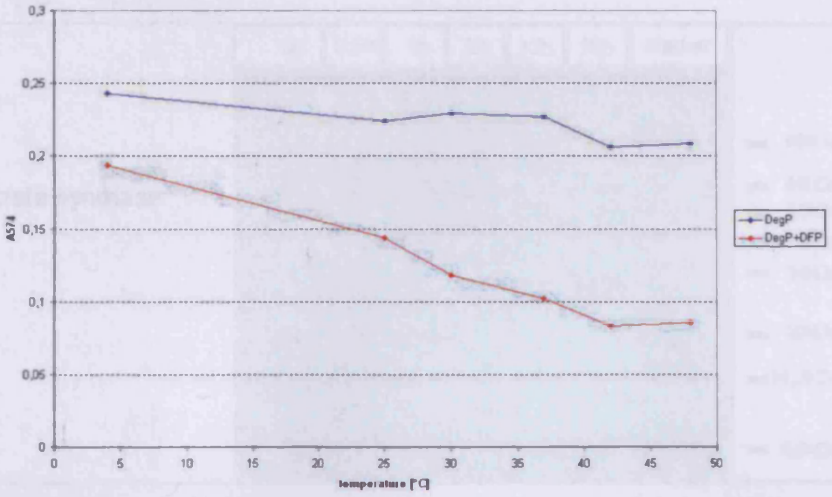

Fig. 4.12. Loss of protease activity of DegP after treatment with DFP at different temperatures. The proteolytic activity for the non-inhibited control experiment is illustrated in blue, the proteolytic activity of inhibited DegP it is illustrated in red.

4.6 DegP degrades its substrates in a processive fashion

The time-dependent degradation of resorufin-labelled casein and citrate synthase by DegP was analyzed using reversed-phase HPLC and SDS-PAGE. The corresponding chromatograms as well as the SDS-PAGE gels are shown in figures 4.13 and 4.14. Note that the figures 4.13 and 4.14 stem from different experiments.

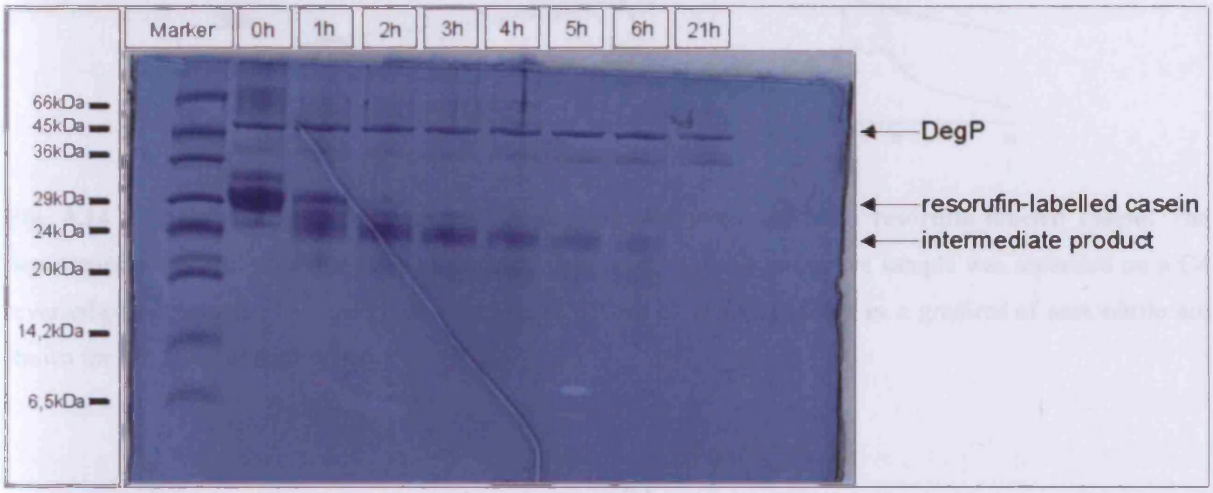

Fig. 4.13 A. SDS-PAGE analysis of resorufin-labelled casein degraded by DegP. Incubation times were 0, 1, 2, 3, 4, 5, 6 and 21.

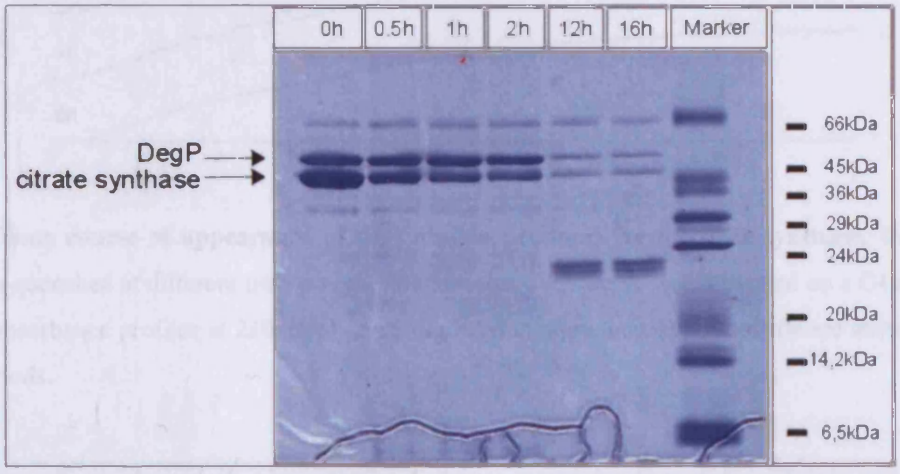

Fig. 4.13 B. SDS-PAGE analysis of citrate synthase degraded by DegP. Incubation times were 0, 0.5, 1, 2, 12 and 16 hours, respectively.

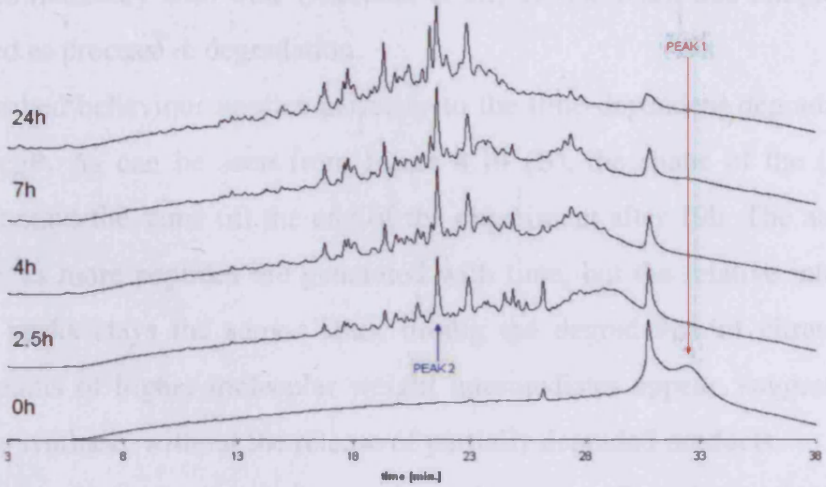

Fig. 4.14 A. Time course of appearance of degradation products from resorufin-labeled casein. The degradation assay has been quenched at different time-points and the respective sample was separated on a C4 reversed-phase column. The absorbance profiles at 210nm of peptides eluted in a gradient of acetonitrile are shown for different incubation periods.

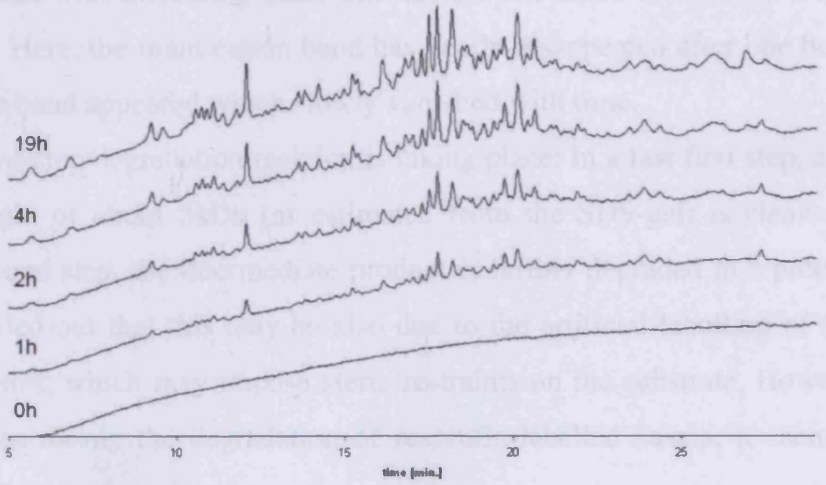

Fig. 4.14 B. Time course of appearance of degradation products from citrate synthase. The degradation assay has been quenched at different time-points and the respective sample was separated on a C4 reversed-phase column. The absorbance profiles at 210nm of peptides eluted in a gradient of acetonitrile are shown for different incubation periods.

When the time-dependent degradation of substrate proteins of cage-forming proteases like the proteasome of ClpAP was analysed by HPLC, it was found out, that the cleavage pattern remained time-invariant (Thompson *et al.*, 1994). This observation is in sharp contrast to the behaviour of monomeric endopeptidases like chymotrypsin, where the pattern of peptides

generated varies markedly with time (Akopian *et al.*, 1997). Thus, this unique property was therefore termed as processive degradation.

The afore described behaviour applies perfectly to the time-dependent degradation of citrate synthase by DegP. As can be seen from figure 4.14 (B), the shape of the initial cleavage pattern after 1h stays the same till the end of the experiment after 19h. The amplitude of the peaks increases as more peptides are generated with time, but the relative intensity between the individual peaks stays the same. Thus, during the degradation of citrate synthase, no significant amounts of higher molecular weight intermediates appear, suggesting that DegP degrades citrate synthase, without the release of partially degraded products

.The results for the casein digest remain somehow obscure. At first glance, the situation seems to be the same. Most of the peaks maintain their relative intensity during the time course, but peak 2 is already present after 2.5 hours and does not change its intensity in the course of the experiment. At the beginning of the experiment, the substrate is present as two peaks, however, peak 1 is completely vanished after 2.5 hours, whereas the other substrate peak steadily decreases with increasing time. The same observation is made on a SDS-PAGE gel (Fig. 4.13(A)). Here, the main casein band has nearly disappeared after one hour, but instead an intermediate band appeared which slowly vanished with time.

Obviously a two-step degradation reaction is taking place: In a fast first step, a peptide with a molecular weight of about 5kDa (as estimated from the SDS-gel) is cleaved off and in a subsequent second step, the intermediate product is further degraded in a processive fashion. It cannot be ruled out that this may be also due to the artificial labelling of the casein with resorufin moieties, which may impose steric restraints on the substrate. However, despite of the irregularities during the degradation of resorufin-labelled casein, it seems permitted to designate DegP as a processive protease.

4.7 The length distribution of DegP's degradation products

Complete substrate digests of chemically denatured MalS, PhoA and TreA and of heat denatured citrate synthase were analysed by mass spectrometry in order to reveal the identity of the product peptides. The published primary sequences of the substrates were used to assign the molecular weights after mass spectrometry to the corresponding peptides. With all the four proteins a large number of different peptides were generated. However, it is obvious, that not all the peptides could have been detected by this 'brute force' method, where it was

attempted to figure out the composition of a complex mixture in a single step. Apparently, some peptides whose existence would be caused by a present neighbouring fragment were not detected. This is illustrated by the representative cleavage pattern of the citrate synthase (Fig. 4.15). It should further be mentioned, that not all peptides were detected with the same reliability (data not shown). Especially larger peptides, containing more than 30 amino acid residues were of lower significance. However the vast majority was clearly detected (personal communication F. Siedler), thus statistical parameters about the reliability of individual peptides were omitted in the analyses. This was also done with respect to the fact, that there is no clear cut-off value, which will allow a differentiation between a 'good' and a 'bad' peptide.

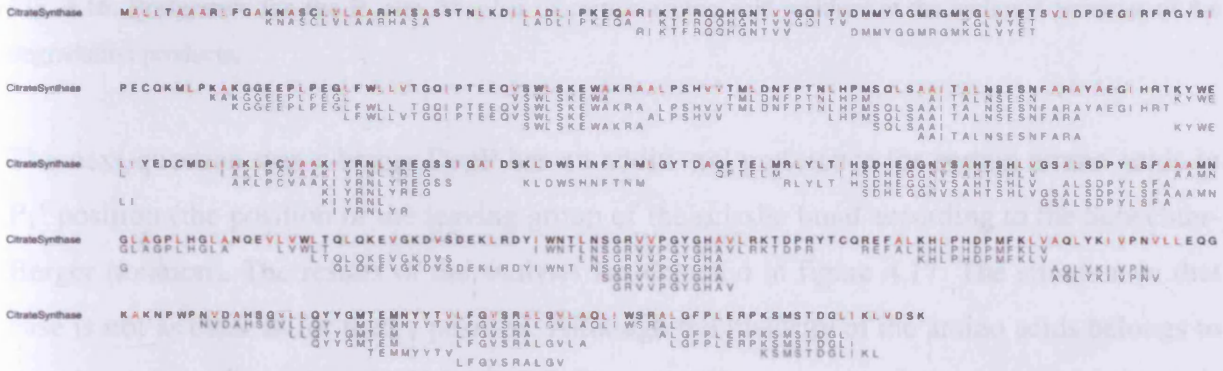

Fig. 4.15. Degradation of citrate synthase by DegP. The sequence is represented in single letter code and all alanine (A), valine (V), isoleucine (I) and leucine (L) residues are coloured in red. The identified peptides are coloured in light grey and are aligned to their respective position in the citrate synthase sequence.

Altogether, there were 177 peptides identified, 37 for MalS, 49 for TreA, 31 for PhoA and 60 for CS. The vast majority of the residues at the carboxy-terminus of the peptide fragments (P_1 position) is constituted by small hydrophobic amino acids like alanine, isoleucine, leucine or valine (Fig. 4.16). This is in good agreement with previous studies (Kolmar *et al.*, 1996; Jones *et al.*, 2002). Additionally a number of threonine and serine residues were identified, but other amino acids were found only incidental, most often belonging to peptides with little reliability. No further preferences for the S2 to S10 subsites could be identified (data not shown).

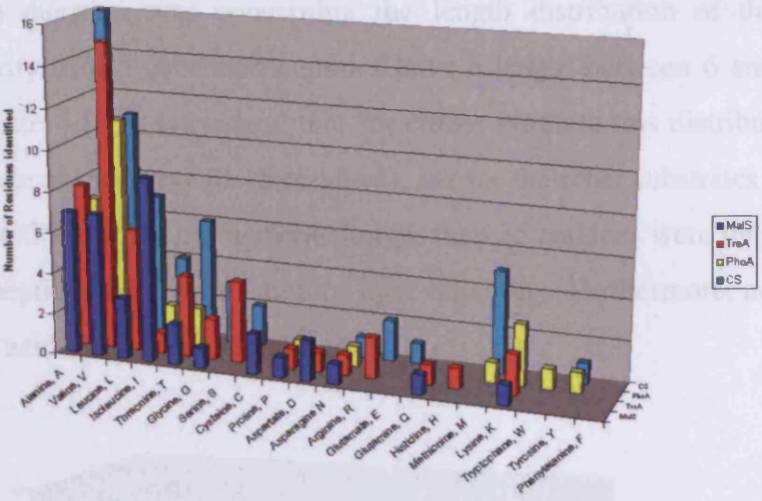

Fig. 4.16. Preference for the P_1 site. Number of certain amino acid residues at the carboxyl terminus of the degradation products.

The next question was whether DegP has an additional preference for certain amino acids in P_1' position (the position of the leaving group of the scissile bond according to the Schechter-Berger notation). The results of the analysis are depicted in figure 4.17. The situation in that case is not as clear as for the P_1 position. Although the majority of the amino acids belongs to the class of small residues (A,V,L,I,T,G,S,C), a significant number belongs to the other half comprising bigger residues The charge of the residues is irrelevant.

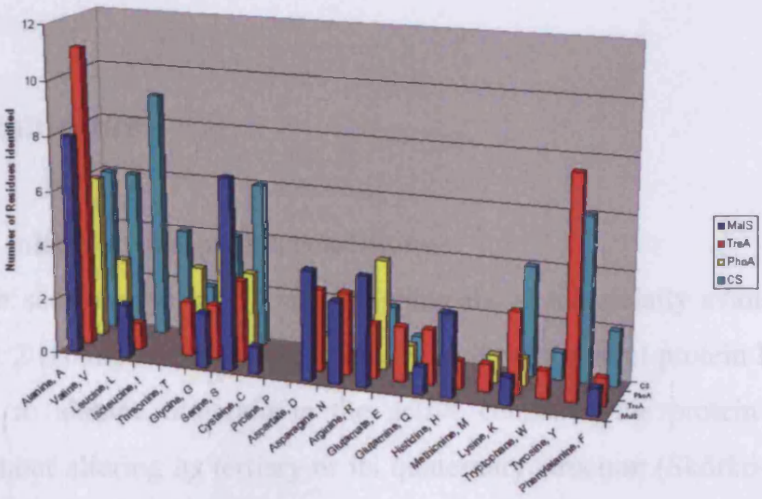

Fig. 4.17. Preference for the P1' site. Number of certain amino acid residues adjacent to the carboxyl-terminus of the degradation products.

However, the main question was concerning the length distribution of the degradation products. The majority of the generated peptides have a length between 6 and 25 residues. When looking at figure 4.18, it is evident, that for citrate synthase this distribution is shifted towards the lower side of this array (6-15 residues), but for the other substrates it is shifted to the high-end side (16-25 residues). Fragments longer than 25 residues were only occasionally detected and these peptides were most often of little reliability. Furthermore, no peptide with less than five amino acid residues was found.

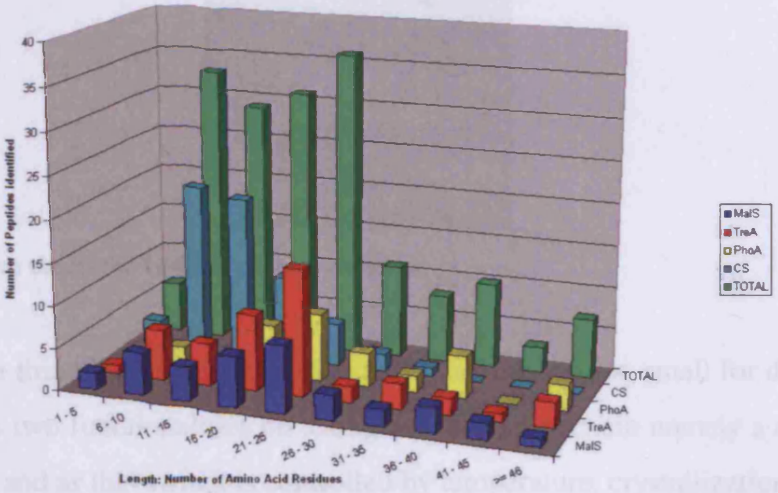

Fig. 4.18. Histogram of the length distribution of all identified peptides. The graphs do not take into account the differences in the relative abundance of a given peptide.

4.8 Protein crystallization

4.8.1 Screening for initial crystallization conditions

Initial crystallization setups have been made by using the commercially available Hampton Crystal Screens 1 & 2 (Hampton Research, USA). Initially the mutant protein $DegP_{S210A}$ was used, as the serine to alanine mutation in the active centre of the protein prevented its autodegradation without altering its tertiary or its quaternary structure (Skórko-Glonek *et al.*, 1995). Autodegradation leads to a heterogeneous protein composition, usually makings protein crystallization impossible (Ferré-d'Amaré and Burley, 1994).

The setups were made using the sitting-drop method at 4°C and at 18°C and the drop ratio has been 2:1, i.e. 2µl of protein and 1µl of reservoir solution. The concentration of the protein

solution was approximately 12mg/ml throughout all the different crystallization trials. After approximately two weeks thin, hexagonal crystal plates (Fig. 4.19) appeared at 18°C in the following condition:

10% (v/v) isopropanol

0.1M Hepes, pH 7.5

20% (w/v) PEG 4000

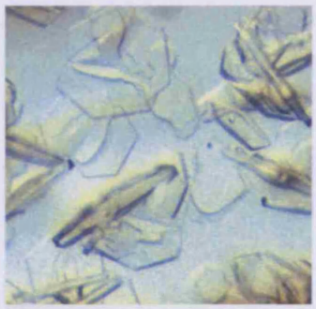

Fig. 4.19. Crystals from the initial crystallization condition.

Unfortunately these thin plates were quite difficult to handle and too small for data collection. As DegP combines two functionalities on a single polypeptide chain namely a chaperone and a protease activity, and as this switch is controlled by temperature, crystallization of $DegP_{S210A}$ was also tried at 36°C, a temperature, where the protein should exist exclusively in the protease conformation. However, initial trials using the Hampton Screens yielded no promising hints for crystallization.

4.8.2 Optimization of crystal quality

In order to grow bigger and more compact crystals, suitable for data collection, the components of the crystallization solution have been systematically varied, i.e. different alcohols, PEGs, buffer substances, different amounts of these components, different pH values and different drop ratios. Moreover, a variety of additives, i.e. salts like $MgCl_2$ or organics like ethanol have been cocrystallized with the initial crystallization solution with a drop ratio of 2µl of protein solution, 0.3µl additive and 1µl of reservoir solution. Finally the following crystallization solution yielded compact hexagonal crystals with an approximate size of 70µm x 200µm x 200µm (Fig. 4.20):

crystallization solution: 10% (v/v) isopropanol
0.1M Tris/HCl, pH8.5
10% (w/v) PEG 2000 MME

drop ratio: 2µl $DegP_{S210A}$ (12mg/ml)
0.12µl 0.1M $MgCl_2$
2µl crystallization solution

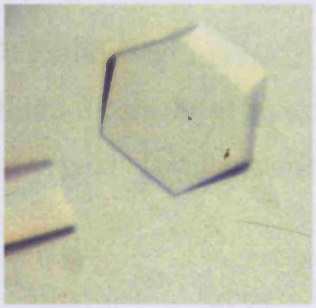

Fig 4.20. Optimized crystals.

4.8.3 Crystallization of different DegP constructs

4.8.3.1 Crystallization of $DegP_{S210A,12\text{-}B}$

$DegP_{S210A,12\text{-}B}$ could be crystallized by using the 'sparse-matrix' approach. The protein crystallized under several conditions which all the following parameters in common: about 10% PEG 4000 to PEG 8000, a buffer with a pH value around neutral and additionally some salts like NaCl or $(NH_4)_2SO_4$. However, all of the initial crystals were rather small and did not show sharp edges, the usual sign for high-quality crystals. A systematic improvement, of the most promising condition, led to 10% PEG 4000, 0.1M HEPES pH 7.5, 0.05M NaCl and a protein to reservoir ratio of 3 to 1. The final crystals are shown in fig. 4.21. Up to now, these crystals did not diffract, however with final dimensions of about 50µm x 50µm x 50µm, they are still quite small. Moreover, due to the limited amount of protein, their number was limited to only a handful of them, thus they may simply have suffered from cryosoaking.

Fig. 4.21. Crystals of $DegP_{S210A,12\text{-}B}$

4.8.3.2 Crystallization of DegP+DFP and $DegP_{(+DFP)}$

The optimized crystallization condition for the $DegP_{S210A}$ protein could be transferred on the DegP+DFP and the $DegP_{(+DFP)}$. The crystals had the same size and shape like the optimized $DegP_{S210A}$ crystals.

4.8.3.3 Crystallization of selenomethionine-substituted DegP's

All the selenomethionine substituted DegP proteins, namely $DegP_{S210A}$(SeMet) and DegP+DFP(SeMet) were crystallized under the same optimized crystallization condition as $DegP_{S210A}$. Only slight adjustments had to be made which were different between different batches from different protein purifications and therefore not a result of the selenomethionine substitution. The shape of the crystals was virtually the same as for the native $DegP_{S210A}$, except that they were more compact (Fig. 4.22).

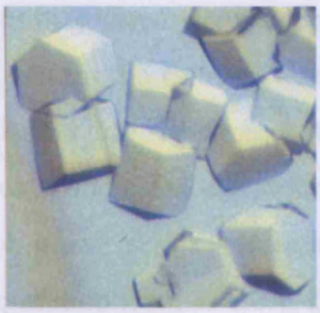

Fig. 4.22. Crystals of $DegP_{S210A}$(SeMet).

4.9 Improvement of the diffraction properties

In Fig. 4.23 a typical diffraction pattern of $DegP_{S210A}$ crystals is shown. The images have been recorded on the in-house X-ray equipment at the MPI for Biochemistry in the Department Strukturforschung.

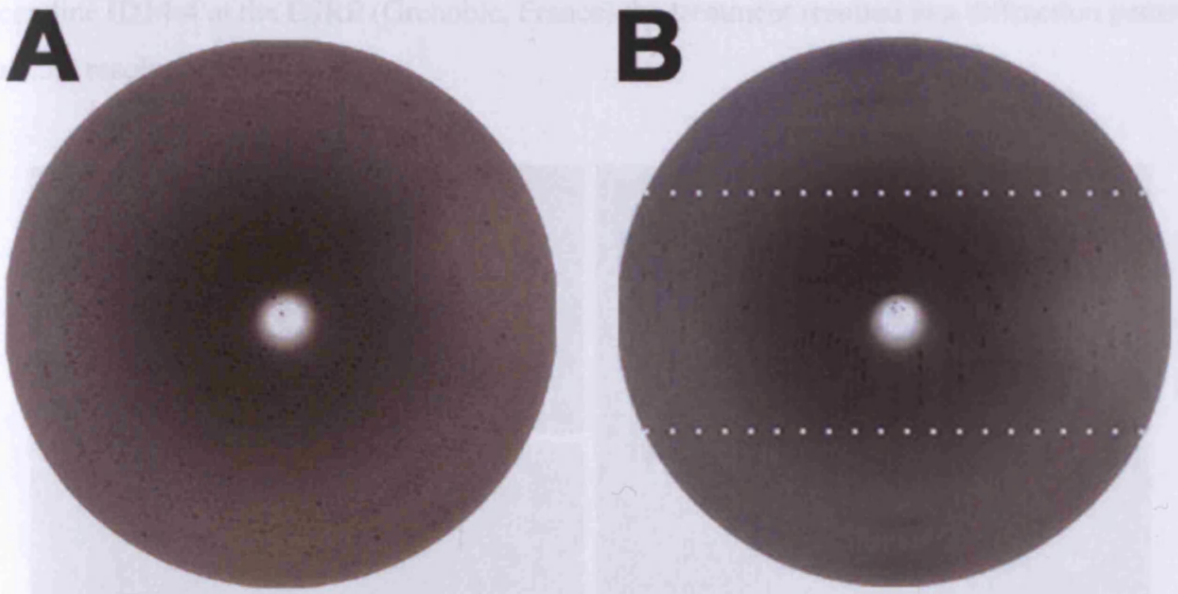

Fig. 4.23. X-ray diffraction images recorded at 100K on a MAR345 image plate; the diffraction limit to the edge of the images is 3.6Å. The images represent two orientations of the same crystal. Once along the real-space c-axis (A) and once perpendicular to the real-space c-axis (B). Note the anisotropy in the diffraction pattern in B, indicated by the dotted white line which marks the resolution limit in the horizontal direction

The spots on the image plate are clearly separated and the crystals diffract to 3.6Å. Unfortunately, the diffraction pattern is extremely anisotropic, i.e. on the left picture, the spots are distributed all over the image, but on the right picture the spots reach the border of the image plate only in the horizontal direction, whereas in the vertical direction no spots are visible beyond 6Å. Anisotropy usually indicates less order in one crystal dimension, in that case along the c-axis. It should be noted, that measurements on beamline BW6 at DESY (Hamburg, Germany) shifted the resolution limit to 2.8Å. But even there the anisotropy in the diffraction pattern limited data collection to 3.8Å.

In order to overcome the anisotropy of the diffraction pattern and to improve the diffraction limit, the crystals grown at 18°C were cooled to 4°C and moreover, the PEG concentration in the reservoir solution was increased to 50%. After approximately one week, an equilibrium between the drop and the reservoir was reached.

The first advantage of the setup is, that the high amount of PEG in the drop serves as a cryoprotectant, so that the crystals were just mounted out of the solution with a cryoloop (Hampton Research, USA) and were directly flash frozen in liquid nitrogen.

The second advantage became obvious when the diffraction properties of the crystals were elucidated. The crystals exhibit now an isotropic diffraction pattern and at the high brilliance beamline ID14-4 at the ESRF (Grenoble, France) the treatment resulted in a diffraction pattern to 2.5Å resolution (Fig. 4.24).

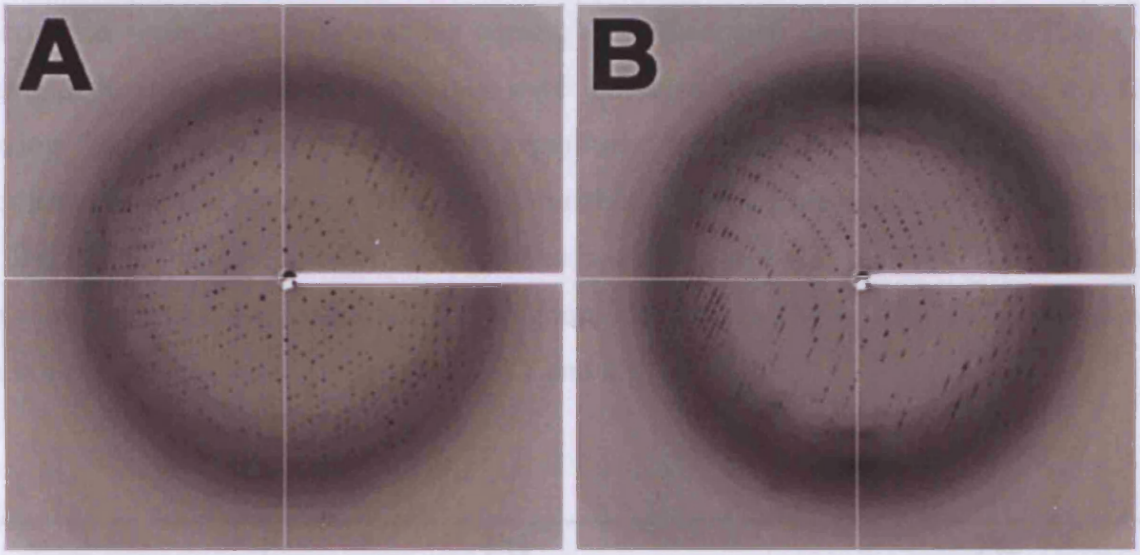

Fig. 4.24. X-ray diffraction images recorded at 100K on a Quantum CCD detector at beamline ID14-4 at the ESRF (Grenoble, France); the diffraction limit to the edge of the images is 2.5Å.

Due to a lack of more systematic experiments, it is not possible to judge if the observed effect arises from the cooling of the crystals to 4°C or if the effect is due to dehydration caused by the high PEG concentration in the drop solution. However, it seems clear that this treatment probably resulted in locking DegP in the "chaperone conformation", due to the low temperature.

Moreover, the treatment resulted in a decrease of the unit cell dimensions along the crystallographic c-axis from 237Å to 233Å. Interestingly, the corresponding c*-axis in the reciprocal space, which one observes in the diffraction pattern, was the axis where data collection was limited to 3.8Å.

4.10 Crystal structure of $DegP_{S210A}$

4.10.1 Data collection and structure determination

Data collection

From now on, all the remaining steps of the structure solution process of $DegP_{S210A}$ have been performed in collaboration with Tim Clausen. Two multiple anomalous diffraction (MAD) experiments have been performed on beamline BW6 at the DESY (Hamburg, Germany). The experiments were made with a native crystal, soaked for two days in 5mM K_2PtCl_4 and a $DegP_{S210A}$(SeMet) crystal. The crystals were briefly soaked into cryobuffer before flash freezing. Moreover a complete MAD experiment on improved $DegP_{S210A}$ crystals was recorded at beamline ID14-4 at the ESRF (Grenoble, France). Data on improved DegP(+DFP) and DegP+DFP crystals were also collected at beamline ID14-4 at the ESRF. All diffraction data were processed and scaled with the programs DENZO and SCALEPACK. A summary of data collection statistics is given in Table 4.7 and 4.8.

	Pt1	**Pt2**
Resolution [Å][1]	20 - 4.5 (4.7 - 4.5)	
Unit cell [Å]	121.9, 121.9, 235.7	
Wavelength [Å]	1.0709	1.0500
$I/\sigma(I)$[1]	15.0 (3.0)	11.0 (2.2)
Completeness[1,3]	94.6 (81.0)	87.8 (70.1)
Redundancy[1,3]	6.0 (3.1)	3.4 (1.9)
R_{sym} (%)[1,2,3]	7.4 (31.1)	7.4 (35.5)

Table 4.7. Data collection statistics: platinum derivative. [1]Highest resolution bin in parenthesis; [2]R_{sym} is the unweighted R-value on I between symmetry mates, [3]Friedel-mates treated as independent reflections.

	Se1	Se2	Se3
Resolution [Å][1]	20 - 2.8 (2.9 - 2.8)		
Unit cell [Å]	121.4, 121.4, 233.7		
Wavelength [Å]	0.9795	0.9797	0.9500
$I/\sigma(I)$[1]	15.1 (2.3)	11.4 (2.0)	14.5 (2.3)
Completeness[1,3]	97.1 (92.2)	89.1 (77.2)	97.0 (92.2)
Redundancy[1,3]	4.4 (3.1)	2.4 (1.8)	4.4 (3.2)
R_{sym} *(%)*[1,2,3]	8.2 (46.6)	7.5 (43.3)	8.3 (46.9)

Table 4.8. Data collection statistics: SeMet. [1]Highest resolution bin in parenthesis; [2]R_{sym} is the unweighted R-value on I between symmetry mates, [3]Friedel-mates treated as independent reflections.

Structure solution

For the two day soak of $DegP_{S210A}$ crystals with 5mM K_2PtCl_4, difference Patterson analysis performed with SOLVE yielded four platinum sites. Subsequently, difference Fourier analyses yielded 21 of the 28 theoretical selenium sites of SeMet crystals. A complete MAD experiment was performed up to 2.8Å resolution. Refinement of heavy atom parameters and phase calculation with SHARP, followed by solvent flattening with SOLOMON resulted in an electron density map of excellent quality (Fig. 4.25). A summary of phasing statistics is given in Table 4.9.

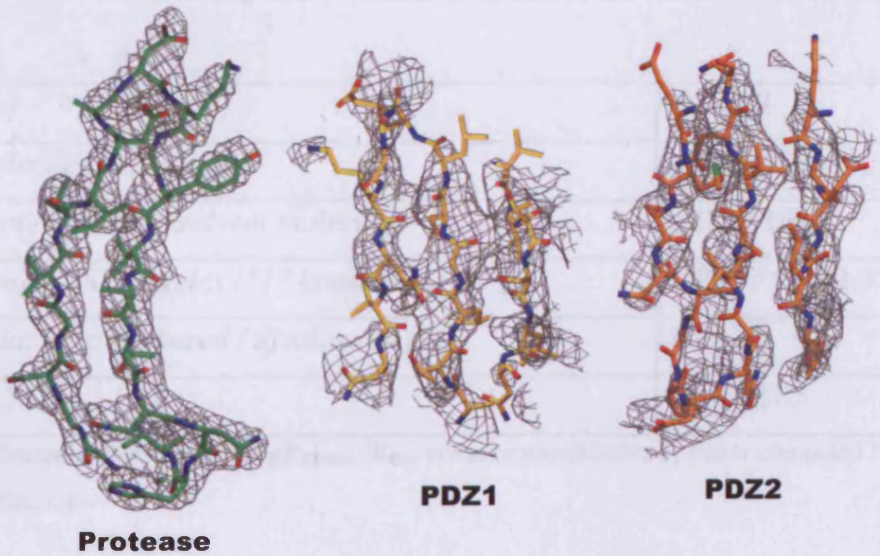

Fig. 4.25. The 2.8Å MAD phased electron density map.

	Se1	Se2	Se3	Pt1	Pt2
Anomalous Scatterer		21 Se		4Pt	
Phasing Power iso/ano	3.14 / 2.74	2.62 / 1.80	- / 2.54	1.23/ 1.17	-/ 0.96
R_{cullis} *iso/ano*	0.51 / 0.67	0.57 / 0.82	- / 0.75	0.75/ 0.78	-/ 0.86
Figure of merit		0.63		0.40	

Table 4.9. Phasing statistics $DegP_{S210A}$.

4.10.2 Model building and refinement

The program O was used to build the initial model. Two molecules are present in the asymmetric unit, which will be referred to as molecule A and B. The electron density map was of excellent quality in the protease region and of lower quality for the PDZ domains. PDZ2 of molecule A was not defined at all. Although the residues of both molecules were generally well defined, residues 1-10, 52-78, 190-194, 354-448 (=PDZ2) in molecule A and residues 1-10, 55-77, 188-195, 370-374, 447, 448 in molecule B could not be fitted to electron density The current model has an R-factor of 21.8% (R_{free}=27.5%). The stereochemistry of the model was validated with PROCHECK. The refinement statistics are summarized in Table 4.10.

Resolution [Å]	20 - 2.8
Number of reflection R_{work}/R_{free}	44793 / 2338
Number of protein atoms / solvent molecules	5228 / 166
Rmsd bond length [Å] / angles [°] / bonded B's [Å^2]	0.012 / 1.7 / 3.3
Ramachandran: most favoured / disallowed [%]	75.4 / 0
R-factor: R_{cryst} / R_{free}[1] [%]	21.8 / 27.5

Table 4.10. Refinement statistics of $DegP_{S210A}$. [1]R_{free} is the crossvalidation R-factor computed for the test set of 5% of unique reflections.

4.10.3 Description of the structure

4.10.3.1 Tertiary structure

After successful crystallization at room temperature, the crystals were transferred to 4°C thereby locking the protein in the "chaperone conformation". This state of $DegP_{S210A}$ will be described in the following chapters. *E.coli* DegP can be divided into three functionally distinct domains, namely a protease (residues 1-259) and two PDZ domains (PDZ1: residues 260-358, PDZ2: residues 359-448) (Fig. 4.26). The previously proposed N-terminal domain (Pallen and Wren, 1997) contributes to the protease fold. Part of this N-terminal segment, the Q-linker (residues 55-79), was too flexible to be traced in the electron density. Like other members of the trypsin family, the protease domain of DegP has two perpendicular β-barrel lobes with a carboxy-terminal helix (e). In both β-barrels, a Greek key motif (strands 1-4, 7-10) is followed by an antiparallel hairpin motif (5-6, 11-12). The catalytic triad is located in the crevice between the two lobes. The N-terminal barrel contributes two catalytic residues, His105 and Asp135, whereas the reactive Ser210 is part of the C-terminal barrel. The overall fold of the two PDZ domains is similar to other PDZ domains of known structure. The first PDZ domain (PDZ1) contains eight β-strands (13-20) and three α-helices (f,g,h). The strands 14-20 form an antiparallel β-sandwich, in which a three-stranded β-sheet (14, 15, 16) crosses a four-stranded β-sheet (18, 17, 19, 20). The short α-helix g and its connecting loop cap one end of the β-sandwich, while helix h caps the other end. Compared to the basic PDZ architecture, PDZ1 contains β13 and αf as additional elements. β13 forms a short two-stranded antiparallel β-sheet with β20 thereby clamping the N- and C-terminal ends of PDZ1. αf is part of a 17-residue insertion that is of great importance for the subunit interplay within the DegP hexamer. In the second PDZ domain (PDZ2), a three-stranded β-sheet (21, 22, 23) is combined with a four-stranded β-sheet (24, 23, 25, 26). The edges of the resulting β-sandwich are capped by the 3_{10} helix i and α-helix j (Fig. 4.26).

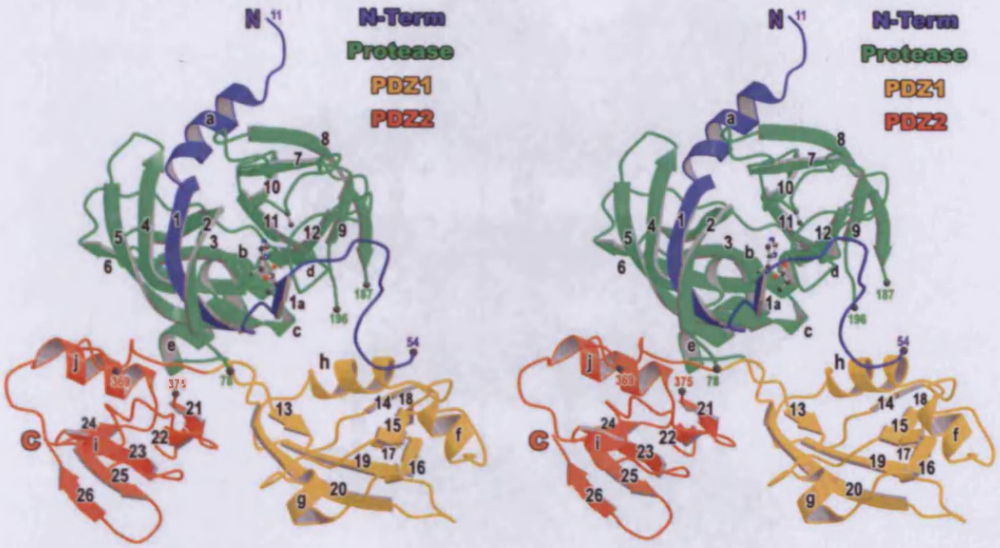

Fig. 4.26. Stereo ribbon presentation of the monomer.

In the asymmetric unit the chaperone form of DegP was observed in two conformations. These will be referred to as molecule A and B. While PDZ2 of molecule A is disordered and not defined by electron density, the PDZ1 domains of molecule A and B differ in their relative orientation and location. Although these differences may be the result of crystal packing constraints, the observation indicates that the PDZ domains are mobile and that the energy barrier to adopt different conformations is low.

4.10.3.2 Assembly of the hexamer

The DegP oligomer is a D3 symmetric hexamer that is formed by a staggered association of trimeric rings. The A and B hexamers are centred in the crystallographic unit cell at (0, 0, 0), (2/3, 1/3, 1/4), respectively (Fig. 4.27) and represents two distinct states.

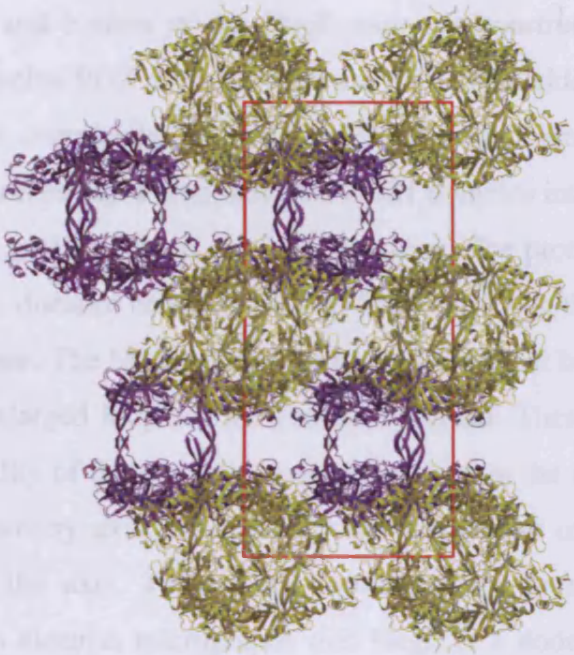

Fig. 4.27. Crystal packing in $DegP_{S210A}$ crystals viewed along the crystallographic b-axis. Two molecules A (lilac) and B (yellow) were observed in the asymmetric unit.

Molecule A is a largely open structure with a wide lateral passage penetrating the entire oligomer, whereas molecule B corresponds to the closed form, in which a cylindrical 45Å cavity is completely shielded from solvent (Fig. 4.28).

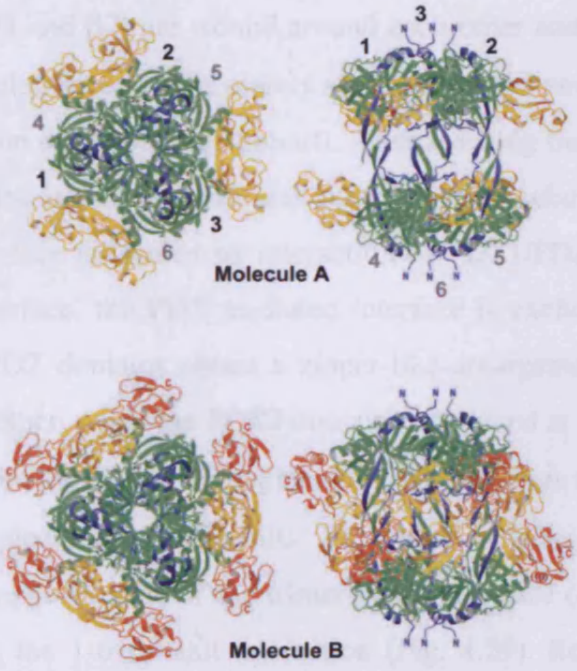

Fig. 4.28. Top (left) and side (right) views of the DegP hexamer constructed by molecules A and B.

In both cases, the top and bottom of the DegP cage are constructed by the six protease domains, whereas the twelve PDZ domains generate the mobile sidewalls. Because the axial pores of the particle are completely blocked, the PDZ domains are the only gates allowing lateral access to the central cavity. Particularly the PDZ1 domains interact with each other and should thus be the main gatekeepers of the inner chamber. The proteolytic sites are enclosed within this chamber. The distance between the sites is about 25Å in the trimeric rings and 40Å across the equatorial plane. The height of the cavity is determined by three molecular pillars, which are formed by enlarged loops of the protease domain. These pillars are also mainly responsible for the stability of the dynamic complex as seen in the open state. When viewed down the threefold symmetry axis, the hexamer has a diameter of 120Å and a maximum extent of 105Å along the axis. The overall toroidal shape is not compatible with the conclusions drawn from electron micrographs that DegP is a dodecamer consisting of two stacks of hexameric rings organized around a large central pore (Kim *et al.*, 1999).

The contacts between the trimeric rings arise almost exclusively from highly flexible structural elements. Thus the DegP hexamer should be considered as a loosely bound dimer of tight trimers. All residues involved in the intersubunit contacts are highly conserved in the DegP family. Monomers associate into trimers by formation of several larger hydrophobic clusters, which are enclosed by a net of intersubunit hydrogen bonds. The resulting trimeric rings interact via the 1-4 and 1-6 interface. In the 1-4 interface, the enlarged protease loops termed LA (connecting β1 and β2) are wound around each other and build the corner pillars of the DegP cage. The molecular spacer is mainly stabilized by the two stranded β-sheet 1'/2* (* denotes the participation of the partner subunit). After reaching the opposite ceiling of the cavity, loop LA protrudes into the active site of its partner subunit. In the hexamer of molecule B, the 1-6 interface is formed by interaction of PDZ1/PDZ2 with their symmetry mates. As in the 1-4 interface, the PDZ mediated interface is exclusively accomplished by polar interactions. The PDZ domains obtain a zipper-like arrangement, in which the PDZ1 domains are facing each other, while the PDZ2 domains are bound at their edges. Residues of this interface originate from the αf-β15-αg segment (PDZ1) and from β22-αi (PDZ2).

In the open form of molecule A, PDZ1 tilts 70° away, thereby moving the αf-β15-αg interaction clamp to the opposite side of the trimer-trimer interface (corresponding to a 30Å movement) and breaking the 1-6 subunit interaction (Fig. 4.29). Reorientation of PDZ1 is achieved by a twist of the polypeptide backbone between residues Arg262/Gly263.

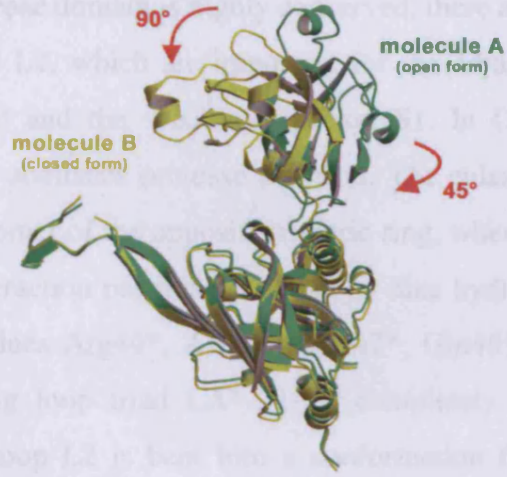

Fig. 4.29. Superposition of molecule A and molecule B.

4.10.3.3 The protease domain

The Dali algorithm was employed to search for structural homologues of the protease domain. *Streptomyces griseus* trypsin (SGT), epidermolytic toxin, β-trypsin and thrombin were identified as the most similar structures in the database. The superposition between DegP and SGT complexed with a tetrapeptide is shown together with the common loop nomenclature in Fig 4.30.

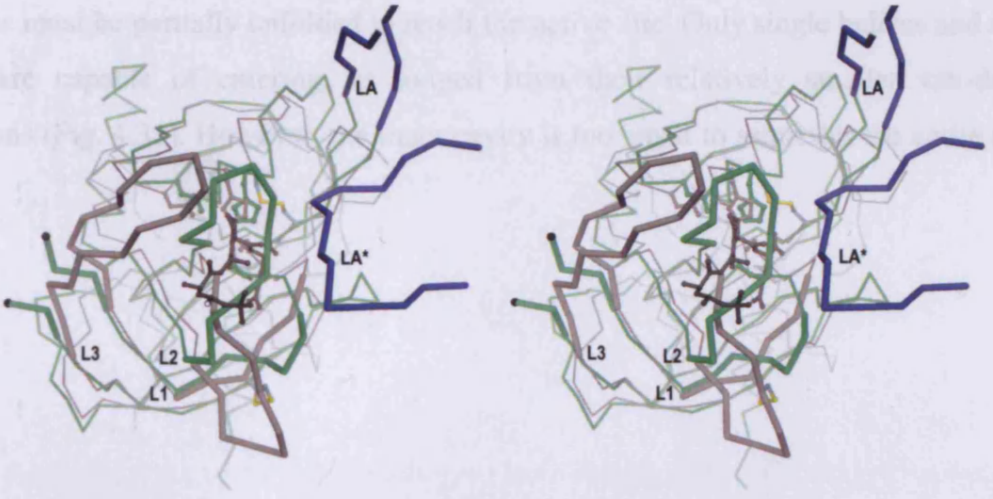

Fig. 4.30. Structural alignment of the DegP protease domain (green and blue) with SGT (gray); in the stereo image both backbones are shown as Cα trace whereas the catalytic triads, the peptide bound to SGT and its disulphide bridges are drawn in ball-and-stick mode; some of the mechanistic important loops (L1, L2, L3, LA and LA*) are emphasized by thicker lines and are labelled; note that the DegP loop LA* (blue) originates from the partner monomer; loop L3 of DegP was only partially defined in the electron density.

While the core of the protease domain is highly conserved, there are striking differences in the surface loops LA, L1 and L2, which are important for the adjustment of the catalytic triad (His105, Asp135, Ser210) and the specificity pocket S1. In DegP, these loops acquire a twisted conformation that abolishes protease function. The enlarged loop LA protrudes into the active site of one monomer of the opposite trimeric ring, where it intimately interacts with loops L1 and L2. This interaction patch is stabilized by nine hydrogen bonds formed between the highly conserved residues Arg44*, Asn45*, Gln47*, Gln48*, Asn209, Asp232, Gly233 and Gly234. The resulting loop triad LA*-L1-L2 completely blocks the entrance to the catalytic site. Especially loop L2 is bent into a conformation that closes the active site, a feature that has not been reported before.

4.10.3.4 Characteristics of the inner cavity

E.coli DegP has the ability to stabilize a number of non-native proteins *in vivo* and *in vitro* (Spiess *et al.*, 1999). Analysis of a mutant lacking both PDZ domains assigned this chaperone function to the protease domain (Spiess *et al.*, 1999). Therefore, possible binding sites for misfolded proteins should be located within the inner cavity which is entirely constructed by residues of the protease domain. The solvent accessible height of this chamber is 15Å at its centre and increases to 18Å near the outer entrance. Due to these geometric constrictions, substrates must be partially unfolded to reach the active site. Only single helices and extended strands are capable of entering, as judged from their relatively smaller van-der-Waals dimensions (Fig. 4.31). However, the inner cavity is too small to sequester the entire unfolded protein.

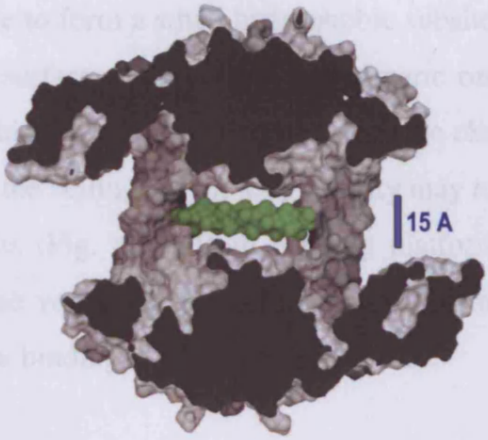

Fig. 4.31. Surface representation of the internal tunnel (side view) illustrating its molecular sieve character; the height of the cavity is indicated; access is restricted to single secondary structure elements as shown by the modeled polyalanine helix (green).

As in other chaperones of known structure, the DegP cavity is lined by hydrophobic residues. In both trimeric rings, three large hydrophobic grooves are organized around the central Gln206/Arg207 cluster and extend towards the PDZ1 domain. (Fig. 4.32)

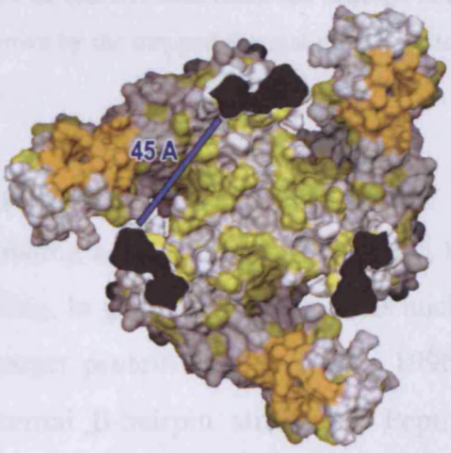

Fig. 4.32. Formation of the hydrophobic binding patches within the cavity (top view); hydrophobic residues of the protease domain are shown in yellow, and the non-polar peptide binding groove of PDZ1 in orange.

The hydrophobic grooves are mainly constructed by residues of loop LA and L2 (Phe46, Phe49, Ile228, Leu229, Pro231, Ile236, Ile238 and Phe240). The three hydrophobic patches are separated from each other by stretches of polar residues. A second potential binding site

was observed on the internal side of the three pillars of the cavity, where Thr35, Val37, Pro40, Met42 and Phe84 assemble to form a small hydrophobic subsite. The alternating arrangement of polar and hydrophobic surfaces, both within one trimeric ring and between trimeric rings, should form the basis for binding exposed hydrophobic side chains and the peptide backbone atoms of substrates. Thus, the ceilings of the DegP cavity may represent docking platforms for partially denatured proteins (Fig. 4.33) Both docking platforms are structurally flexible as indicated by their backbone variations and by their high thermal motion factors (Fig. 4.24). This plasticity should allow binding of diverse polypeptides.

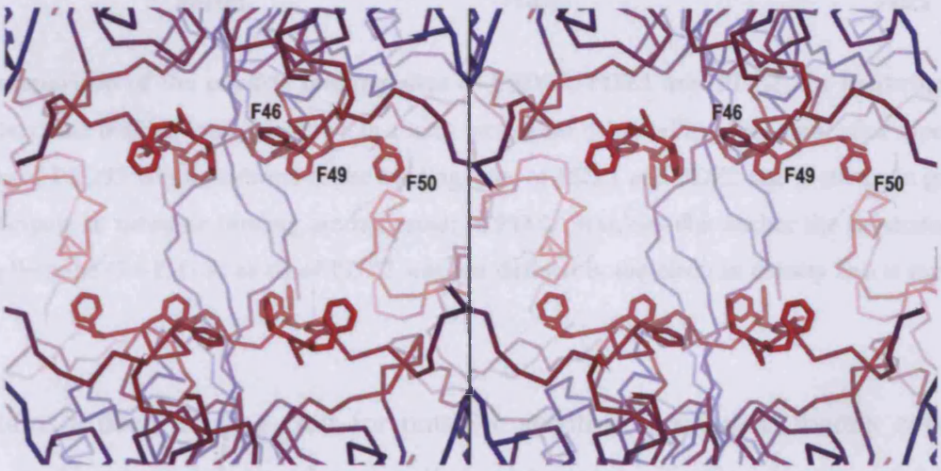

Fig. 4.33. Stereo view of the two Phe clusters that form the ceilings of the internal cavity (side view); these platforms are highly mobile, as shown by the mapped thermal motion factors (blue, rigid; red, flexible); the Phe triplet of one monomer is labelled.

4.10.3.5 Substrate binding properties of the PDZ domains

Beside their function in regulating access to the inner cavity, the PDZ domains of DegP may be involved in substrate binding. In general, PDZ domains anchor the three or four C-terminal residues of the respective target protein (Doyle *et al.*, 1996; Songyang *et al.*, 1997), but sometimes also bind to internal β-hairpin structures. Peptide ligands interact with PDZ domains in a β-augmentation process, in which a peptide is bound as an additional β-strand to the small PDZ β-sheet (Harrison, 1996). To identify the determinants of substrate specificity of PDZ1 and PDZ2, we aligned both structures with the peptide complex of PSD95 (Doyle *et al.*, 1996) and modelled the bound ligand to both PDZ domains of DegP (Fig. 4.34)

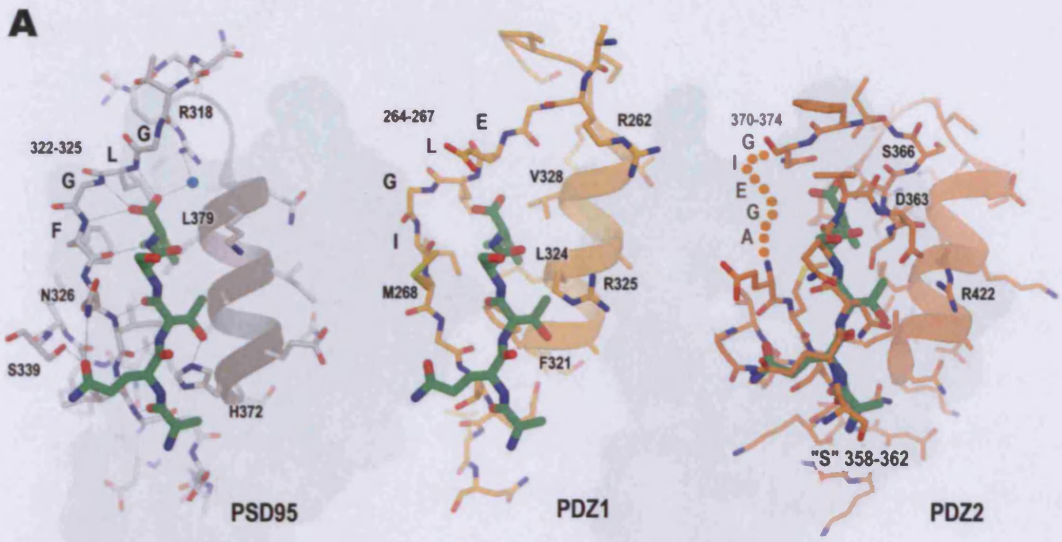

Fig. 4.34. Comparison of the peptide binding sites of PSD95, PDZ1 and PDZ2; the bordering β-strand, α-helix and carboxylate binding loop are shown in a stick model are coloured by atom type; after superposition, the bound peptide of PSD95 was transferred to the binding sites of PDZ1 and PDZ2 and is shown in green; residues that may participate in substrate binding are indicated; in PDZ2, residues that anchor the substrate-like segment ("S") are labelled; the G-I-E-G-A motif of PDZ2 was not defined in the electron density and is represented by a dashed coil.

PDZ1 contains a deep binding cleft for putative substrates, which is mainly constructed by strand 14, its N-terminal loop (the so-called carboxylate binding loop) and helix h. The carboxylate-binding loop is located in a highly positively-charged region and is formed by an E-L-G-I motif, which is similar to the frequently observed G-L-G-F motif (Cabral *et al.*, 1996). The strictly conserved Arg262 is properly oriented to bind further the carboxylate loop of the substrate. Since this arginine is also forming the hinge between protease and PDZ1, substrate binding to Arg262 might trigger reorientation of PDZ1. Binding specificity is mainly conferred by the specific configuration of the 0, -2, -3 binding pockets, where pocket 0 anchors the sidechain of the carboxy-terminal residue. In PDZ1, all the pockets are built by mainly hydrophobic residues (Fig. 4.35).

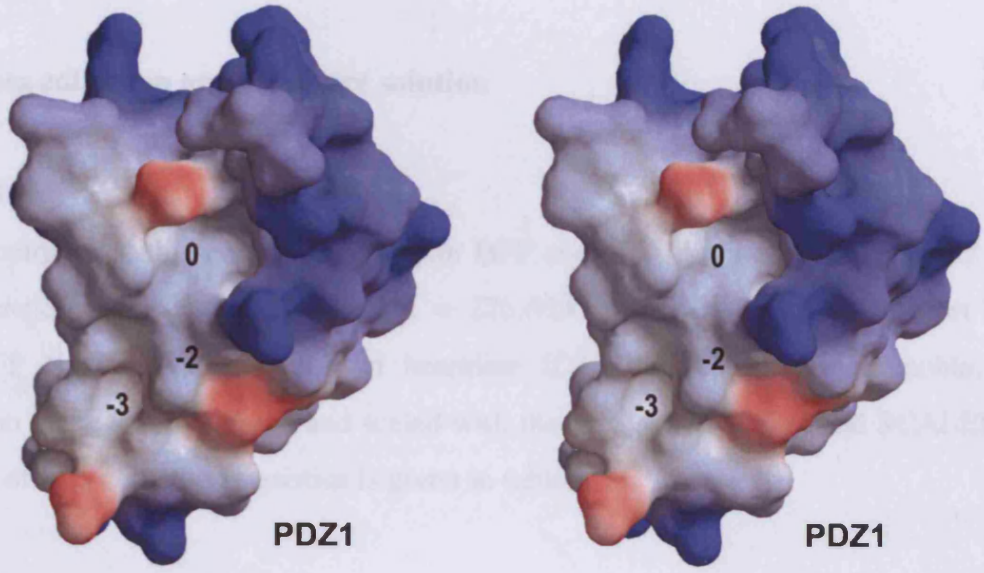

Fig. 4.35. Electrostatic surface representation of the PDZ1 binding site in stereo; acidic regions are shown in red, basic regions in blue and non-polar regions in white; the location of the hydrophobic 0, -2 and -3 binding pocket is indicated.

Thus, PDZ1 seems to be well adapted to bind stretches of hydrophobic peptide ligand. One of the main differences to other PDZ domains is the flexibility of strand 14 and its associated carboxylate binding loop, indicating the plasticity of the binding site. It was proposed that the carboxylate-binding loop serves as a steric block preventing interaction with peptides that extend beyond the recognition sequence (Hillier *et al.*, 1999). The linear arrangement of the peptide binding sites of PDZ1 and PDZ2, might suggest that this block could be rearranged allowing the concerted interaction of one substrate with both PDZ domains.

The comparison of PSD95 and PDZ2 suggests a novel mechanism of substrate binding and release. In PDZ2, the loop 355-369 adopts a conformation that mimics a bound substrate molecule (Fig. 4.34). Residues 358-362 are bound in extended conformation, antiparallel to strand 21, and several interaction occur with helix j. This substrate-like segment is connected by the carboxylate-binding loop (comprising the G-I-E-G-A motif) to the remainder of PDZ2. The highly flexible carboxylate-binding loop might allow reorientation of the substrate-like segment and thereby opening and closing of the peptide-binding site. As the occluded pocket was observed in the closed form of the DegP cage, displacement of the bound substrate by the loop 355-369 might be involved in the process of substrate translocation.

4.11 Crystal structure of DegP+DFP

4.11.1 Data collection and structure solution

Strategy 1:

DegP in complex with the covalent inhibitor DFP crystallized in space group $P6_322$ with unit cell dimensions of a = b = 121.43Å, c = 226.65Å. A highly redundant dataset of native DegP+DFP crystals was collected at beamline ID14-2 at the ESRF (Grenoble, France). Diffraction data were integrated and scaled with the programs DENZO and SCALEPACK. A summary of data collection statistics is given in table 4.11.

	DegP+DFP
Resolution [Å][1]	20 – 2.90 (2.97 – 2.90)
Unit cell [Å]	121.43, 121.43, 226.65
Wavelength [Å]	0.9330
$I/\sigma(I)$[1]	34.8 (5.6)
Completeness[1]	91.7 (58.8)
Redundancy[1]	9.7
R_{sym} (%)[1,2]	6.2 (25.1)

Tab. 4.11. Data collection statistics: DegP+DFP. [1]Highest resolution bin in parenthesis; [2]R_{sym} is the unweighted R-value on I between symmetry mates.

The observed diffraction limit of 2.9Å, was comparable to the $DegP_{S210A}$ crystals. However, for the DegP+DFP crystals, the dehydration protocol, described in section 4.10.3, which was crucial for the determination of the initial $DegP_{S210A}$ structure, did not significantly reduce the anisotropy in the diffraction pattern.

For structure solution, the $DegP_{S210A}$ model was used. Residues 205 to 215 were omitted in both monomers of the asymmetric unit and the temperature factors for all residues were reset to 70Å^2. Moreover, 2-fold non-crystallographic symmetry restraints were imposed on the remaining residues of the protease domains. After one round of rigid-body refinement (each protease and PDZ domain in the asymmetric unit was considered as a rigid entity), followed

by several rounds of positional and grouped B-factor refinement, R_{cryst} and R_{free} dropped to 31.1% and 35.9%, respectively

Strategy 2:

The DegP+DFP(SeMet) protein crystallized in space group $P6_322$ with unit cell dimensions a = b = 121.38Å, c = 226.87Å. A complete MAD dataset was collected at beamline ID29 at the ESRF (Grenoble, France). Diffraction data were integrated and scaled with the programs DENZO and SCALEPACK. A summary of data collection statistics is given in table 4.12.

	Se1	Se2	Se3
Resolution [Å][1]		20 – 3.15 (3.22 – 3.15)	
Unit cell [Å]		121.38, 121.38, 226.87	
Wavelength [Å]	0.9791	0.9792	0.9753
$I/\sigma(I)$[1]	20.1 (3.0)	15.1 (2.3)	13.7 (1.8)
Completeness[1,3]	96.2 (80.5)	91.7 (68.0)	90.7 (57.6)
Redundancy[1,3]	6.8	3.4	3.2
R_{sym} *(%)*[1,2,3]	9.5 (43.7)	8.7 (43.8)	9.3 (45.5)

Table 4.12. Data collection statistics: DegP+DFP(SeMet). [1]Highest resolution bin in parenthesis; [2]R_{sym} is the unweighted R-value on I between symmetry mates, [3]Friedel-mates treated as independent reflections.

The determination of the heavy atom substructure was done with direct methods using Shake and Bake (Weeks and Miller, 1999) and 15 out of 28 expected selenium positions could be identified. Refinement of heavy atom parameters, phase calculations and solvent flattening were done with CNS (Brunger *et al.*, 1998). A summary of phasing statistics is given in table 4.13.

	Se1	Se2	Se3
Anomalous Scatterer		15 Se	
Phasing Power iso/ano	0.76 / 1.48	1.01 / 1.64	- / 1.11
R_{cullis} *iso/ano*	0.88 / 0.74	0.84 / 0.72	- / 0.82
Figure of merit		0.57	

Table 4.13. Phasing statistics DegP+DFP(SeMet).

The resulting electron density was only well defined in the rigid parts of the protease domain (Fig. 4.36). Unfortunately, there was hardly any density for the PDZ domains.

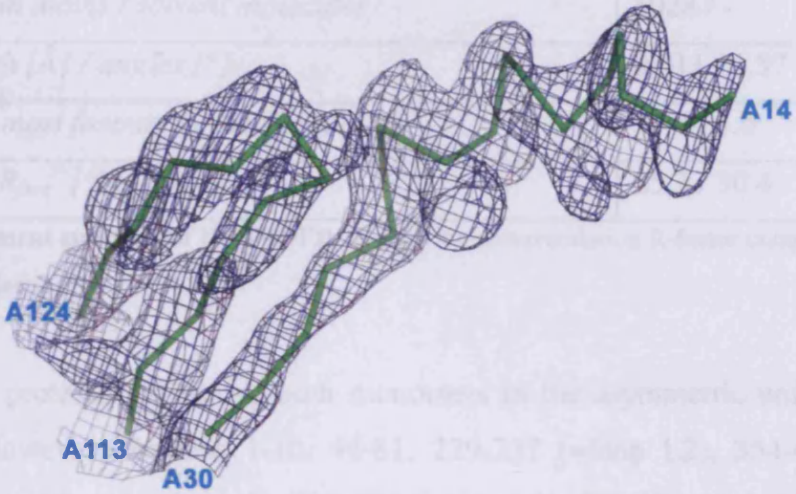

Fig. 4.36. The 3.15Å phased electron density map of the protease domain of DegP+DFP(SeMet).

4.11.2 Model building and refinement

The model obtained after the application of strategy 1 was used as the starting point. Inspection of the resulting 2fofc (contoured at 1.0 sigma) and fofc (contoured at ±3.0 sigma) electron density maps indicated structural changes throughout the active site and rather weak density in the region of the PDZ domains of both molecules. After several rounds of manual rebuilding and subsequent refinement of loops L1, L2, L3 and LA it became evident, that their course could not be modelled unambiguously. Although there was residual density in the fofc-map (contoured at +3.0 sigma) for the position where the DFP moiety was anticipated, its exact orientation did not become clear. Due to the impossibility to build an accurate model with the available data, it was considered, that a possible source for these ambiguities might be phase bias, originating from the input model.

Thus, additional experimental phases from a three-wavelength MAD experiment were incorporated (strategy 2). Although the electron density map after phasing was of minor quality, the resulting map after combining experimental and model phases led to a reasonable electron density map. After several rounds of refinement using the MLHL target and subsequent manual rebuilding, refinement converged to an R_{cryst} of 25.4% and R_{free} of 30.4%, respectively. The refinement statistics are summarized in table 4.14.

Resolution [Å]	15.0 – 2.9
Number of reflection R_{work}/R_{free}	19616 / 994
Number of protein atoms / solvent molecules	5028 / -
Rmsd bond length [Å] / angles [°]	0.011 / 1.57
Ramachandran: most favoured / disallowed [%]	74.4 / 1.0
R-factor: R_{cryst} / R_{free}[1] [%]	25.3 / 30.4

Table 4.14. Refinement statistics of DegP+DFP. [1]R_{free} is the crossvalidation R-factor computed for the test set of 5% of unique reflections.

Residues of the protease domain of both monomers in the asymmetric unit were generally well defined. However, residues 1-10, 44-81, 229-237 (=loop L2), 354-448 (=PDZ2) in molecule A and residues 1-10, 43-77, 229-233 (=loop L2), 370-374, 447, 448 in molecule B could not be fitted to electron density. It should be noted, that most of the side chains of the PDZ domains in molecule A and molecule B were only weakly defined. However, omitting them from the model led to an increase in R_{cryst} and R_{free} by about two percent, thus they were included into the final model. Only the main chain course of loop L3 (residues 187-196) could be traced. The six residues in the disallowed region of the Ramachandran plot, namely 41(A), 191(A), 193(A), 192(B), 411(B), 440(B) all lie in regions that are not well defined in the electron density.

4.11.3 Overall structure and comparison with the $DegP_{S210A}$ structure

Structural differences between DegP+DFP and $DegP_{S210A}$ are confined to loops, associated with the active site, the rest of the structure remained unaffected by the covalent incorporation of the diisopropyl moiety. The root-mean-square difference between the protease domains of DegP+DFP and $DegP_{S210A}$ is 0.75Å over 196 C_α atoms. When the two structures were aligned, the most obvious differences are visible in loop D, comprising residues 170 to 175 and in loop L1 comprising residues 203 to 212 (Fig. 4.37). Loop L2 is not defined in the DegP+DFP model.

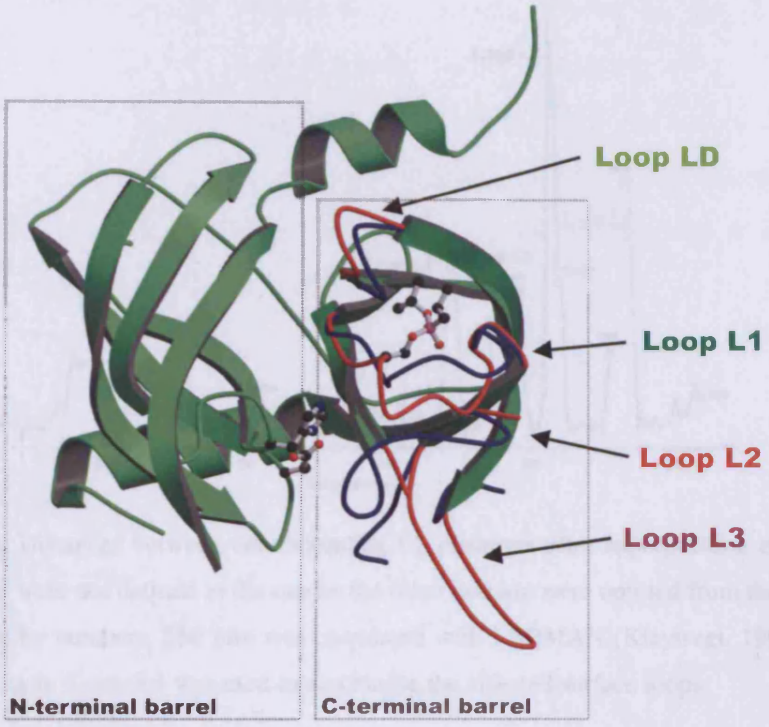

Fig. 4.37. Structural alignment of the protease domains of $DegP_{S210A}$ and DegP+DFP. The unaffected parts of the protease domain are coloured in green, loops with a different conformation are coloured in red (DegP+DFP) and blue ($DegP_{S210A}$), respectively. Residues of the active site (His105, Asp135 and DFP bound to Ser210) are shown as ball-and-stick models. Note that loop L3 is absent in the $DegP_{S210A}$ model; in the DegP+DFP model this loop could only be built as a poly-alanine trace, its side-chain conformations did not become clear.

Loops L1 and LD are somehow pushed apart in order to accommodate the inhibitor. The distances between individual C_α atoms after alignment of the two protease domains are illustrated in figure 4.38. The maximum distance for loop LD is about 3Å, but for loop L1, the maximum distance is nearly 7Å.

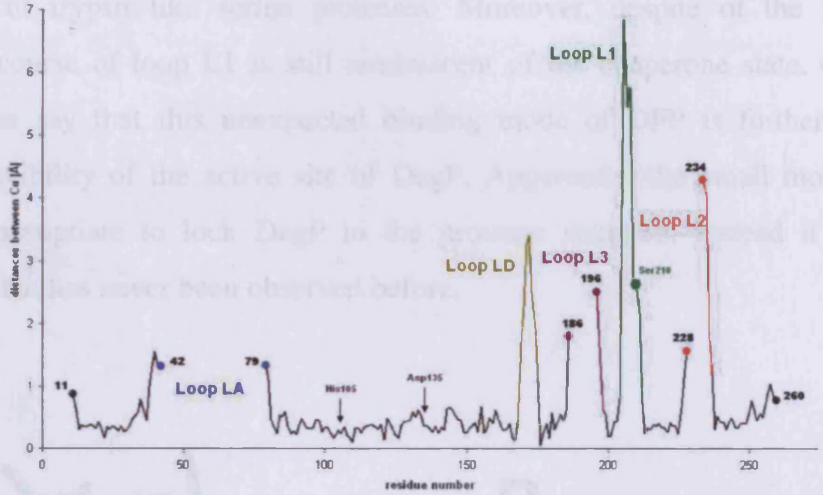

Fig. 4.38. Distance plot. Distances between corresponding C_α positions after superposition of $DegP_{S210A}$ and DegP+DFP. Residues that were not defined in the one or the other domain were omitted from the calculation and are indicated in the figure by numbers. The plot was calculated with LSQMAN (Kleywegt, 1996). Please note, that the same colour code as in figure 5.1 was used to emphasise the affected surface loops.

Loop L2, which in concert with loops L1 and LA* blocks access to the active site in the chaperone form of DegP was not defined from residue 229 to residue 235. Consequently loop LA could only be traced to residue Met42. The situation is different for the previously not defined loop L3. Though its side chain interactions did not get clear, its approximate course could be modelled as a poly-alanine trace (Fig. 4.37).

4.11.4 The active site

In a crystal structure of γ-chymotrypsin in complex with DFP, the uncharged phosphate oxygen atom of the DFP group is locked in the oxyanion hole by forming two hydrogen bonds to Ser195 N and Gly193 N. One of the isopropyl moieties of DFP extends into the hydrophobic S1-specificity pocket of chymotrypsin (Fig. 4.39(B)) (Harel *et al.*, 1991). However, in the DegP+DFP structure, the situation is completely different. The serine 210 with the covalently bound diisopropyl moiety is flipped by about 180° in contrast to chymotrypsin and the wild-type DegP structure (Fig. 4.39 (A) and Fig. 4.40). Thus, instead of making hydrogen bonds to the oxyanion hole, the phosphate oxygen atom of the inhibitor makes a hydrogen bond the backbone amide of Ile205 and the diisopropyl moiety protrudes into the direction of loop D (Fig. 4.37). Thus the active site of DegP has not adopted the

conformation of trypsin-like serine proteases. Moreover, despite of the binding of the inhibitor, the course of loop L1 is still reminiscent of the chaperone state. Comprising the results one can say that this unexpected binding mode of DFP is further indicating the remarkable flexibility of the active site of DegP. Apparently, the small molecule inhibitor DFP was inappropriate to lock DegP in the protease state but instead it is bound in a conformation that has never been observed before.

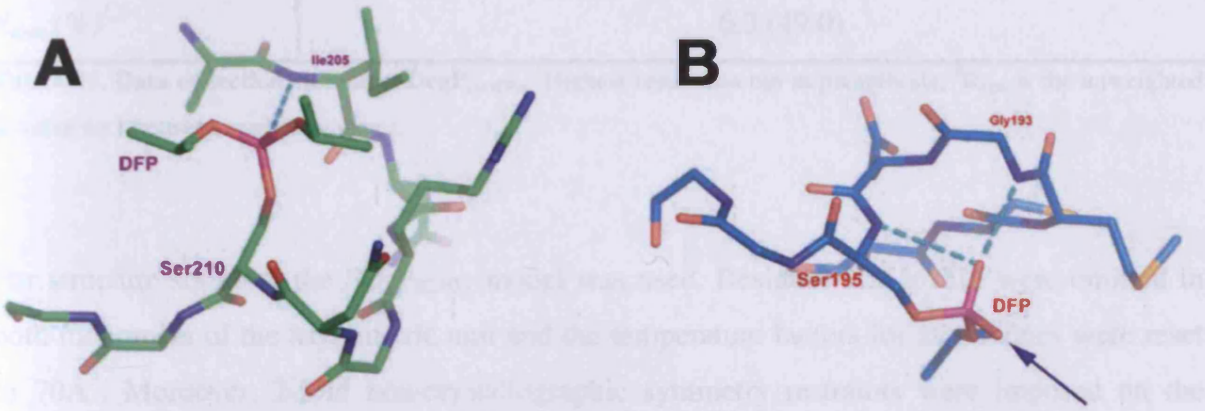

Fig. 4.39. Active site loop. A: DegP+DFP. **B:** γ-chymotrypsin+DFP (PDB entry code 1GMH). Note that the second isopropyl moiety is absent in B, because it was not properly defined in the electron density map (see arrow) (Harel *et al.*, 1991). Hydrogen bonds are depicted as dotted lines coloured in cyan.

4.12 Crystal structure of wild-type DegP ($DegP_{(+DFP)}$)

4.12.1 Data collection and structure determination

The wild-type DegP ($Degp_{(+DFP)}$) crystallized in space group $P6_322$ with unit cell dimensions a = b = 121.43Å, c = 226.65Å. A dataset of native $DegP_{(+DFP)}$ crystals was collected at beamline ID14-4 at the ESRF (Grenoble, France) and diffraction data were integrated and scaled with the programs DENZO and SCALEPACK. After the application of the dehydration procedure, the crystals diffracted isotropically up to 2.5Å. A summary of data collection statistics is given in table 4.15.

	$DegP_{(+DFP)}$
Resolution [Å][1]	20 - 2.5 (2.56 - 2.50)
Unit cell [Å]	120.79, 120.79, 233.51
Wavelength [Å]	0.9393
$I/\sigma(I)$[1]	35.7 (3.2)
Completeness[1]	99.1 (99.6)
Redundancy[1]	4.2 (2.1)
R_{sym} *(%)*[1,2]	6.3 (49.0)

Tab. 4.15. Data collection statistics: $DegP_{(+DFP)}$. [1]Highest resolution bin in parenthesis; [2]R_{sym} is the unweighted R-value on I between symmetry mates.

For structure solution, the $DegP_{S210A}$ model was used. Residues 205 to 215 were omitted in both monomers of the asymmetric unit and the temperature factors for all residues were reset to 70Å^2. Moreover, 2-fold non-crystallographic symmetry restraints were imposed on the remaining residues of the protease domains. After one round of rigid-body refinement (each protease and PDZ domain in the asymmetric unit was considered as a rigid entity), followed by several rounds of positional and grouped B-factor refinement, both, R_{cryst} and R_{free} dropped well below 40%.

4.12.2 Model building and refinement

The initial electron density map was of excellent quality throughout the protease domain and there was clear density in the fofc map (contoured at +3 sigma) for the γ-oxygen atom of serine 210. Surprisingly, despite of the excellent resolution limit, both PDZ domains of molecule A were not defined in the electron density. The PDZ1 domain of molecule A was therefore omitted from the final model. After several rounds of refinement and subsequent manual rebuilding, refinement converged to an R_{cryst} of 26.5% and R_{free} of 29.7%, respectively. The refinement statistics are summarized in table 4.16.

Resolution [Å]	15.0 – 2.5
Number of reflection R_{work}/R_{free}	33389 / 1740
Number of protein atoms / solvent molecules	4477 / 1740
Rmsd bond length [Å] / angles [°]	0.007 / 1.28
Ramachandran: most favoured / disallowed [%]	80.7 / 0
R-factor: R_{cryst} / R_{free} [%]	26.5 / 29.7

Table 4.16. Refinement statistics of $DegP_{(+DFP)}$. [1]R_{free} is the crossvalidation R-factor computed for the test set of 5% of unique reflections.

Like in the other DegP structures, several parts of the molecule were not defined in the electron density. In molecule A, residues 1-10, 50-78, 187-196 (=loop L3), 264-448 (=PDZ1 & PDZ2) and in molecule B, residues 1-10, 51-77, 187-195 (=loop L3), 370-374 could not be fitted to electron density.

4.12.3 Overall structure and comparison with the $DegP_{S210A}$ structure

When the structures of $DegP_{(+DFP)}$ and $DegP_{S210A}$ are superimposed, there are hardly any apparent differences concerning the course of the main chain. Consequently, the root-mean-square difference between the protease domains of $DegP_{(+DFP)}$ and $DegP_{S210A}$ is 0.30Å over 213 C_α atoms. Modest structural differences are restricted to the environment of serine 210. Position and orientation of PDZ1 and PDZ2 of molecule B are the same as in the $DegP_{S210A}$ structure.

4.12.4 The active site

The Ser 210 is clearly oriented towards the loop L2. The active site loop does not acquire the typical conformation of trypsin-like serine proteases. There is still no oxyanion hole formed and access towards the active site stays blocked by the afore mentioned loop triad, namely loop L1, L2 and LA*. This inactive conformation is further stabilized by a hydrogen-bonding network of the Ser210 Oγ towards the backbone oxygen atoms of Gly233 and Gly234. Furthermore the oxygen of Gly 233 makes a hydrogen bond to the amide nitrogen of Ser210

(Fig.4.40). Although the structure of wild-type DegP does not give any clues about the conformation of the protease state of DegP, it does clearly show that the unusual geometry observed in the crystal structure of $DegP_{S210A}$ is not due to the artificial serine to alanine mutation, which renders the protein proteolytically inactive. This seems to be an inherent feature of the chaperone state of DegP in order to prevent unwanted proteolysis at room temperature.

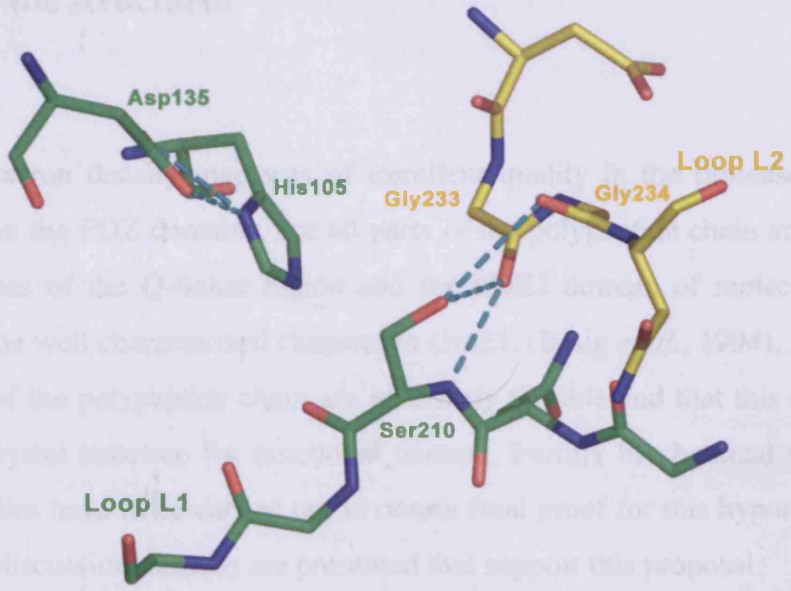

Fig. 4.40. Active site of DegP(+DFP). Hydrogen bonds are depicted as dotted lines coloured in cyan.

Chapter 5

Discussion

5.1 Quality of the structures

5.1.1 $DegP_{S210A}$

Although the electron density map was of excellent quality in the protease region and of medium quality in the PDZ domains, not all parts of the polypeptide chain are well resolved. Especially residues of the Q-linker region and the PDZ2 domain of molecule A were not defined. As for the well characterized chaperonin GroEL (Braig *et al.*, 1994), we also propose that these parts of the polypeptide chain are inherently flexible and that this region is poorly ordered in the crystal structure for functional reasons. Further biochemical work as well as mutagenesis studies have to be carried out to obtain final proof for this hypothesis. However, in the following discussion, reasons are presented that support this proposal.

5.1.2 DegP+DFP

The quality of the final phase combined electron density was excellent in the rigid parts of the protease domain. Throughout the PDZ domains, the main chain was several times interrupted and side chains were rarely observed. Indeed, the position of the domains was clear and side chains were kept in the final model as in the $DegP_{S210A}$ structure, because when these residues were omitted, R_{free} increased by about two percent during refinement. The mechanistically important surface loops, which were already quite flexible in the chaperone state as indicated by increased thermal motion factors became even more flexible and loops L2 and LA completely disappeared in the electron density map. The rough course of the previously not defined loop L3 could be traced as a poly-alanine model. However, sidechain interaction did not become clear for this part of the structure.

5.1.3 $DegP_{(+DFP)}$

Although the crystals of $DegP_{(+DFP)}$ were by far the best diffracting of the three proteins, the PDZ domains of molecule A were not defined at all. Of course, the side chains, especially in the protease domain were much better defined than in $DegP_{S210A}$ or DegP+DFP, however, the structure did not give any novel insights into parts of the structure that were missing in the previously solved $DegP_{S210A}$ model. For the resolution limit of 2.5Å, the final R-factors of refinement are quite high with an R_{cryst} of 26.5% and an R_{free} of 29.7%, respectively. This may be attributed to the fact that significant parts of the protein are highly flexible. Although these parts could not be modelled, they are of course present and contribute partially to the diffraction pattern.

5.2 Quaternary structure of DegP in solution

Through size-exclusion chromatography and dynamic light scattering measurements, it could be confirmed that $DegP_{S210A}$ is predominantly present as a hexamer in solution. In addition, small amounts of higher molecular weight species of DegP were detected and it could be shown that none of these fractions is interconvertible. However, previous studies concerning the oligomeric state of DegP in solution gave contradicting results to the ones presented in this study. It has been proposed that DegP is either a hexamer (Sassoon *et al.*, 1999), a dodecamer (Kim *et al.*, 1999) or that both forms are interconvertible (Kolmar *et al.*, 1996).

Recently, CastilloKeller and Misra (CastilloKeller and Misra, 2003) could demonstrate that $DegP_{S210A}$ can sequester a misfolded mutant OmpC protein. They did not provide any data about the molecular weight of the resulting complex, but in this study it was found out, that binding of OmpC is restricted to these higher molecular weight species. However, it is not clear from the present data whether the herein termed proteins $DegP_{S210A,12\text{-}A}$ and $DegP_{S210A,12\text{-}B}$ are also existing in the absence of unfolded OmpC and have unique structural features that enables them to bind OmpC or if they are formed as a consequence of OmpC binding. The molecular weight of these complexes is still unclear but the results from SEC and DLS suggest that $DegP_{S210A,12\text{-}B}$ is present either as a nonamer or a dodecamer and $DegP_{S210A,12\text{-}A}$ is present as an octadecamer. Moreover, no conclusion can be drawn about the stoichiometry of the DegP – OmpC complex, thus it is not clear whether these species exist as a nonamer, dodecamer or an octadecamer.

It is quite interesting that the results of this survey fit well to the initial characterization of DegP made by Goldberg and co-workers in 1983 (Swamy *et al.*, 1983). This is even more interesting as their investigation was the only study about DegP where the natural enzyme and not the recombinant gene product was analysed. It is therefore quite unlikely that the higher molecular weight species represent an artefact due to protein overexpression. Although crystals from $DegP_{S210A,12\text{-}B}$ could be obtained, they did not diffract in the X-ray beam. Additional experiments have to be undertaken in order to clarify the relevance of this unusual behaviour of DegP.

5.3 DegP, a novel type of cage-forming protease

It has been proposed that the DegP oligomer should have the same shape as the previously described cage-forming proteases (Kolmar *et al.*, 1996; Pallen and Wren, 1997). Crystal structures for HslV (Bochtler *et al.*, 1997), ClpP (Wang *et al.*, 1997), the 20S proteasome (Groll *et al.*, 1997) and of the bleomycin hydrolase (Gal6) (Joshua-Tor *et al.*, 1995) show a common molecular architecture which is also shared by the chaperonins GroEL (Braig *et al.*, 1994) and the thermosome (Ditzel *et al.*, 1998). All these proteins have a cylindrical core and a central pore at the top of the cylinder that provides access to the proteolytic sites which are located within this channel that traverses the whole particle from top to bottom. Electron microscopic (EM) studies of DegP also support an identical arrangement that is not consistent with the herein presented crystal structure (Kim *et al.*, 1999).

This EM-structure postulates a dodecamer, in which a dimer of hexamers displays 62-point symmetry (Kim *et al.*, 1999). Although the protein has twice the molecular mass as the hexameric DegP and the dodecameric HslV, it has the same overall particle dimensions (Bochtler *et al.*, 1997). Moreover, the recently published crystal structures of human HtrA2 (Li *et al.*, 2002), the protease domain of HtrA from *Thermotoga maritima* (Kim *et al.*, 2003) and of DegS from *Escherichia coli* (Wilken *et al.*, 2004) all confirmed a "protease trimer" as the basic building block of all members of the HtrA family. Therefore, it seems unlikely, that the EM-structure does represent a relevant physiological state of DegP.

As seen from the crystal structure, the DegP hexamer is formed by a staggered association of trimeric rings, comparable to the bleomycin hydrolase (Gal6) (Joshua-Tor *et al.*, 1995). However, the fundamental difference to the other cage-forming proteases is that the height of

the inner cavity is determined by three molecular pillars, a unique structural feature that has never been reported before. Moreover, substrates cannot gain access towards the active sites from axial pores of the particle, because the DegP particle is only laterally accessible. Entry of substrates seems to be controlled by twelve PDZ domains which build the mobile sidewalls of the DegP oligomer. It is still not unambiguously clear how the PDZ domains manage this complex task, but the crystal structure provides some hints for this function. The PDZ domains represent the most flexible parts of the structure. Residues of the PDZ domains possess increased thermal motion factors, with an average of 130Å^2 for PDZ1 and of 125Å^2 for PDZ2 respectively compared to 60Å^2 for the protease domain. Second, two conformations of DegP are present in the asymmetric unit of the crystals, i.e. PDZ1 was observed in two strikingly different conformations. The PDZ1 domains differ significantly in their relative orientation and location. A superposition of the molecules revealed a 31.4Å distance for Glu275 between the molecules. Although this difference may be the result of crystal packing constraints, this observation points to the high en-bloc mobility of these protein-protein interaction modules. The scenario is consistent with the previously proposed 'anemone' model for DegP (Pallen and Wren, 1997), where the PDZ domains act as tentacular arms feeding the proteolytic core.

5.4 A janus-headed molecular machine

Refolding and degradation mediated by chaperones and proteases are two antagonistic activities, which establish protein quality control in the cell. During the recent years, it became evident that the regulatory ATPase subunits of the ClpAP, ClpXP and HslUV proteolytic complexes act as chaperones, when these molecules were studied isolated in solution (Wickner *et al.*, 1999). However, the dual activities are provided by two different molecules and it seems likely that these ATPases work as 'unfoldases', thus binding denatured proteins, unfolding them and subsequently delivering them to the proteolytic core of the complex (Wickner *et al.*, 1999).

DegP is one of the rare examples besides the Lon (Wickner *et al.*, 1999) and the FtsH proteases (Schumann, 1999), where both activities are present in a single polypeptide chain. However, the significant difference between DegP and Lon and FtsH is the fact, that DegP is an ATP-independent system and that the switch between digestive and remodelling activity is

controlled by temperature (Spiess *et al.*, 1999). At low temperatures, the protein behaves like a chaperone while at elevated temperatures the proteolytic activity dominates. Beside its protective role against heat-induced folding stress, DegP also participates in the defense against oxidative stress and various other external stress conditions (Skórko-Glonek *et al.*, 1999). Therefore, the question arises how all these different functions are regulated on a molecular level.

The crystal structure of $DegP_{S210A}$ presented in this study corresponds to the chaperone conformation. The fact that the chaperone form does not act as a protease is the consequence of the interplay of loops LA*-L1-L2, resulting in a severe distortion of the active site conformation and blockage of substrate access. The consequence of the distortion is that there is no functional catalytic triad, no oxyanion hole and no S_1 specificity pocket. Displacement of loop LA is proposed to lead to a collapse of the interplay of loops LA*-L1-L2 and subsequently to the proper adjustment of the S_1-binding pocket and the catalytic triad. On loop LA, polar and non-polar residues are asymmetrically arranged. While the former are contacting the interior side of the protein, three phenylalanines (Phe46, Phe49, Phe50) are projecting into the inner cavity, forming the base of the proposed protein-binding platform. As hydrophobic interactions increase with temperature, these phenylalanines may condense at elevated temperatures, thus causing the disruption of the LA*-L1-L2 complex and subsequently activation of the protease. Cage-forming proteases and chaperones can be energy-dependent or energy-independent. In the former group, ATPase activity is important for recognition of target proteins, their dissociation and unfolding, their translocation within the complex and various gating mechanisms. The crystal structure indicates why these functions are not relevant for DegP. DegP preferentially degrades substrates, which are *per se* partially unfolded and which might accumulate under extreme conditions (Strauch *et al.*, 1989). Alternatively, threading of substrates through the inner chamber could promote unfolding into an extended conformation. Removal of higher order structural elements would allow the substrate to reinitiate folding after exit from DegP. In GroEL, the peptide binding sites are arranged in a ring structure allowing simultaneous association with various segments of a polypeptide chain. This multivalent binding is believed to be required for unfolding of protein substrates (Farr *et al.*, 2000). Although an unfoldase activity has not yet been demonstrated, the ring-like architecture of possible peptide binding sites is conserved in DegP. The height of the internal cavity implies that substrates interact with the hydrophobic patches of both trimeric rings. We therefore propose that substrate binding is multivalent,

requiring elements from all subunits of the hexamer. The hydrophobic binding sites of the PDZ domains are properly oriented to augment the number of potential binding patches.

5.5 Possible modes of regulation

The crystal structure of the chaperone state of DegP revealed the structural organization of the protein and showed why DegP is inactive as a protease at room temperature. However, direct structural information about the protease conformation of DegP is still not available, because crystallization at elevated temperatures failed. In addition, preservation of the protease state by a DegP+DFP protease–inhibitor complex failed, the structure rather represents an intermediate between both states. However, the herein presented crystal structures and the recently published structures of other members of the HtrA family allow some suggestions about its regulation.

A structure based sequence alignment with the closest homologues of DegP indicates that residues Ile205 and Ile228 are equivalents of the positions 192 and 216 in trypsin-like serine proteases. These amino acids usually extend into the S_1 pocket, where they partially or fully block access to the base of the pocket. We propose that, in DegP, the two conserved isoleucines play a similar role in the active protease, providing the basis for the interaction with small, hydrophobic P_1 residues of substrates as proposed previously (Kolmar *et al.*, 1996). Consistently, Kolmar and co-workers (Kolmar *et al.*, 1996) reported that DegP has a preference for an aliphatic β-branched residue such as valine at the P_1 position but no obvious preference for the P_1' position. Moreover, the presence of intramolecular disulphide bonds in the denatured form of an oxidized substrate seems to prevent degradation (Kolmar *et al.*, 1996). The height of the inner cavity which harbours the active sites is limited, so that substrates must be at least partially unfolded to reach the active site. Parts of the polypeptide chain containing disulphide bridges are therefore probably too large to gain access to the proteolytic active sites.

Studies which were published throughout the last year pointed to an additional aspect concerning the regulation of DegP. Experiments with DegS (Walsh *et al.*, 2003) and human HtrA2 (Martins *et al.*, 2003) demonstrated that binding of certain peptides to the PDZ domains of these proteins can stimulate their protease activity. In the crystal structure of DegS (Wilken *et al.*, 2004), Clausen and co-workers revealed that binding of an activating peptide

to the PDZ domain of DegS leads to an interaction between the -1 residue of the peptide and the backbone atoms of loop L3. This interaction induces a 15Å movement of loop L3, and a concomitant rearrangement of loops L1, L2 and LD, that ultimately leads to the formation of a functional active site. Moreover, this study determined, that this conformational change after peptide binding is reversible (Wilken *et al.*, 2004). Thus, the question arises whether DegP employs a similar mechanism for its switch from chaperone to protease activity.

A structure based sequence-alignment between DegP and DegS shows a high degree of amino acid conservation throughout the protease domain (Fig. 5.1). Therefore, one may speculate about a similar mode of action. However, so far there is no experimental evidence for this assumption, although it has been reported and furthermore demonstrated in this study, that DegP can be activated by a carboxyl-terminal pilin subunit peptide (Jones *et al.*, 2002). Unlike DegS, peptides derived from the C-terminus of OmpC did not have any effect on the protease activity. Although the similarities between the two systems are quite obvious, with respect to their primary (Fig. 5.1) and tertiary structure (Fig. 5.2), one must also emphasize the pronounced differences.

In contrast to DegS, which has only one PDZ domain and appears as a trimer, DegP has two PDZ domains and it is a hexamer with its active sites buried in an internal cage. Furthermore, access to the active site is controlled by two exceptional loops: Loop LA, which is much longer than in any other member of the HtrA family (Clausen *et al.*, 2002) and loop L1, which obtains a rather unique inactive conformation. Interestingly, an alignment of the third PDZ domain from the synaptic protein Psd-95 in complex with a peptide to the PDZ1 domains of DegP+DFP in the open and in the closed state revealed that only the closed state would allow an interaction between loop L3 and residues of the peptide. In the open state, the peptide binding site of PDZ1 is too far away from the protease domain.

Additionally, $DegP_{S210A}$ was co-crystallized with the activating peptide SPCJ-1 (data not shown). The resulting crystals had a different shape than the $DegP_{S210A}$ crystals but had the same space group and similar unit cell constants. When the structure was determined, hardly any density was visible for the PDZ domains and in the other parts of the protease, no apparent differences could be observed.

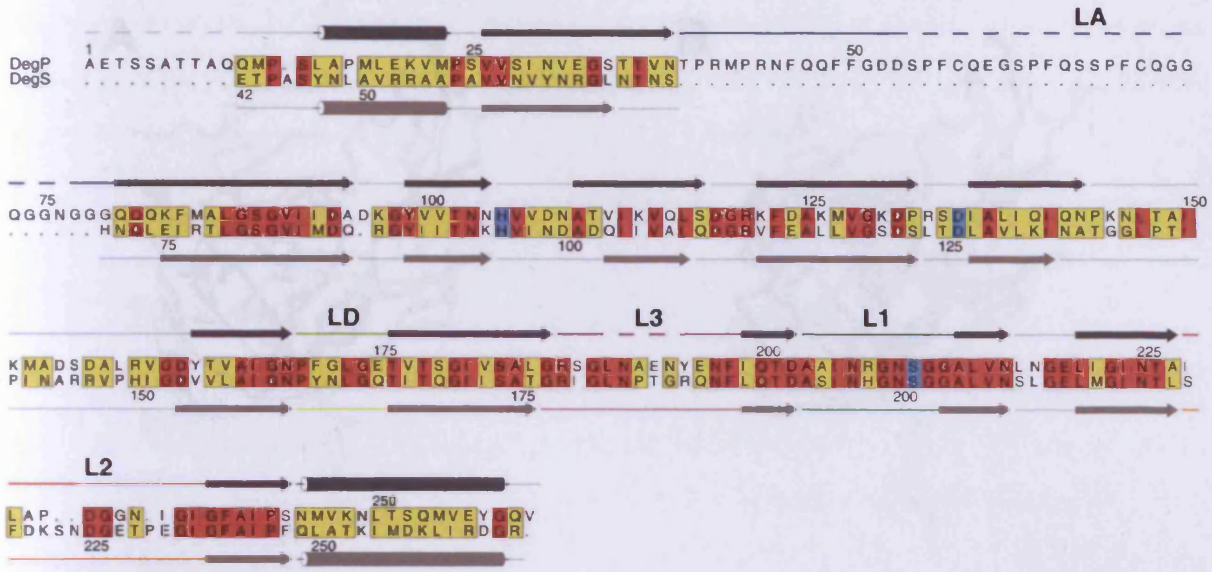

Fig. 5.1. Sequence alignment of *E.coli* DegP and DegS. Identical and homologous residues are shown in red and yellow, respectively. Secondary structure elements, residues of the catalytic triad and mechanistically important loops are indicated.

DegP undergoes autoproteolytic cleavage in solution after Cys69 and Gln82 (Skórko-Glonek *et al.*, 1995). Recently, Skórko-Glonek and coworkers observed that autoproteolysis can be controlled by the redox potential and is restricted to a reducing environment (Skórko-Glonek *et al.*, 2003). It is therefore tempting to speculate that the redox potential might be another determinant, modulating at least the proteolytic activity of DegP. This scenario is reminiscent of Hsp33, a redox-regulated chaperone that appears to protect the cell against the lethal effects of oxidative stress (Jakob *et al.*, 1999). DegP from *E.coli* has two cysteine residues, namely Cys57 and Cys69 which are both located on loop LA and were not visible in any of the herein presented crystal structures. At the moment it is not clear if a disulphide bond between the two residues exists and if the formation or reduction of such a covalent modification corresponds to an additional mechanism for regulating the protease activity of DegP.

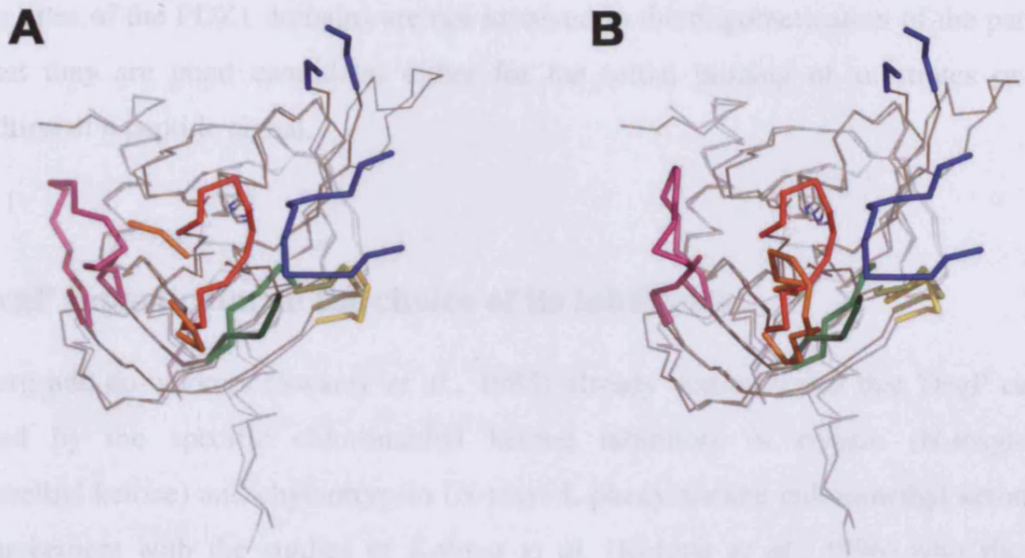

Fig. 5.2. Structural alignment of the $DegP_{S210A}$ with the DegS protease domain. A: DegS (inactive) – $DegP_{S210A}$. **B:** DegS (active) – $DegP_{S210A}$. DegP is coloured in dark grey, DegS in light grey. The colour code for loops L1, L2, L3, LA and LD is the same as in figure 5.1.

5.6 PDZ domains of DegP: molecular gate keepers

PDZ domains are well characterized protein-protein interaction modules, which were first discovered in eukaryotic proteins. PDZ domains play a central role in organizing diverse signalling pathways. They function as scaffolding proteins and therefore increase the efficiency of such pathways (Fanning and Anderson, 1996). Extensive sequence analysis studies indicated that these domains are also present in bacterial proteins (Ponting, 1997). DegP has two PDZ domains and it was discussed whether their main task is substrate recognition or oligomerization (Pallen and Wren, 1997). The deletion of both PDZ domains results in strongly decreased proteolytic activity (Spiess *et al.*, 1999). The deletion also leads to a breakdown of the oligomeric assembly, e.g. the PDZ2 deletion mutant appears to exist as a dimer or trimer in solution (Sassoon *et al.*, 1999). The crystal structure presented in this study presents for the first time the localization of the PDZ domains within the oligomeric complex and provides clues for their function. The PDZ domains build up the flexible side-walls of the particle, therefore controlling lateral access towards the active sites. The flexibility is also illustrated by the fact that the PDZ1 domains of molecule A and B differ in their relative orientation and location. As a consequence, the DegP particle can acquire two distinct states in solution, namely an open and a closed state. Moreover, the canonical peptide

binding sites of the PDZ1 domains are not involved in the oligomerization of the particle, so that that they are good candidates either for the initial binding of substrates or for the recognition of a peptide signal.

5.7 DegP is particular in the choice of its inhibitors

Goldberg and co-workers (Swamy *et al.*, 1983) already demonstrated that DegP cannot be inhibited by the specific chloromethyl ketone inhibitors of trypsin (N-tosyl-L-lysine chloromethyl ketone) and chymotrypsin (N-tosyl-L-phenylalanine chloromethyl ketone). This is in agreement with the studies of Kolmar *et al.* (Kolmar *et al.*, 1996) who showed the requirement for an aliphatic, β-branched residue at the P_1 position, as well as with the crystal structure presented in this study. Despite these findings, it was not possible, to inhibit DegP with the elastase specific inhibitor N-Methoxysuccinyl-ALA-ALA-PRO-VAL chloromethyl ketone, which has a valine residue in the P_1 position. Also, the use of peptide aldehyde inhibitors (MG-115, MG-132, PSI) did not have any effect. Furthermore, inhibitor trials in the presence of the activating peptide SPCJ-1 did not change these results (data not shown).

In this study, the inhibition of DegP by the small, covalent inhibitor DFP was confirmed and DIC was found as an additional inhibitor. Although DIC has several advantages, namely its relatively little toxicity and its enhanced reactivity towards DegP, the copurification as described for DFP was not successful.

It was quite surprising, that the protease activity of DegP was completely abolished, by the hydrophobic probe BisANS, an observation that was previously reported by Kim *et al.* (Kim *et al.*, 1999). At the moment, it can be only speculated, were this molecule may bind to DegP and thus block the proteolytic mechanism. The herein presented crystal structure suggests that BisANS might bind to the hydrophobic patches that line the inner cavity. Through the limited height of the inner cavity and the relatively large van-der-Waals radius of the BisANS molecules, they may somehow obstruct the inner cavity and thereby fix the loop LA. Thus, a displacement of loop LA might become impossible, so that the loop triad remains stable under all conditions. However, additional biochemical experiments have to be carried out in order to locate the binding site(s) of BisANS to verify this hypothesis. Crystallization trials in the presence of BisANS were so far not successful (data not shown).

5.8 The inhibition by DFP is temperature-dependent

The temperature-dependent reaction of the covalent, irreversible inhibitor DFP with DegP corroborated the model derived from the crystal structure. At 4°C, DFP only slightly inhibited DegP, but with increasing temperatures, the rate of inhibition significantly improved. The loops which control access to the active site open the entrance for the substrates (and inhibitors) in correlation with the increasing temperature.

However, it should be mentioned that the reaction of DFP with DegP, or with serine proteases in general might be temperature-dependent and not specifically dependent on the arrangement of DegP's loop triad L1, L2 and LA*. To exclude this possibility, chymotrypsin a structurally similar serine protease could be used as a control. Furthermore, DegP should be incubated for different periods of time at different temperatures to exclude artefacts due to a possible temperature dependence of the inhibitor reaction. Another feature of the inhibitor DFP, namely the limited half-life in aqueous solution should also be noted.

5.9 DFP supports the crystallization of the active protease

The very first DegP/DFP co-purification led to well-diffracting crystals with similar unit cell constants compared to the previously crystallized $DegP_{S210A}$. When the proteolytic activity of this protein was determined in parallel to the crystallization trials it turned out that the protein, although incubated with an excess amount of DFP was still active as a protease. In order to verify the composition of the crystals, they were washed, dissolved and analysed by mass spectrometry. The analysis yielded, that at least 95% of the crystal consisted of DegP and DegP+DFP was only present to a minor extent.

After the analysis of the experiment, it became clear, that one should have considered the short half-life as well as the slow reaction of DFP in the experimental setup. However, this does not explain why the protein could crystallize forming regular, well diffracting crystals, because the autoproteolytic activity which is detrimental for the crystallization of the wild-type DegP (data not shown) may still be present. Anyway, this "accident" yielded for the very first time diffraction quality crystals of wild-type DegP.

The protein crystallized in the same space group and with similar unit cell constants as the $DegP_{S210A}$ mutant protein. At the ESRF (Grenoble, France) the crystals diffracted up to 2.5Å and a complete and highly redundant dataset could be collected. Despite of the good resolution of the structure, no new features could be detected. Especially, the course of loop

LA and L3 was not properly resolved. Moreover, it was not possible to trace any of the PDZ domains of molecule A. The conformation of loop L1, which harbours the active site serine 210 cannot be distinguished from the $DegP_{S210A}$ structure. Thus, the inactive conformation observed in the $DegP_{S210A}$ structure is not a consequence of the serine to alanine mutation but an inherent feature of the chaperone state. As the structure of wt-DegP shows, this inactive conformation is further stabilized by hydrogen bonds between the hydroxyl group of serine 210 and the backbone carbonyl groups of glycine 233 and glycine 234 which are both part of loop L2. This unusual active site geometry is a unique property of *E. coli* DegP which is not apparently shared by other members of the HtrA family. The crystal structures of human HtrA2, HtrA from *Thermotoga maritima* and inactive DegS show proteases in an inactive state because their oxyanion hole is not properly formed. The NH donor of residue 193 (trypsin nomenclature) is flipped to the opposite side of the protein backbone precluding stabilization of the oxyanion intermediate (Fig. 5.3 A). However, in the activated form of DegS, the backbone atoms are rearranged and loop L1 adopts a conformation similar to trypsin (Fig. 5.3 B). In the case of DegP loop L1 has an unusual stretched conformation that has no equivalent in any other structure (Fig. 5.3 C), suggesting again a different mode of activation compared to the other proteins.

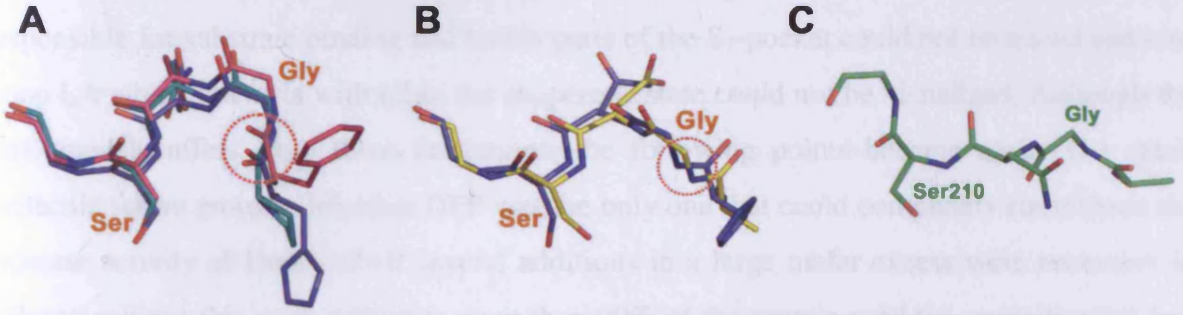

Fig. 5.3. The oxyanion hole of HtrA proteins. A: aligment of human HtrA2 (purple, residues 170-174, PDB entry code 1LCY), HtrA from *Thermotoga maritima* (cyan, residues 203-207, PDB entry code 1L1J) and inactive DegS from *Escherichia coli* (blue, residues 198-202, PDB entry code 1SOT). **B:** alignment of trypsin from *streptomyces griseus* (yellow, residues 192-196, PDB entry code 1SGT) and active DegS from *Escherichia coli* (blue, residues 198-202, PDB entry code 1SOZ). **C:** residues 207-211 of $DegP_{(+DFP)}$ in a similar orientation as the alignments in A and B. The orientation of the NH donor is indicated by a dotted red circle.

For DegS and HtrA2 no chaperone activity has been reported so far but for HtrA from *Thermotoga maritima* in vivo experiments demonstrated that this protein also possesses both functionalities (Kim *et al.*, 2003). However, the crystal structure comprises only the protease domain, the PDZ domains were omitted by genetic engineering. Thus, a direct structural comparison of the two protease-chaperone systems may not be possible due to the pronounced different assemblies.

5.10 DFP could not preserve the protease state of DegP

A crystal structure of an enzyme-inhibitor complex is a general tool to assess the mechanism of a certain enzyme. In the case of DegP, the crystal structures of $DegP_{S210A}$ as well as of wild-type DegP ($DegP_{(+DFP)}$) showed DegP in an inactive protease conformation, which is not surprising, because both proteins were crystallized at room temperature, a condition, where the protease activity of DegP is hardly detectable (Spiess *et al.*, 1999). Crystallization at elevated temperatures failed, therefore a covalent protease-inhibitor complex was considered to be the method of choice, in order to lock the protein in the protease conformation.

Much effort was spent into the determination of the DegP+DFP structure in order to clarify the structural changes which accompany DFP binding. Nevertheless, loop L2 which is responsible for substrate binding and builds parts of the S_1-pocket could not be traced and also Loop LA which interacts with L2 in the chaperone state could not be visualized. Although the final model suffers from these limitations, the following points became clear: The small molecule serine protease inhibitor DFP was the only one that could completely circumvent the protease activity of DegP, albeit several additions in a large molar excess were necessary in order to achieve this goal. Although more than 99% of the protein used for crystallization had incorporated the DFP moiety as judged by mass spectrometry, this inhibitor was incapable to preserve the protease state. The DFP moiety adopts a conformation that has never been observed before in that it points to the opposite of the expected direction into a free space between the loops L1 and LD. Thus, what can be learned from that structure?

Although loop L3 could not be built in molecular detail, its trace did become clear. Unfortunately, no clues about its side chain interactions could be obtained, which are probably a key for the understanding of the protease activation of DegP. The area around loop L3 represents a 'hot spot' of the particle as this is the 'meeting point' of several important

structural features, namely loop L1, L3 and the PDZ1 domain. Especially in the light of the novel DegS structure in its inactive form and with a peptide bound to the PDZ1 domain (see discussion above). However, as the metaphor of the 'hot spot' already implies, these parts of the structure have high thermal motion factors and remain so far ambiguous in their interpretation.

Recapitulating the results concerning the DegP+DFP structure one may say that DFP was unable the lock DegP in the protease conformation. The structure rather represents a dynamic snapshot of a transition state between the protease and the chaperone form as indicated by the increased flexibility of the responsible loops.

5.11 DegP is a processive protease

When the mode of action of multi-subunit protease complexes like the 20S proteasome from *Thermoplasma acidophilum* (Akopian *et al.*, 1997) or the ClpAP protease-chaperone complex from *Escherichia coli* (Thompson *et al.*, 1994) were investigated, it turned out that, unlike classical single-hit proteases (e.g. trypsin), they degrade their protein substrates in a processive fashion and that they produce peptide fragments with a narrow length distribution comprising five to ten residues. Goldberg and co-workers compared the time-dependent substrate degradation with conventional endopeptidases like chymotrypsin and found that for these single-hit proteases, the pattern of peptides in the HPLC chromatogram varied markedly with time. Thus, these enzymes initially produced fragments of their substrates which were degraded further at later times, as expected for non-processive proteases that detach from the polypeptide substrates after each cleavage (Thompson *et al.*, 1994). Processive proteases on the other hand degrade their substrates without the release of significant amounts of higher molecular weight degradation intermediates as indicated by HPLC chromatograms which preserve the same pattern of peptides with time.

The crystal structure of $DegP_{S210A}$ showed that DegP is also a member of the family of cage-forming proteases but with a completely different architecture compared to ClpP or the proteasome. Up to now nothing was known about the potential processivity of DegP and about the length distribution of its degradation products. According to the experiments presented in this study, DegP is a further example for a processive cage-forming protease, though the results for the degradation of resorufin-labelled casein were not completely

unambiguous. Concerning the distribution of the fragment length it is interesting that most products range from 6 to 25 residues and that the distribution is different for heat denatured citrate synthase or the chemically denatured substrates MalS, PhoA and TreA. For citrate synthase the length distribution was found on the low-end of this region comprising 6 to 15 residues, but for the other substrates the result was shifted to the high-end comprising 16 to 25 residues. The significance of this finding is not entirely clear at the moment. The statistical reliability of the citrate synthase assay was greater than for the other substrates (F. Siedler, personal communication), which may be attributed to the fact that the initial substrate concentration was higher in this assay so that the detection of peptides by mass spectrometry was relieved. It is noticeable that in all assays, parts of the substrate amino acid sequence could not be assigned to the fragments detected by mass spectrometry. Though these peptides must be present, this problem may be attributed to the pre-separation of the cleavage products by HPLC. Refinement of the initial peptide separation conditions by HPLC may improve this situation. Furthermore, the result of this assay confirmed that DegP has a preference for small hydrophobic amino acids such as valine, isoleucine, leucine and additionally alanine in the P_1 position und a less pronounced preference for the P_1' position. Additional preferences for the P_2 to P_{10} subsites could not be detected (data not shown). However, when looking at figure 4.15 it is evident that there are numerous additional, putative scissile bonds but none of the corresponding peptides could be identified. At the moment, we do not know if we were just unable to detect them with the current mass spectrometry protocol or if the peptides are really absent because additional substrate preferences prevent cleavage at these sites. Further improvements concerning the experimental setup have to be made to allow statements about subtle details and differences in the results. In addition, we should be able to introduce a robust weighting scheme to be able to exclude peptides with little reliability.

Is it now possible to relate the biochemical results to the crystal structure? The distance between the active sites within one trimer is approximately 24Å and between the respective sites at the opposite site of the cavity the distance is 44Å, 46Å and 51Å. The distance between two adjacent C_α atoms in a peptide is 3.8Å, thus the distances would correspond to peptides with a minimal length of 7, 12, 13 and 14 amino acids. Furthermore, when the third PDZ domain from the synaptic protein Psd-95 (Protein Data Bank entry code: 1BE9) is aligned to the PDZ1 domains of the open and the closed form of $DegP_{S210A}$ and the distance between the active site of DegP and the C_α atom of the C-terminal residue of the peptide bound to the PDZ domain is measured, the corresponding values are 25Å and 33Å. These distances correspond

to polypeptides of 7 and 9 residues in length. Indeed, all these values account for a direct connection. In reality, the distances must be longer, because the active sites within a trimer face into different directions and also when a peptide would bind to a PDZ domain, it has to be bended. Thus, several additional residues have to be added to the calculated lengths for theoretical products (suggesting multiple and parallel binding and cleavage events). While speaking of 'several' residues is very inaccurate, single polypeptide chains are inherently flexible and binding is further determined by the specificities of the active site and the PDZ domains. Although we do not have direct evidence for a correlation between the observed length distribution and the spatial arrangement of putative binding/cleavage sites it is evident that the data obtained from mass spectrometry fit well to the above described restraints that are present within the DegP particle. Moreover, the inability of tripeptidic aldehyde or chloromethylketone inhibitors to inactivate DegP is a further hint for the necessity of a second binding event, away from the active site that may additionally be accompanied with a conformational change opening access towards the active site. Thus one can speculate that once DegP has captured a substrate protein, it is threaded step by step into the internal cavity, and chopped to peptides, without releasing it between the individual cleavage reactions. However, we do not know at the moment, whether the six active sites within the internal cavity act cooperatively. Furthermore, can we really think of the PDZ domains and of the active site arrangement within the central cavity as a molecular ruler as it has been proposed for the proteasome (Wenzel *et al.*, 1994)? When the residue distribution for P_2 to P_{10} subsites was investigated, no clear trend could be observed (data not shown). However, the existence of secondary substrate specificity sites far away from the actual active site cannot be excluded, though peptide analysis may not answer this question because coiling of the polypeptide chain on its way between two binding sites does not allow a direct comparison. Despite of this fact, by extending the analyses to more substrates one may be able to reveal a clustering of certain residues in the P_6 to P_{10} range.

Although the diverse crystal structures of DegP have enlarged our knowledge about the DegP protease-chaperone machine, the number of open questions regarding the regulation of this unique system increased. Therefore, the herein presented results represent the framework for future experiments that will hopefully allow a detailed view into its mode of action.

Chapter 6

References

Abrahams, J. P. and A. G. W. Leslie (1996). "Methods used in the structure determination of bovine mitochondrial F-1 ATPase." Acta Crystallographica Section D-Biological Crystallography **52**: 30-42.

Ades, S. E., L. E. Connolly, *et al.* (1999). "The *Escherichia coli* sigma(E)-dependent extracytoplasmic stress response is controlled by the regulated proteolysis of an anti- sigma factor." Genes & Development **13**(18): 2449-2461.

Akopian, T. N., A. F. Kisselev, *et al.* (1997). "Processive degradation of proteins and other catalytic properties of the proteasome from *Thermoplasma acidophilum*." Journal of Biological Chemistry **272**(3): 1791-1798.

Alba, B. M., H. J. Zhong, *et al.* (2001). "*degS* (*hhoB*) is an essential *Escherichia coli* gene whose indispensable function is to provide sigma(E) activity." Molecular Microbiology **40**(6): 1323-1333.

Andersen, C. L., A. MattheyDupraz, *et al.* (1997). "A new *Escherichia coli* gene, *dsbG*, encodes a periplasmic protein involved in disulphide bond formation, required for recycling DsbA/DsbB and DsbC redox proteins." Molecular Microbiology **26**(1): 121-132.

Anfinsen, C. B. (1973). "Principles That Govern Folding of Protein Chains." Science **181**(4096): 223-230.

Bailey, S. (1994). "The CCP4 Suite - Programs For Protein Crystallography." Acta Crystallographica Section D-Biological Crystallography **50**: 760-763.

Banner, D. W. and P. Hadvary (1991). "Crystallographic Analysis At 3.0-Å Resolution of the Binding to Human Thrombin of 4 Active Site-Directed Inhibitors." Journal of Biological Chemistry **266**(30): 20085-20093.

Bardwell, J. C. A., J. O. Lee, *et al.* (1993). "A Pathway For Disulfide Bond Formation *In vivo*." Proceedings of the National Academy of Sciences of the United States of America **90**(3): 1038-1042.

Barrett, A. J. (1994). Classification of Peptidases. Proteolytic Enzymes: Serine and Cysteine Peptidases. **244:** 1-15.

Barton, G. J. (1993). "Alscript - a Tool to Format Multiple Sequence Alignments." Protein Engineering **6**(1): 37-40.

Bass, S., Q. M. Gu, *et al.* (1996). "Multicopy suppressors of *prc* mutant *Escherichia coli* include two HtrA (DegP) protease homologs (HhoAB), DksA, and a truncated RlpA." Journal of Bacteriology **178**(4): 1154-1161.

Beebe, K. D., J. N. Shin, *et al.* (2000). "Substrate recognition through a PDZ domain in tail-specific protease." Biochemistry **39**(11): 3149-3155.

Ben-Zvi, A. P. and P. Goloubinoff (2001). "Review: Mechanisms of disaggregation and refolding of stable protein aggregates by molecular chaperones." Journal of Structural Biology **135**(2): 84-93.

Bochtler, M., L. Ditzel, *et al.* (1997). "Crystal structure of heat shock locus V (HslV) from *Escherichia coli*." Proceedings of the National Academy of Sciences of the United States of America **94**(12): 6070-6074.

Braig, K., Z. Otwinowski, *et al.* (1994). "The Crystal-Structure of the Bacterial Chaperonin Groel At 2.8- Angstrom." Nature **371**(6498): 578-586.

Brandstetter, H., J. S. Kim, *et al.* (2001). "Crystal structure of the tricorn protease reveals a protein disassembly line." Nature **414**(6862): 466-470.

Brunger, A. T., P. D. Adams, *et al.* (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallographica Section D-Biological Crystallography **54**: 905-921.

Burmeister, W. P. (2000). "Structural changes in a cryo-cooled protein crystal owing to radiation damage." Acta Crystallographica Section D-Biological Crystallography **56**: 328-341.

Cabral, J. H. M., C. Petosa, *et al.* (1996). "Crystal structure of a PDZ domain." Nature **382**(6592): 649-652.

CastilloKeller, M. and R. Misra (2003). "Protease-deficient DegP suppresses lethal effects of a mutant OmpC protein by its capture." Journal of Bacteriology **185**(1): 148-154.

Cavard, D., C. Lazdunski, *et al.* (1989). "The Acylated Precursor Form of the Colicin-a Lysis Protein Is a Natural Substrate of the DegP Protease." Journal of Bacteriology **171**(11): 6316-6322.

Chen, R. and U. Henning (1996). "A periplasmic protein (Skp) of *Escherichia coli* selectively binds a class of outer membrane proteins." Molecular Microbiology **19**(6): 1287-1294.

Clausen, T., C. Southan, *et al.* (2002). "The HtrA family of proteases: Implications for protein composition and cell fate." Molecular Cell **10**(3): 443-455.

Czapinska, H. and J. Otlewski (1999). "Structural and energetic determinants of the S-1-site specificity in serine proteases." European Journal of Biochemistry **260**(3): 571-595.

Danese, P. N. and T. J. Silhavy (1998). "CpxP, a stress-combative member of the Cpx regulon." Journal of Bacteriology **180**(4): 831-839.

Danese, P. N., W. B. Snyder, *et al.* (1995). "The Cpx 2-Component Signal-Transduction Pathway of *Escherichia Coli* Regulates Transcription of the Gene Specifying the Stress-Inducible Periplasmic Protease, DegP." Genes & Development **9**(4): 387-398.

Dartigalongue, C. and S. Raina (1998). "A new heat-shock gene, *ppiD*, encodes a peptidyl-prolyl isomerase required for folding of outer membrane proteins in *Escherichia coli*." Embo Journal **17**(14): 3968-3980.

Dauter, Z., M. Dauter, *et al.* (2002). "Jolly SAD." Acta Crystallographica Section D-Biological Crystallography **58**: 494-506.

delaFortelle, E. and G. Bricogne (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Macromolecular Crystallography, Pt a. **276:** 472-494.

DeLano, W. L. (2004). Pymol. San Carlos, California, U.S.A, DeLano Scientific LLC.

Ditzel, L., J. Lowe, *et al.* (1998). "Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT." Cell **93**(1): 125-138.

Doyle, D. A., A. Lee, *et al.* (1996). "Crystal structures of a complexed and peptide-free membrane protein-binding domain: Molecular basis of peptide recognition by PDZ." Cell **85**(7): 1067-1076.

Ellis, R. J. and F. U. Hartl (1996). "Protein folding in the cell: Competing models of chaperonin function." Faseb Journal **10**(1): 20-26.

Engh, R. A. and R. Huber (1991). "Accurate Bond and Angle Parameters For X-Ray Protein-Structure Refinement." Acta Crystallographica Section a **47**: 392-400.

Evans, G. and R. F. Pettifer (2001). "CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra." Journal of Applied Crystallography **34**: 82-86.

Evans, S. V. (1993). "Setor - Hardware-Lighted 3-Dimensional Solid Model Representations of Macromolecules." Journal of Molecular Graphics **11**(2): 134-&.

Faccio, L., C. Fusco, *et al.* (2000). "Characterization of a novel human serine protease that has extensive homology to bacterial heat shock endoprotease HtrA and is regulated by kidney ischemia." Journal of Biological Chemistry **275**(4): 2581-2588.

Fanning, A. S. and J. M. Anderson (1996). "Protein-protein interactions: PDZ domain networks." Current Biology **6**(11): 1385-1388.

Farr, G. W., K. Furtak, *et al.* (2000). "Multivalent binding of nonnative substrate proteins by the chaperonin GroEL." Cell **100**(5): 561-573.

Ferré-d'Amaré, A. R. and S. K. Burley (1994). "Use of Dynamic Light-Scattering to Assess Crystallizability of Macromolecules and Macromolecular Assemblies." Structure **2**(5): 357-359.

Fersht, A. (1999). Structure and Mechanism in Protein Science. New York, W.H. Freeman and company.

Gill, S. C. and P. H. von Hippel (1989). "Calculation of Protein Extinction Coefficients From Amino-Acid Sequence Data." Analytical Biochemistry **182**(2): 319-326.

Groll, M., L. Ditzel, *et al.* (1997). "Structure of 20S proteasome from yeast at 2.4 angstrom resolution." Nature **386**(6624): 463-471.

Harel, M., C. T. Su, *et al.* (1991). "Refined Crystal-Structures of Aged and Non-Aged Organophosphoryl Conjugates of Gamma-Chymotrypsin." Journal of Molecular Biology **221**(3): 909-918.

Harrison, S. C. (1996). "Peptide-surface association: The case of PDZ and PTB domains." Cell **86**(3): 341-343.

Hillier, B. J., K. S. Christopherson, *et al.* (1999). "Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex." Science **284**(5415): 812-815.

Holmgren, A., M. J. Kuehn, *et al.* (1992). "Conserved Immunoglobulin-Like Features in a Family of Periplasmic Pilus Chaperones in Bacteria." Embo Journal **11**(4): 1617-1622.

Hu, S. I., M. Carozza, *et al.* (1998). "Human HtrA, an evolutionarily conserved serine protease identified as a differentially expressed gene product in osteoarthritic cartilage." Journal of Biological Chemistry **273**(51): 34406-34412.

Hung, A. Y. and M. Sheng (2002). "PDZ domains: Structural modules for protein complex assembly." Journal of Biological Chemistry **277**(8): 5699-5702.

Ishikawa, T., F. Beuron, *et al.* (2001). "Translocation pathway of protein substrates in ClpAP protease." Proceedings of the National Academy of Sciences of the United States of America **98**(8): 4328-4333.

Jakob, U., W. Muse, *et al.* (1999). "Chaperone activity with a redox switch." Cell **96**(3): 341-352.

Jones, C., B. Mulloy, *et al.*, Eds. (1996). Crystallographic Methods and Protocols. Methods in Molecular Biology. Totowa, New Jersey, Humana Press.

Jones, C. H., P. Dexter, *et al.* (2002). "*Escherichia coli* DegP protease cleaves between paired hydrophobic residues in a natural substrate: the PapA pilin." Journal of Bacteriology **184**(20): 5762-5771.

Jones, T. A., J. Y. Zou, *et al.* (1991). "Improved Methods For Building Protein Models in Electron- Density Maps and the Location of Errors in These Models." Acta Crystallographica Section a **47**: 110-119.

Joshua-Tor, L., H. E. Xu, *et al.* (1995). "Crystal-Structure of a Conserved Protease That Binds DNA - the Bleomycin Hydrolase, Gal6." Science **269**(5226): 945-950.

Karthikeyan, S., T. Leung, *et al.* (2001). "Crystal structure of the PDZ1 domain of human Na+/H+ exchanger regulatory factor provides insights into the mechanism of carboxyl-terminal leucine recognition by class I PDZ domains." Journal of Molecular Biology **308**(5): 963-973.

Kim, D. Y., D. R. Kim, *et al.* (2003). "Crystal structure of the protease domain of a heat-shock protein HtrA from *Thermotoga maritima*." Journal of Biological Chemistry **278**(8): 6543-6551.

Kim, K. I., S. C. Park, *et al.* (1999). "Selective degradation of unfolded proteins by the self-compartmentalizing HtrA protease, a periplasmic heat shock protein in *Escherichia coli*." Journal of Molecular Biology **294**(5): 1363-1374.

Kleywegt, G. J. (1996). "Use of non-crystallographic symmetry in protein structure refinement." Acta Crystallographica Section D-Biological Crystallography **52**: 842-857.

Knight, S. D. (2000). "RSPS version 4.0: a semi-interactive vector-search program for solving heavy-atom derivatives." Acta Crystallographica Section D-Biological Crystallography **56**: 42-47.

Kolmar, H., P. R. H. Waller, *et al.* (1996). "The DegP and DegQ periplasmic endoproteases of *Escherichia coli*: Specificity for cleavage sites and substrate conformation." Journal of Bacteriology **178**(20): 5925-5929.

Kraulis, P. J. (1991). "Molscript - a Program to Produce Both Detailed and Schematic Plots of Protein Structures." Journal of Applied Crystallography **24**: 946-950.

Laemmli, U. K. (1970). "Cleavage of Structural Proteins During Assembly of Head of Bacteriophage-T4." Nature **227**(5259): 680-&.

Laskowski, R. A., M. W. Macarthur, *et al.* (1993). "Procheck - a Program to Check the Stereochemical Quality of Protein Structures." Journal of Applied Crystallography **26**: 283-291.

Lee, G. J., A. M. Roseman, *et al.* (1997). "A small heat shock protein stably binds heat-denatured model substrates and can maintain a substrate in a folding-competent state." Embo Journal **16**(3): 659-671.

Lesk, A. M. (2001). Introduction to Protein Architecture. Oxford, New York, Oxford University Press.

Leslie, A. G. W. (1992). "Recent changes to the MOSFLM package for processing film and image plate data." Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography **26**.

Li, W. Y., S. M. Srinivasula, *et al.* (2002). "Structural insights into the pro-apoptotic function of mitochondrial serine protease HtrA2/Omi." Nature Structural Biology **9**(6): 436-441.

Lindquist, S. and E. A. Craig (1988). "The Heat-Shock Proteins." Annual Review of Genetics **22**: 631-677.

Lipinska, B., O. Fayet, *et al.* (1989). "Identification, Characterization, and Mapping of the *Escherichia-Coli* Htra Gene, Whose Product Is Essential For Bacterial-Growth Only At Elevated-Temperatures." Journal of Bacteriology **171**(3): 1574-1584.

Lipinska, B., S. Sharma, *et al.* (1988). "Sequence-Analysis and Regulation of the *htra* Gene of *Escherichia-Coli* - a Sigma-32-Independent Mechanism of Heat- Inducible Transcription." Nucleic Acids Research **16**(21): 10053-10067.

Liu, J. and C. T. Walsh (1990). "Peptidyl Prolyl Cis-Trans-Isomerase From *Escherichia-Coli* - a Periplasmic Homolog of Cyclophilin That Is Not Inhibited By Cyclosporine-a." Proceedings of the National Academy of Sciences of the United States of America **87**(11): 4028-4032.

Lupas, A., J. M. Flanagan, *et al.* (1997). "Self-compartmentalizing proteases." Trends in Biochemical Sciences **22**(10): 399-404.

Martins, L. M., B. E. Turk, *et al.* (2003). "Binding specificity and regulation of the serine protease and PDZ domains of HtrA2/Omi." Journal of Biological Chemistry **278**(49): 49417-49427.

Matsuyama, S., T. Tajima, *et al.* (1995). "A Novel Periplasmic Carrier Protein Involved in the Sorting and Transport of *Escherichia-Coli* Lipoproteins Destined For the Outer-Membrane." Embo Journal **14**(14): 3365-3372.

Matsuyama, S., N. Yokota, *et al.* (1997). "A novel outer membrane lipoprotein, LolB (HemM), involved in the LolA (p20)-dependent localization of lipoproteins to the outer membrane of *Escherichia coli*." Embo Journal **16**(23): 6947-6955.

Matthews, B. W. (2001). Anomalous Dispersion. Volume F: Crystallography of Biological Macromolecules. A. E. Rossmann W.G., Kluwer Academic Publishers: 293-298.

Merritt, E. A. and D. J. Bacon (1997). Raster3D: Photorealistic molecular graphics. Macromolecular Crystallography, Pt B. **277:** 505-524.

Messerschmidt, A. and R. Huber (2000). X-ray Crystallography of Biological Macromolecules. Encyclopedia of Analytical Chemistry. M. R.A. Chichester, John Wiley & Sons Ltd.: 6061-6107.

Missiakas, D., M. P. Mayer, *et al.* (1997). "Modulation of the *Escherichia coli* sigma(E) (RpoE) heat-shock transcription-factor activity by the RseA, RseB and RseC proteins." Molecular Microbiology **24**(2): 355-371.

Missiakas, D. and S. Raina (1998). "The extracytoplasmic function sigma factors: role and regulation." Molecular Microbiology **28**(6): 1059-1066.

Moradian-Oldak, J., W. Leung, *et al.* (1998). "Temperature and pH-dependent supramolecular self-assembly of amelogenin molecules: A dynamic light-scattering analysis." Journal of Structural Biology **122**(3): 320-327.

Morimoto, R., A. Tissieres, *et al.* (1994). Progress and Perspectives on the Biology of Heat Shock Proteins and Molecular Chaperones. The Biology of Heat Shock Proteins and Molecular Chaperones, Cold Spring Harbor Laboratory Press.

Nicholls, A., R. Bharadwaj, *et al.* (1993). "Grasp - Graphical Representation and Analysis of Surface- Properties." Biophysical Journal **64**(2): A166-A166.

Oliver, D. (1996). Periplasm. *Escherichia* and *Salmonella* Cellular and Molecular Biology. F. C. Neidhardt. Washington, D.C., ASM Press.

Oschkinat, H. (1999). "A new type of PDZ domain recognition." Nature Structural Biology **6**(5): 408-410.

Otwinowski, Z. and W. Minor (1997). Processing of X-ray diffraction data collected in oscillation mode. Macromolecular Crystallography, Pt a. **276:** 307-326.

Pallen, M. J. and B. W. Wren (1997). "The HtrA family of serine proteases." Molecular Microbiology **26**(2): 209-221.

Philippsen, A. (2002). DINO: Visualizing Structural Biology.

Pogliano, K. J. and J. Beckwith (1994). "Genetic and Molecular Characterization of the *Escherichia-Coli* SecD Operon and Its Products." Journal of Bacteriology **176**(3): 804-814.

Ponting, C. P. (1997). "Evidence for PDZ domains in bacteria, yeast, and plants." Protein Science **6**(2): 464-468.

Ponting, C. P. and M. J. Pallen (1999). "beta-Propeller repeats and a PDZ domain in the tricorn protease: predicted self-compartmentalisation and C-terminal polypeptide-binding strategies of substrate selection." Fems Microbiology Letters **179**(2): 447-451.

Raivio, T. L. and T. J. Silhavy (1999). "The sigma(E) and Cpx regulatory pathways: overlapping but distinct envelope stress responses." Current Opinion in Microbiology **2**(2): 159-165.

Raivio, T. L. and T. J. Silhavy (2001). "Periplasmic stress and ECF sigma factors." Annual Review of Microbiology **55**: 591-624.

Ramachandran, R., C. Hartmann, *et al.* (2002). "Functional interactions of HslV (ClpQ) with the ATPase HslU (ClpY)." Proceedings of the National Academy of Sciences of the United States of America **99**(11): 7396-7401.

Ramm, K. and A. Pluckthun (2000). "The periplasmic *Escherichia coli* peptidylprolyl cis,trans- isomerase FkpA - II. Isomerase-independent chaperone activity in vitro." Journal of Biological Chemistry **275**(22): 17106-17113.

Rawlings, N. D. and A. J. Barrett (1994). Families of Serine Peptidases. Proteolytic Enzymes: Serine and Cysteine Peptidases. **244:** 19-61.

Remaut, H., C. Bompard-Gilles, *et al.* (2001). "Structure of the *Bacillus subtilis* D-aminopeptidase DppA reveals a novel self-compartmentalizing protease." Nature Structural Biology **8**(8): 674-678.

Rhodes, G. (1999). Crystallography made Crystal Clear. San Diego, Academic Press.

Rietsch, A. and J. Beckwith (1998). "The genetics of disulfide bond metabolism." Annual Review of Genetics **32**: 163-184.

Rosen, B. P. (1987). Washington D.C., American Society for Microbiology.

Rouviere, P. E. and C. A. Gross (1996). "SurA, a periplasmic protein with peptidyl-prolyl isomerase activity, participates in the assembly of outer membrane porins." Genes & Development **10**(24): 3170-3182.

Saibil, H. and S. Wood (1993). "Chaperonins." Current Opinion in Structural Biology **3**(2): 207-213.

Sambongi, Y. and S. J. Ferguson (1994). "Specific Thiol Compounds Complement Deficiency in C-Type Cytochrome Biogenesis in *Escherichia-Coli* Carrying a Mutation in a Membrane-Bound Disulfide Isomerase-Like Protein." Febs Letters **353**(3): 235-238.

Sanner, M. F., A. J. Olson, *et al.* (1996). "Reduced surface: An efficient way to compute molecular surfaces." Biopolymers **38**(3): 305-320.

Sassoon, N., J. P. Arie, *et al.* (1999). "PDZ domains determine the native oligomeric structure of the DegP (HtrA) protease." Molecular Microbiology **33**(3): 583-589.

Schneider, C. and F. U. Hartl (1996). "Cell biology - Hats off to the tricorn protease." Science **274**(5291): 1323-1324.

Schneider, T. R. and G. M. Sheldrick (2002). "Substructure solution with SHELXD." Acta Crystallographica Section D-Biological Crystallography **58**: 1772-1779.

Schumann, W. (1999). "FtsH - a single-chain charonin?" Fems Microbiology Reviews **23**(1): 1-11.

Shtilerman, M., G. H. Lorimer, *et al.* (1999). "Chaperonin function: Folding by forced unfolding." Science **284**(5415): 822-825.

Skórko-Glonek, J., K. Krzewski, *et al.* (1995). "Comparison of the Structure of Wild-Type HtrA Heat-Shock Protease and Mutant HtrA Proteins - a Fourier-Transform Infrared Spectroscopic Study." Journal of Biological Chemistry **270**(19): 11140-11146.

Skórko-Glonek, J., B. Lipinska, *et al.* (1997). "HtrA heat shock protease interacts with phospholipid membranes and undergoes conformational changes." Journal of Biological Chemistry **272**(14): 8974-8982.

Skórko-Glonek, J., A. Wawrzynow, *et al.* (1995). "Site-Directed Mutagenesis of the Htra(Degp) Serine-Protease, Whose Proteolytic Activity Is Indispensable For *Escherichia-Coli* Survival At Elevated-Temperatures." Gene **163**(1): 47-52.

Skórko-Glonek, J., D. Zurawa, *et al.* (1999). "The *Escherichia coli* heat shock protease HtrA participates in defense against oxidative stress." Molecular and General Genetics **262**(2): 342-350.

Skórko-Glonek, J., D. Zurawa, *et al.* (2003). "The N-terminal region of HtrA heat shock protease from *Escherichia coli* is essential for stabilization of HtrA primary structure and maintaining of its oligomeric structure." Biochimica Et Biophysica Acta-Proteins and Proteomics **1649**(2): 171-182.

Songyang, Z., A. S. Fanning, *et al.* (1997). "Recognition of unique carboxyl-terminal motifs by distinct PDZ domains." Science **275**(5296): 73-77.

Spiess, C., A. Beil, *et al.* (1999). "A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein." Cell **97**(3): 339-347.

Stoscheck, C. M. (1990). "Quantitation of Protein." Methods in Enzymology **182**: 50-68.

Strauch, K. L., K. Johnson, *et al.* (1989). "Characterization of *degp*, a Gene Required For Proteolysis in the Cell-Envelope and Essential For Growth of *Escherichia-Coli* At High-Temperature." Journal of Bacteriology **171**(5): 2689-2696.

Suzuki, Y., Y. Imai, *et al.* (2001). "A serine protease, HtrA2, is released from the mitochondria and interacts with XIAP, inducing cell death." Molecular Cell **8**(3): 613-621.

Swamy, K. H. S., H. C. Chin, *et al.* (1983). "Isolation and Characterization of Protease Do From *Escherichia- Coli*, a Large Serine Protease Containing Multiple Subunits." Archives of Biochemistry and Biophysics **224**(2): 543-554.

Taylor, G. (2003). "The phase problem." Acta Crystallographica Section D-Biological Crystallography **59**: 1881-1890.

Terwilliger, T. C. and J. Berendzen (1999). "Automated MAD and MIR structure solution." Acta Crystallographica Section D-Biological Crystallography **55**: 849-861.

Thompson, M. W., S. K. Singh, *et al.* (1994). "Processive Degradation of Proteins By the ATP-Dependent Clp Protease From *Escherichia-Coli* - Requirement For the Multiple Array of Active-Sites in ClpP But Not ATP Hydrolysis." Journal of Biological Chemistry **269**(27): 18209-18215.

Tilly, K., H. Murialdo, *et al.* (1981). "Identification of a 2nd *Escherichia-Coli groE* Gene Whose Product Is Necessary For Bacteriophage Morphogenesis." Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences **78**(3): 1629-1633.

Todd, M. J., G. H. Lorimer, *et al.* (1996). "Chaperonin-facilitated protein folding: Optimization of rate and yield by an iterative annealing mechanism." Proceedings of the National Academy of Sciences of the United States of America **93**(9): 4030-4035.

Waller, P. R. H. and R. T. Sauer (1996). "Characterization of *degQ* and *degS*, *Escherichia coli* genes encoding homologs of the DegP protease." Journal of Bacteriology **178**(4): 1146-1153.

Walsh, N. P., B. M. Alba, *et al.* (2003). "OMP peptide signals initiate the envelope-stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain." Cell **113**(1): 61-71.

Wang, J. M., J. A. Hartling, *et al.* (1997). "The structure of ClpP at 2.3 angstrom resolution suggests a model for ATP-dependent proteolysis." Cell **91**(4): 447-456.

Weeks, C. M. and R. Miller (1999). "The design and implementation of SnB version 2.0." Journal of Applied Crystallography **32**: 120-124.

Wenzel, T., C. Eckerskorn, *et al.* (1994). "Existence of a Molecular Ruler in Proteasomes Suggested By Analysis of Degradation Products." Febs Letters **349**(2): 205-209.

Wickner, S., M. R. Maurizi, *et al.* (1999). "Posttranslational quality control: Folding, refolding, and degrading proteins." Science **286**(5446): 1888-1893.

Wilken, C., K. Kitzing, *et al.* (2004). "Crystal structure of the DegS stress sensor: How a PDZ domain recognizes misfolded protein and activates a protease." Cell **117**(4): 483-494.

Wootton, J. C. and M. H. Drummond (1989). "The Q-Linker - a Class of Interdomain Sequences Found in Bacterial Multidomain Regulatory Proteins." Protein Engineering **2**(7): 535-543.

Yao, T. T. and R. E. Cohen (1999). "Giant proteases: Beyond the proteasome." Current Biology **9**(15): R551-R553.

Zumbrunn, J. and B. Trueb (1996). "Primary structure of a putative serine protease specific for IGF-binding proteins." Febs Letters **398**(2-3): 187-192.