Autoproteolysis accelerated by conformational strain - a novel biochemical mechanism.

Denny Johansson

UNIVERSITY OF GOTHENBURG

Institute of Biomedicine
Department of Medical Biochemistry and Cell Biology

2008

ISBN 978-91-628-7545-9
© Denny Johansson, August 2008
Institute of Biomedicine
Department of Medical Biochemistry and Cell Biology
Sahlgrenska Academy at University of Gothenburg

Printed by Chalmers reproservice, Chalmers tekniska högskola
Gothenburg, Sweden

Cover illustration: The MUC1 SEA domain

Denny Johansson (2008): **Autoproteolysis accelerated by conformational strain - a novel biochemical mechanism.** Institute of Biomedicine, Department of Medical Biochemistry and Cell Biology, Sahlgrenska Academy at University of Gothenburg, Sweden

Abstract

Natural fragmentation of polypeptide chains by autoproteolysis occurs in a number of protein families. It is a vital step in the maturation of several enzymes and in the formation of membrane-associated mucins that constitute a part of the protective mucus barrier lining epithelial cells. These reactions follow similar routes involving an initial N→O or N→S acyl shift starting with a nucleophilic attack by a hydroxyl or thiol group on a carbonyl carbon followed by resolution of the ester intermediate. Previous studies indicate that distortion of the scissile peptide bond may play a role in autoproteolysis. Our structural, biochemical and molecular dynamics studies of the autoproteolyzed SEA domains from human membrane-bound mucin MUC1 and human receptor GPR116 confirmed this by revealing a novel biochemical mechanism where the folding free energy accelerates cleavage by imposing conformational strain in the precursor structure. This mechanism may well be general for autoproteolysis.

The structure of the cleaved MUC1 SEA domain was determined using NMR spectroscopy. It consists of four alpha-helices packed against the concave surface of a four-stranded anti-parallel beta-sheet. There are no disordered loops. The site of autoproteolysis is a conserved GSVVV sequence located at the ends of beta-sheets 2 and 3 where the resulting N- and C-terminal residues become integrated parts of these sheets after cleavage. The structure does not reveal any charge-relay system or oxyanion hole as would be expected if catalysis proceeded by way of transition state stabilization. The surface of the domain contains two hydrophobic patches that may serve as sites of interaction with other proteins, giving it a potential function in the regulation of the protective mucus layer.

Combined studies of autoproteolysis and adoption of native fold show that these mechanisms proceed with the same rate and that the autoproteolysis has a global effect on structure. Studies of the stability and cleavage kinetics were performed by destabilizing core mutations or addition of denaturing co-solvents. Analysis revealed that ~7 kcal mol^{-1} of conformational free energy is partitioned as strain in the precursor. The results corroborate a mechanism where the autoproteolysis is accelerated by the concerted action of a conserved serine residue and strain imposed on the precursor structure upon folding, that is, the catalytic mechanism is substrate destabilization.

The autoproteolysis of SEA is pH dependent. This is in line with a proposed mechanism with an initial N→O acyl shift, involving transient protonation of the amide nitrogen, and subsequent hydroxyl-mediated hydrolysis of the resulting ester. The mechanistic link between strain and cleavage kinetics is that strain induces a pyramidal conformation of the amide nitrogen which results in an increase of the pK_a and thereby an acceleration of the N→O acyl shift. Furthermore we propose a water hydronium as proton donor in this step. This explains the absence of conserved acid-base functionality within the SEA structure.

Keywords: SEA domain, autoproteolysis, MUC1, conformational strain, protein folding.

© Denny Johansson 2008

List of publications

This thesis is based upon the following four articles, which are referred to in the text by their Roman numerals (I-IV). The articles are enclosed at the end of the thesis.

I. Macao, B., **Johansson, D. G. A**., Hansson, G. C., Härd, T. (2006) Autoproteolysis coupled to protein folding in the SEA domain of the membrane-bound MUC1 mucin. *Nature Structural and Molecular Biology* **13**: 71-76.

II. **Johansson, D. G. A.**, Macao, B., Sandberg, A., Härd, T. (2008) SEA domain autoproteolysis accelerated by conformational strain: mechanistic aspects. *Journal of Molecular Biology* **377**: 1130-1143.

III. Sandberg, A., **Johansson, D. G. A.**, Macao, B., Härd, T. (2008) SEA domain autoproteolysis accelerated by conformational strain: energetic aspects. *Journal of Molecular Biology* **377**: 1117-1129.

IV. **Johansson, D. G. A.**, Sandberg, A., Härd, T. (2008) SEA domain autoproteolysis: paths, intermediates and linkage between conformational strain and reaction rate. *Manuscript.*

All publications are reproduced with permission from the copyright holders.

Contents

Abbreviations

CD	circular dichroism
GdmCl	guanidinium chloride
HSQC	heteronuclear single quantum coherence
IMAC	immobilized metal affinity chromatography
NMR	nuclear magnetic resonance
Ntn	N-terminal nucleophile
PAGE	polyacrylamide gel electrophoresis
SDS	sodium dodecyl sulfate
SEC	size exclusion chromatography
TS	transition state
wt	wild type

1 Introduction

This thesis describes a biochemical mechanism where autoproteolysis of a protein domain is accelerated by conformational strain imposed on the precursor structure, leading to a distortion of the scissile peptide bond. The distortion is mechanistically linked to the cleavage as it leads to amide nitrogen pyramidalization and thereby an increase in the pK_a of the peptide nitrogen.

Data obtained from studies of the SEA domain of human mucin protein MUC1 and the human receptor protein GPR116 are summarized here and presented in detail in the four enclosed articles, referred to by their Roman numerals (I-IV). The results are also related to other systems and it is suggested that acceleration through conformational strain is a general feature in autoproteolytic proteins.

1.1 Protein autoproteolysis and the N→O acyl shift

Autoproteolysis is here defined as the disruption of a peptide bond in a protein domain by an intramolecular reaction, as opposed to enzymatic proteolysis defined as a mechanism in which an enzyme is involved in an intermolecular reaction. Autoproteolysis occurs in a number of proteins, including the maturation of enzymes and the process of protein splicing (Blair and Semler 1991; Perler, Xu *et al*. 1997; Paulus 2000).

The autoproteolysis of proteins is generally believed to be initiated by an N→O acyl shift (Perler 1998). In this process the scissile peptide bond is converted to an ester via a nucleophilic attack by an adjacent hydroxyl or thiol on the carbonyl carbon of the protein backbone. The resulting ester is resolved by a consecutive reaction that differs between autoproteolytic systems. This indicates that the N→O acyl shift is a preserved mechanism while the resolution of the ester has evolved to meet the needs of nature (Perler 1998; Paulus 2000) (figure 1).

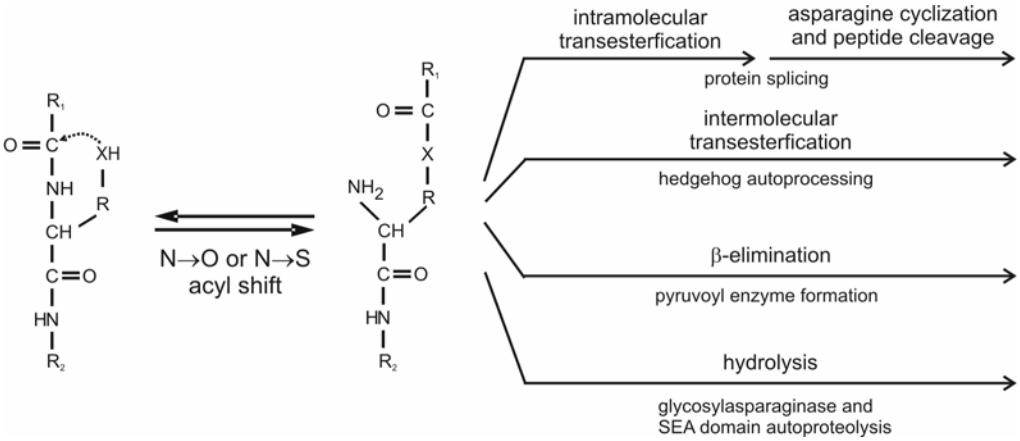

Figure 1. Examples of autoproteolytical reactions to illustrate the diversity of systems all initiated by an N→O acyl shift. A nucleophilic attack of an adjacent amino acid containing a hydroxyl or thiol group (denoted XH) on the carbonyl carbon leads to formation of an ester. The resulting ester is resolved by consecutive reactions that differ between the systems.

Mutational studies of the systems outlined in figure 1 have shown that a serine, threoinine or cysteine residue adjacent to the scissile peptide bond is necessary for the reaction to proceed (Perler, Xu *et al*. 1997; Paulus 2000).

The possibility to induce an N→O acyl shift at peptide bonds adjacent to amino acids containing a hydroxyl or thiol has been utilized to obtain specific cleavage of peptides since the resulting ester is more prone to hydrolysis than the amide bond. However, at physiological conditions the equilibrium of the N→O acyl shift is in favor of the amide (Iwai and Ando 1967) and spontaneous cleavage at peptide bonds preceding a serine, threonine or cysteine residue is too slow to be of major biological importance. Hence, the amide-ester equilibrium must be shifted towards the ester and the N→O acyl shift must be accelerated in some way in autoproteolytical systems. One possible explanation for such acceleration is a destabilization of the precursor involving distortion of the scissile peptide bond. Conformational strain as a way of accelerating reactions has been suggested previously (Holley 1953) and extensive work has been performed on the increased reactivity of distorted amides within the field of organic chemistry (Blackburn, Skaife *et al*. 1980; Somayaji and Brown 1986; Lopez, Mujika *et al*. 2003; Tani and Stoltz 2006). Also, possible involvement of strained precursor conformations has been suggested for some of the autoproteolytical systems outlined in figure 1, but the

biochemical and energetic aspects of a strained precursor has never been completely sorted out in these systems. Here the details of the SEA domain autoproteolysis are elucidated and the involvement of a strained precursor becomes evident. The energetic aspects of the mechanism are also discussed based on transition state (TS) theory. At the end of this thesis I will return to the autoproteolytical systems outlined in figure 1 and discuss the generality of acceleration by strain imposed on the scissile peptide bond of the precursor.

1.2 The peptide bond

Spontaneous hydrolysis of peptide bonds is an extremely slow reaction that proceeds with a half-life of ~2100 years at 21 °C and pH 6.8 (Radzicka and Wolfenden 1996). In order to understand the connection between conformational strain and increased reactivity one needs to consider the nature of the peptide bond. Peptide bonds linking the amino acids in proteins are planar due to resonance stabilization leading to a double bond character between the nitrogen and the carbonyl carbon (Pauling, Corey *et al*. 1951) (figure 2.).

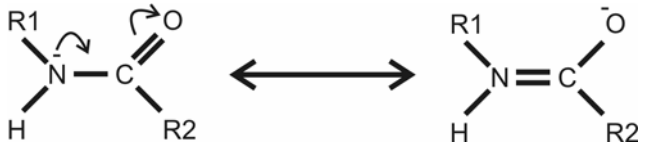

Figure 2. The major resonance forms of the peptide bond.

The exact nature of the peptide bond and the reason for it being stable has been a matter of debate but it is generally agreed that distortion of the peptide bond leads to increased reactivity (Pauling, Corey *et al*. 1951; Somayaji and Brown 1986; Bennet, Wang *et al*. 1990; Milner-White 1997). This is because distortion leads to a loss in resonance stabilization and a pyramidalization of the amide nitrogen. Furthermore, experimental data provide a link between amide nitrogen pK_a and peptide geometry. Twisted amides have higher gas phase pK_a values than planar amides (Greenberg, Moore *et al.* 1996). Also, the solution pK_a for a distorted amide has been estimated to between 3.5 and 3.8 and it decreases towards the pKa of a planar amide as the distortion decreases

(Wang, Bennet *et al*, 1991). However, distortion is energetically unfavored (Fersht 1999). In an autoproteolytic mechanism involving strain imposed on the scissile bond, the energetic cost of a distorted peptide bond must be compensated in some way. In the cleavage of the SEA domain we propose that the folding free energy accounts for this compensation. Hence, part of the folding free energy is partitioned as strain in the scissile peptide bond upon folding thereby accelerating peptide cleavage by facilitating amide nitrogen protonation.

1.3 MUC1

The MUC1 mucin is a type 1 membrane-bound protein. It is expressed in epithelial cells and is a part of the protective mucosal layer. It consists of a large, *O*-glycosylated N-terminal domain with a repeated sequence rich in proline, threonine and serine (PTS-domain). The SEA domain is located just outside the membrane-spanning domain and the cytosolic domain that is involved in signaling (Gendler 2001; Carraway, Ramsauer *et al*. 2003). The number of repeats in the PTS-domain varies in the population. Here the numbering of residues in the SEA domain of MUC1 is based on a variant containing 42 repeats (SwissProt accession number P15941) (figure 3). The MUC1 protein is over-expressed in cancer tissue and a lot of research has therefore focused on the role of this protein in cancer and the possibilities of using it as a therapeutic target (reviewed by Taylor-Papadimitriou, Burchell *et al*. 2002; Hollingsworth and Swanson 2004; Hattrup and Gendler 2007). Cleavage of MUC1 takes place in the endoplasmatic reticulum within minutes of translation (Hilkens and Buijs 1988) and the cleavage site was mapped to the $G^{1097}S^{1098}$VVV sequence within the SEA domain already in 2001 (Parry, Silverman *et al*. 2001). Also, mutation of the conserved serine residue at the cleavage site has been shown to affect cleavage (Lillehoj, Han *et al*. 2003) but the autoproteolytical mechanism was not elucidated until recently (article I and Levitin, Stern *et al*. 2005). When cleaved, the N- and C-terminal fragments stay associated through non-covalent interactions and the MUC1 protein is presented as a heterodimer on the apical surface of epithelial tissue (Ligtenberg, Kruijshaar *et al*. 1992). The large N-terminal can then be shed off implying a biological function for the cleaved SEA domain as discussed below.

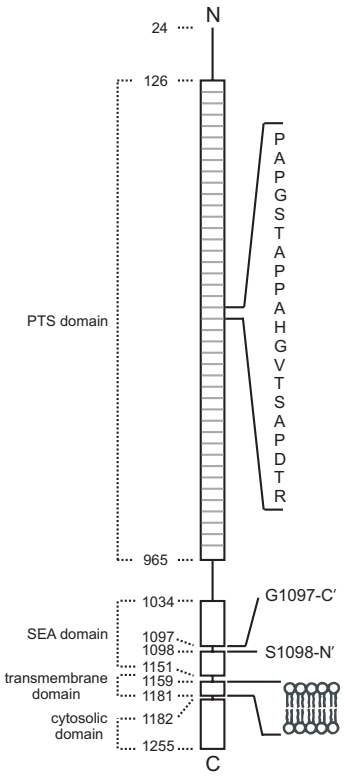

Figure 3. Overview of the full-length MUC1 protein

1.4 GPR116

To confirm the autoproteolytic mechanism the analysis of MUC1 SEA is supported by additional results from the SEA domain of the human GPR116. GPR116 is the human homologue to the IgHepta protein in rat (Abe, Suzuki *et al*. 1999). It is a G-protein coupled receptor containing seven transmembrane helices (figure 4(a)). Cleavage occurs within the SEA domain (Abe, Fukuzawa *et al*. 2002), which is located close to the N-terminal. The sequence identity between the human MUC1 SEA and GPR116 SEA is 19% and includes the conserved cleavage site containing the GSVVV motif (figure 4(b)). NMR experiments reveal a packed hydrophobic core of the GPR116 indicating a well folded protein (III). Also, the far-UV circular dichroism (CD) spectra of GPR116 and MUC1 SEA are very similar (unpublished results). An autoproteolytic mechanism involving strain is thereby expected to be the same as in MUC1 SEA despite the differences in sequence and location in the full-length proteins. Although the mechanism of cleavage indeed proved to be the same the biochemical properties of the domains

differ to some extent. GPR116 SEA turned out to be better suited for stability measurements as denaturation in urea is reversible in contrast to MUC1 SEA, which turned out to be too stable to be fully denatured in urea. The GPR116 SEA was therefore used in the studies of cleavage kinetics in constructs containing mutations in the protein core or in the presence of denaturing agents (III).

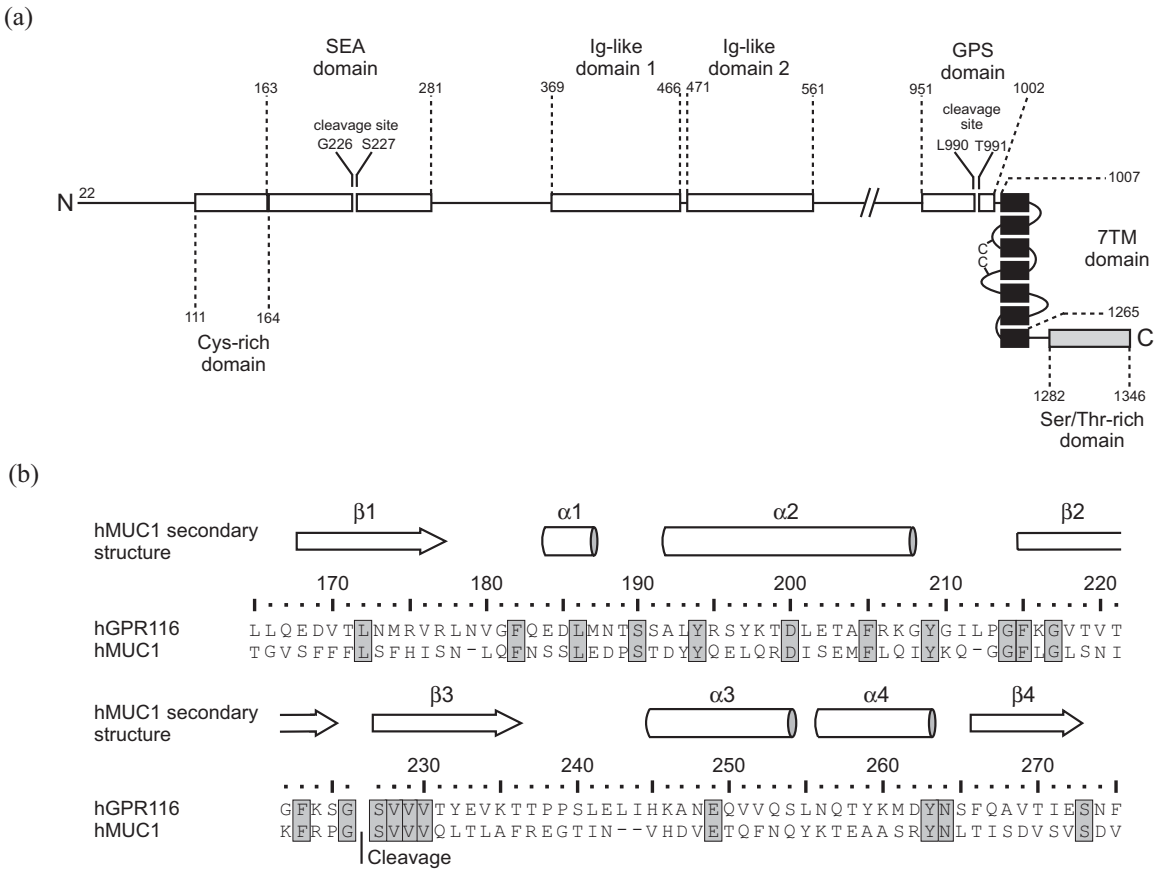

Figure 4. (a) Overview of the GPR116 protein. Extracellular domains are colored white, the transmembrane domain black and the intracellular peptide fragment grey. All domains are denoted as in the rat homologue. (b) The GPR116 and MUC1 SEA subset taken from a multiple sequence alignment between SEA domains (II). Numbering refers to the GPR116 SEA domain and secondary structural elements refer to the MUC1 SEA structure (I). Figure adapted from the supplementary material to article III.

1.5 The SEA domain

The SEA domain was named from the proteins where it was first discovered: sea urchin sperm protein, enterokinase and agrin (Bork and Patthy 1995). It consists of 100-150 amino acids and exists as an intact domain or as a cleaved heterodimer. The sequence

identity between SEA domains vary. Notably, there is a conserved cysteine pair but it is not present in the cleaved SEA domains while the SEA domains having the conserved cysteine pair lack a potential cleavage site (I). This implies diversity in evolution where some of the SEA domains have evolved to be stable while others have evolved to enable cleavage. The cleavage is autoproteolytic and initiated by an N→O acyl shift, as shown below and also by an independent study (Levitin, Stern *et al*. 2005). Since the SEA domain is present in a variety of proteins it is not possible to ascribe a single function to it. The cleaved SEA domain of the membrane bound MUC1 protein has been suggested to function in "receptor-ligand alliances" (Wreschner, McGuckin *et al*. 2002). In article I it is suggested to be involved in the protection of the mucosal layer protecting the apical surface of the airways as part of the innate defense mechanism of mucus clearance (Knowles and Boucher 2002). The suggestion is based on previous observations connecting the shedding of the N-terminal to intracellular signaling. It is further proposed that the shedding is a result of mechanical forces acting on the extended MUC1 N-terminal, thereby disrupting the non-covalent interactions in the cleaved SEA domain (I). However, there are also suggestions of enzymatic action leading to shedding of the N-terminal (Thathiah, Blobel *et al*. 2003). In either case the biological importance of SEA domain autoproteolysis is confirmed as cleavage of the SEA domain is a prerequisite for the subsequent shedding at the epithelial surface. Also, the structure of the MUC1 SEA domain, discussed later in this thesis, contains possible sites for protein-protein interaction. Hence, shedding of the N-terminal may lead to altered affinity for a binding protein, which in turn could enable signaling (figure 5). No binding partner has yet been found but if discovered it may provide additional knowledge of the biological function of the MUC1 SEA domain.

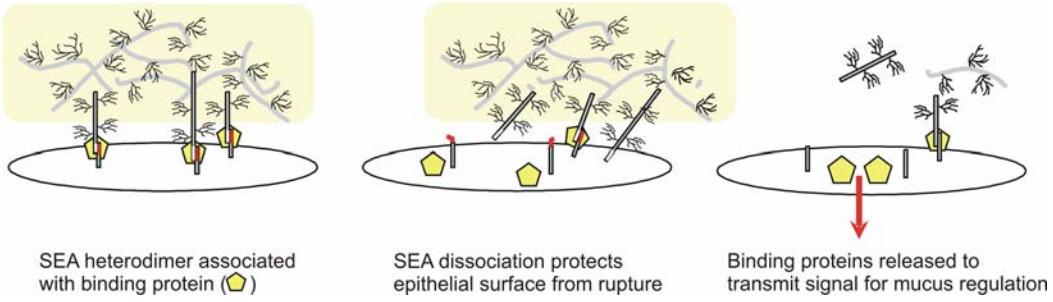

Figure 5. Mechanism for protection of epithelial membranes and signaling of shedding.

2 Methods

Detailed descriptions of the methodology are presented in articles I to IV and will not be repeated here. Instead this section aims to motivate the choice of methods and the specific considerations associated with them.

2.1 Protein production and purification

Today it is possible to produce recombinant proteins in a number of expression systems such as bacteria, yeast, insect cells and mammalian cells. Bacteria are used most frequently and there is a number of established strains and promoter systems available (Terpe 2006). Also, the amounts of stable isotope-labeled material needed for structural studies by nuclear magnetic resonance (NMR) spectroscopy is costly and difficult to obtain in mammalian and yeast strains without extensive optimization of growth conditions. Considering this *Escherichia coli* was chosen as host for expression of the SEA domain.

All protein constructs were designed using standard methods. Mutant constructs were named from the wild type (wt) residue, followed by position and mutant residue, *e.g*. in the S1098A mutant the serine at position 1098 was replaced by an alanine. In cases where glycine residues were inserted at the cleavage site the constructs were named according to the number of residues inserted *e.g.* in the 4G four glycine residues were inserted. A His-tag containing a $GSSH_6$ sequence was incorporated at the N-terminal of all SEA-domain constructs to enable purification using immobilized metal affinity chromatography (IMAC).

Purification by IMAC was sufficient to confirm cleavage by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE), but further purification by size exclusion chromatography (SEC) was needed to perform NMR spectroscopy and CD spectroscopy. An additional advantage of SEC is the ability to detect multimer conformations, exemplified in the analysis of the N1051A mutant, where a substantial fraction of the purified protein formed soluble dimers unable to undergo autoproteolysis (II).

2.2 Circular dichroism (CD)

Considerations when performing CD measurements are reviewed by Kelly, Jess *et al*. (2005). CD measures the difference in absorbance of the left and right fractions of plane-polarized light. In proteins the main optically active groups are the peptide bonds and the aromatic side-chains. Due to the specific arrangement of peptide bonds in the α-helices and β-sheets it is possible to gain information on the secondary structure of proteins by measuring CD in the far-UV range (190-240 nm). There is also software available to determine the amount of α-helix and β-sheet from far-UV spectra (Greenfield 2004). However, in the study of the SEA domain no quantitative analyses were necessary since CD was utilized to compare the secondary structure content of mutant structures and the wt SEA domain and to determine protein stability by thermal and chemical denaturation.

CD may also be used to gain information on the tertiary structure of proteins by measurements in the near-UV region (240-350 nm). No signal is detected for unstructured proteins in this range but packing of the hydrophobic core renders an achiral environment that gives rise to a measurable signal. This was utilized, together with data from NMR experiments, to show the concomitant cleavage and adoption of native fold in the MUC1 SEA domain (II, III and IV). The rate of formation of a native protein core was manifested as an increase in CD-signal at 249 nm. The details regarding the specific structural changes giving rise to this signal has not been elucidated, but the wavelength suggests that the signal arises from the packing of phenylalanine side chains (Kelly, Jess *et al*. 2005).

2.3 Nuclear magnetic resonance (NMR) spectroscopy

The theoretical background to NMR spectroscopy is extensive (see Keeler 2005) and not within the frame of this work. It is established as a powerful tool to study protein structure and a good complement to crystallographic studies as it enables studies of proteins in solution (Wüthrich 1990). It is also an evolving technique and advances are made towards better understanding of dynamic events in proteins (illustrated in Ishima and Torchia 2000; Lange, Lakomek *et al*. 2008; Loria, Berlow *et al*. 2008). In the present study NMR was utilized to determine the solution structure of the MUC1 SEA domain

and to detect interconversions in the precursor structure. Furthermore the formation of native fold was studied, including detailed information of formation of the protein core, manifested as distinct peaks in the methyl region of the ^{1}H NMR spectrum (II).

2.4 Protein denaturation

Denaturation of a protein domain is the transition from native fold (N) to denatured state (D). The denatured state is assumed to be equivalent in energy to the completely unfolded state. This is a simplified view since both states involve an ensemble of conformations and residual secondary structure may be present even at higher concentrations of denaturant (Tsumoto, Ejima *et al*. 2003; Bowler 2007). The folding free energy is here defined as the free energy of the denatured state relative to the free energy of the folded state (denoted $\Delta G_{D\text{-}N}$). As mentioned above CD spectroscopy was the method of choice to study protein denaturation. Confirming results were also obtained by fluorescence spectroscopy. These are two of several techniques based on the same principles, where a physical property, reflecting the content of secondary or tertiary structure, is measured as a function of denaturant (chemical denaturation) or as a function of temperature (thermal denaturation). However, thermal denaturation of the MUC1 SEA domain turned out to be irreversible and no reliable values for the folding free energy could be obtained. Instead information regarding folding free energies was deduced from chemical denaturation of the GPR116 SEA using either urea or guanidinium hydrochloride (GdmCl). In the analysis the fraction folded protein was plotted versus concentration of denaturant and fitted to the Santoro-Bolen equation (Santoro and Bolen 1988). It is important to point out that the experimental value obtained for the folding free energy is extrapolated from an assumed linear relation between concentration of denaturant and fraction of folded protein in the transition between the native and the denatured state (see article III for details). This may lead to uncertain values since the extrapolation is long for stable proteins and may deviate from a linear dependence as the concentration of denaturant reaches zero (Fersht 1999). The difference in folding free energy ($\Delta\Delta G_{D\text{-}N}$) between SEA constructs can be determined with higher accuracy since any curvatures are expected to be equal for the wt and mutated constructs. However, comparison of the SEA constructs

studied here calls for correction for the entropic effects associated with the closing of the loop in the 4G mutant as explained in article III (Nagi and Regan 1997; Zhou 2004).

2.5 Protein folding

Although a lot of progress has been made over the last years the details of protein folding reactions are still under debate (Demchenko 2001; Fersht 2008). But it is generally established that folding of small domains is rapid with folding rates on millisecond to microsecond timescales. It is thereby reasonable to assume that the folding of the SEA domain is very fast compared to the rate of cleavage ($t_{1/2}$ = 18 min at pH 7.5 (III)). It must also be noted that the concomitant folding and cleavage reaction for the SEA discussed here refers to the changes in conformation from monomer precursor to cleaved heterodimer. The conversion from unfolded protein to monomer precursor conformation is very fast compared to cleavage and does not affect the kinetics of the autoproteolysis in wt or the 1G constructs. However, it is possible to affect the equilibrium between unfolded and monomer precursor states by manipulation of the difference in free energy of the two states. This is further discussed in article III and in relation to the kinetic model presented below.

Refolding experiments were performed to confirm that the cleavage of SEA is autoproteolytic (I). Protein produced as inclusion bodies was dissolved in GdmCl and purified by IMAC. This also enabled studies of the kinetics of wt autoproteolysis which is too fast to follow for SEA domain present in the soluble fraction of lyzed *E. coli* (III). However, protein refolding is a complex mechanism and competitive reactions leading to degradation and aggregation takes place simultaneously (Tsumoto, Ejima *et al*. 2003). This was partly prohibited by using a method involving immobilized protein (III). The drawback of this method is that it only allows for measurements within a limited pH range. The pH dependence was therefore performed in solution using a slow-cleaving mutant.

3 Results

The following section describes the key results from analyses of SEA domain autoproteolysis. First the solution structure of the MUC1 SEA domain is presented together with evidence for autoproteolysis and initial indications of strain in the precursor structure (I). The involvement of strain as an accelerator of the autoproteolysis raises a number of implications which are described and analyzed in articles II and III. Here they are discussed in terms of TS theory. Finally a detailed mechanism for SEA domain cleavage is presented and the various ionization reactions involved in the N→O acyl shift are sorted out (IV).

3.1 Cleavage of the SEA domain is autoproteolytic

Purified MUC1 SEA domain produced in bacteria and analyzed by SDS-PAGE reveals three fragments. These correspond in size to the full-length and the N- and C-terminal fragments of the SEA domain cleaved between the Gly1097 and Ser1098 residue at the $G^{1097}S^{1098}$VVV site (I). Cleavage is also detected in MUC1 SEA produced in mammalian cells and in *in vitro* translation assays and it is not reduced in presence of protease inhibitors (article I and Ligtenberg, Kruijshaar *et al*. 1992). Furthermore, cleavage has been confirmed in SEA produced in bacterial systems by independent research groups (Levitin, Stern *et al*. 2005). All these observations indicate that the cleavage is autoproteolytic since an unknown protease, present in all systems above and active at a wide range of pH and temperature (I), is highly unlikely.

Even more convincing evidence of an autoproteolytical mechanism is revealed by refolding studies of the SEA domain. Figure 6 illustrates SDS-PAGE analysis of SEA obtained from inclusion bodies dissolved in GdmCl, loaded on an IMAC column (L) and eluted under denaturing conditions (D) or refolded by dilution (R). Upon refolding the cleavage fragments are readily detected. As a control a construct lacking the conserved serine residue was used. Later studies showed that cleavage can be detected in this mutant as well but it proceeds through an alternative reaction pathway that is very slow at neutral pH (see below).

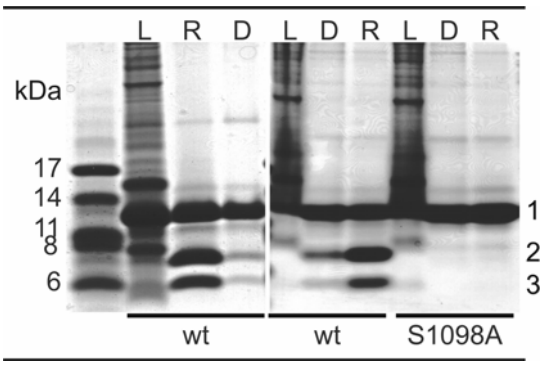

Figure 6. Purification and refolding of wt SEA and the S1098A mutant from bacterial inclusion bodies. The lanes are marked to indicate the material loaded on IMAC column (L) and protein eluted in GdmCl (D) or refolded in native buffer conditions (R). Protein bands marked 1,2 and 3 correspond to full-length SEA and N- and C- terminal fragments of the cleaved heterodimer, respectively. Figure adapted from article I.

3.2 The structure of the SEA domain

The structure of the cleaved SEA domain was determined using NMR spectroscopy (I). It is a very tight structure without any disordered loops. The original pair of N- and C-termini and the N´- and C´-termini arising from proteolysis are located on opposite lobes of the heart-shaped domain (figure 7(a)). The surface contains two hydrophobic patches that might be involved in protein-protein interactions as discussed above. Also a very acidic region of the surface is formed by Glu1059, Asp1060, Asp1064 and Glu1068. These amino acid side chains comprise a negatively charged surface area at physiological pH. The function of this area is not known but it could be involved in orientation of the domain relative to the cell membrane. Furthermore, the potential sites for N-glycosylation at Asn1055 and Asn1133 are exposed (figure 7(b)). N-glycosylation has been mapped to this area (Parry, Hanisch *et al*. 2006) but is not expected to affect the autoproteolytic process since cleavage is detected in systems lacking enzymes for glycosylation.

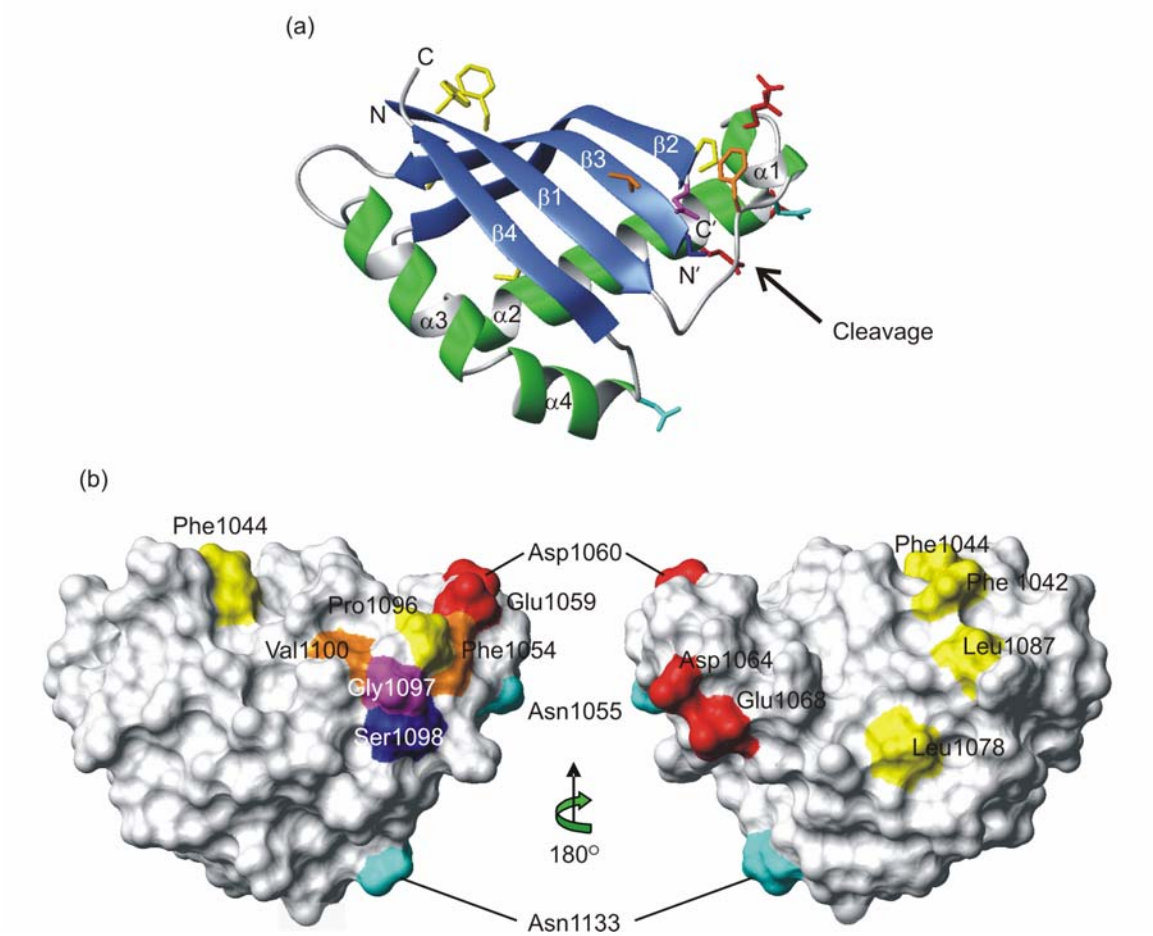

Figure 7. The structure of human MUC1 SEA. (a) Ribbon representation of the SEA structure. The original N- and C-termini and the new N′- and C′-termini arising from proteolysis are shown. The Gly1097 and Ser1098 at the novel N′- and C′-termini (colored magenta and blue, respectively) are integrated parts of β-sheets 2 and 3. (b) Surface representation of MUC1 SEA. Side chains of amino acids forming hydrophobic patches (orange and yellow) and side chains creating an acidic environment (red) are shown. The potential sites for N-glycosylation (cyan) are exposed at the surface. Figure adapted from article I.

The structure at the cleavage site is well defined and the Gly1097 and Ser1098 are integrated into β-sheets 2 and 3 after cleavage. Attempts using computer modeling to introduce a *trans* peptide bond between Gly1097 and Ser1098 followed by energy minimization in which backbone peptides were restrained to remain planar results in a distortion affecting the β-sheet hydrogen bonding throughout the structure (figure 8 and articles I and II).

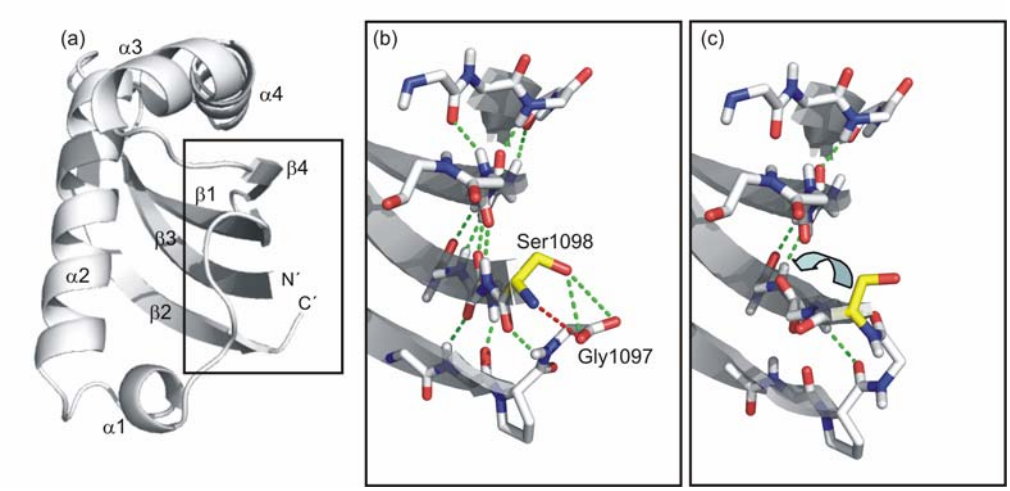

Figure 8. Introduction of a *trans* peptide bond between Gly1097 and Ser1098 in the cleaved MUC1 SEA structure. (a) Ribbon representation illustrating the location of the cleavage site. (b) Detailed view showing backbone conformation and selected side chains in the vicinity of the cleavage site of the experimentally determined, cleaved, MUC1 SEA structure. (c) Modeling of an uncleaved conformation obtained by introducing a *trans* peptide bond linking Gly1097 and Ser1098, followed by energy minimization in which backbone peptides are restrained to remain planar. Heavy atom colors indicate nitrogen (blue), oxygen (red), carbon (white) and C^{α} and C^{β} of Ser1098 (yellow). Backbone hydrogen bonds are shown to illustrate peptide bond geometry and backbone hydrogen bonds are represented by green dashed lines. Figure adapted from article II.

The tight structure of the MUC1 SEA domain and the inability to retain proper β-sheet conformation upon introduction of a planar *trans* peptide bond between Gly1097 and Ser1098 led to the suggestion that folding induces strain on the scissile peptide bond. The involvement of strain is further supported by comparison of the cleaved MUC1 SEA with the uncleaved SEA domain from a murine homologue of human MUC16 (Maeda, Inoue *et al*. 2004). The latter domain contains a loop at the site corresponding to the cleavage site in MUC1 SEA (figure 9). This implicates that a native conformation can be reached without breakage of the peptide backbone due to increased flexibility at the cleavage site. On the other hand a crystal structure of the SEA domain of a transmembrane protease from mouse[*] suggests that the domain is uncleaved. However,

[*] Xie, Y., Kishishita, S., Murayama, K., Hori-Takemoto, C., Shirozu, M., Yokoyama, S., Chen, L., Liu, Z.J., Wang, B.C., RIKEN Structural Genomics/Proteomics Initiative (RSGI). Protein Data Bank accession number 2e7v "*To be published"*

this domain has a nucleophilic side chain of threonine in a GTGVV motif located in a tight loop at the site corresponding to the MUC1 SEA cleavage site. These prerequisites do not agree with an uncleaved domain. Also, biochemical data on the latter structure are still to be published and these might in fact reveal cleavage.

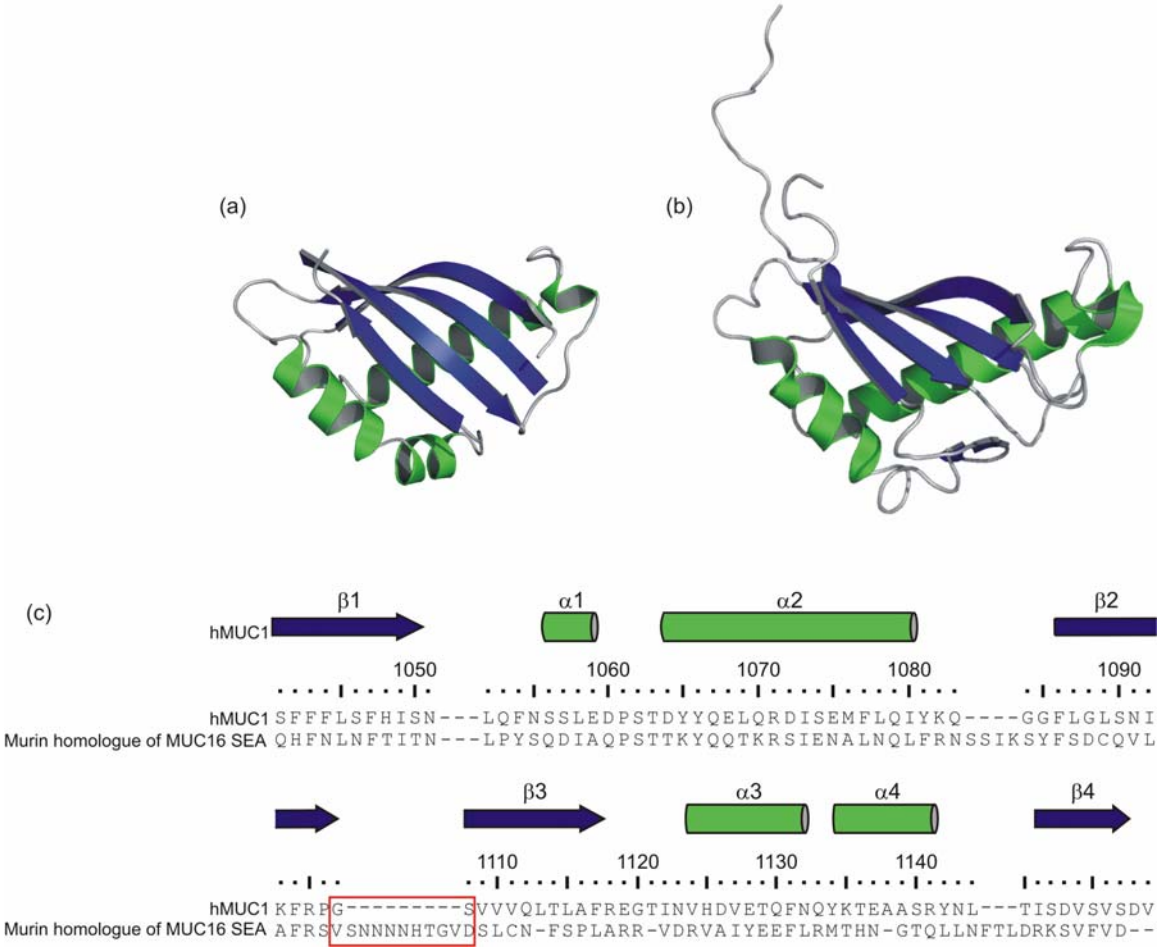

Figure 9. Ribbon representations of the SEA domain. (a) The solution structure of the cleaved MUC1 SEA domain. (b) Solution structure of the uncleaved SEA domain from the murine homologue of human MUC16. (c) Alignment of the sequences taken from a multiple sequence alignment between several SEA domains (article I). The structure in (b) contains a loop at the site corresponding to the MUC1 SEA cleavage site (highlighted by a red box in the alignment). The loop enables formation of native conformation without introducing strain in the structure.

3.3 The 4G mutant

To test whether a loop conformation can relieve strain in the SEA precursor structure and thereby prohibit cleavage four glycine residues were introduced at the cleavage site of MUC1 SEA (generating a GGGGG^{-1}S^{+1}VVV motif and denoted 4G). As shown in figure 10 insertion of four glycine residues leads to abolished cleavage. Interestingly, NMR spectroscopy shows that the 4G mutant adopts a fold identical to the cleaved wt while the uncleaved S1098A mutant is unable to adopt native structure (figure 10(b)).

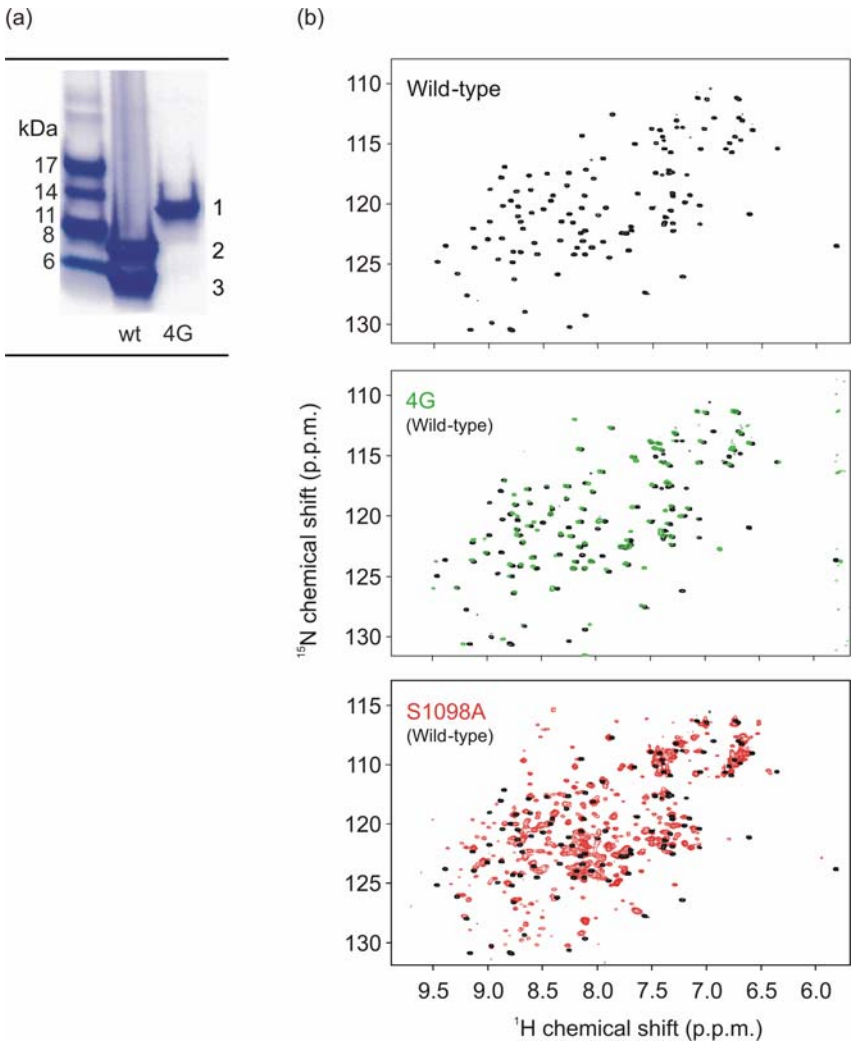

Figure 10. (a) SDS-PAGE analysis of a SEA construct where four glycine residues were inserted at the cleavage site (4G) compared to wt SEA. (b) ^{15}N HSQC spectra of wt SEA (black), 4G (green) and S1098A (red). Figure adapted from article I.

The biochemical and structural data presented so far indicate a connection between adoption of native fold and autoproteolysis. The results are in agreement with a strained precursor conformation as the S1098A has troubles with folding due to inability to undergo N→O acyl shift and thereby relieve the strain in the precursor. Furthermore the 4G construct adopts the native fold but remains intact due to the release of strain resulting from the insertion of glycine residues.

3.4 Mutation of polar residues in the vicinity of the cleavage site

It has previously been shown that replacement of the conserved serine in MUC1 SEA leads to an uncleaved construct (Levitin, Stern *et al*. 2005). The effect was confirmed here by biochemical studies at neutral pH where the serine residue was replaced by an alanine leading to a construct stuck in the precursor conformation (figures 6 and 10). However, cleavage of the S1098A mutant can be accelerated by elevated pH and temperature indicating an alternative reaction pathway not involving the N→O acyl shift (see below). The effect of other polar residues in the vicinity of the cleavage site was also tested by replacing one by one with alanine. None of the mutations, except for the S1098A and N1051A had any major effect on autoproteolysis (figure 11).

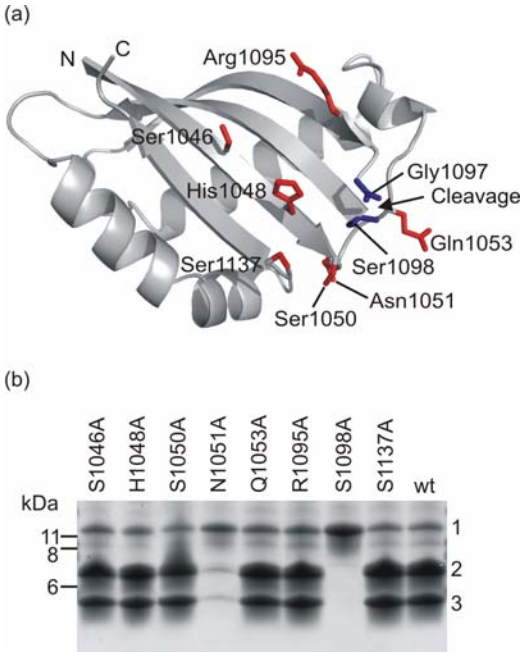

Figure 11. The effect of mutations of polar residues in the vicinity of the cleavage site. (a) Ribbon representation of SEA showing amino acid residues subjected to alanine-scanning mutagenesis (red sticks). Side chains of the amino acids at the cleavage site are shown in blue. (b) SDS-PAGE analysis of the effect on autoproteolysis of alanine substitutions of the residues in (a). Figure adapted from article II.

The effect seen in the N1051A mutant is explained as due to destabilization of the precursor, leading to an increase in the rate of aggregation and an increased formation of soluble, uncleaved dimer conformation rather than a direct effect on autoproteolysis (II). This is also supported by the solution structure in which the Asn1051 side chain is directed away from the cleavage site and appears to participate in at least one hydrogen bond with the backbone and/or side chain of Asn1133.

The result of the alanine scanning mutagenesis has interesting consequences as it rules out the involvement of an active deprotonation system of the attacking serine residue. Also, it confirms the inability to find any structural prerequisites for involvement of side chains other than the conserved serine in the cleavage reaction (I). Hence, the structural and biochemical results suggest that the reaction is accelerated by a destabilized precursor rather than by stabilization of the TS, which is a very common feature of enzymatic reactions.

The results from 4G and S1098A establish the connection between adoption of native fold and autoproteolysis. However, there are further implications associated with a mechanism where strain imposed in the precursor structure accelerates autoproteolysis. These are discussed in terms of TS theory below.

3.5 Kinetic model of SEA autoproteolysis

The equilibrium constant (K) for a reaction is dependent on the difference in Gibbs free energy (denoted ΔG) between the start and end states according to equation (1)

$$\Delta G = -RT \ln(K) \tag{1}$$

where R is the gas constant and T is the temperature in Kelvin. The relationship enables manipulation of the equilibrium constants for chemical reactions by affecting the stability of the end states. It is also applicable on the mechanism of protein folding where the equilibrium of unfolded and folded states is dependent on the differences in conformational free energy (Fersht 1999).

TS theory considers the ground state and the most unstable species on the reaction pathway, the transition state, occurring at the peak in a plot of free energy versus reaction coordinate (see figure 12). The importance of TS theory is that it relates the rate of a reaction to the difference in free energy between the TS and the ground state (the free energy of activation, denoted $\Delta G^{\ddagger}$). If the ground state and the TS are considered to be in thermodynamic equilibrium the rate of a uni-molecular reaction (k_1) is given by multiplying the concentration of the TS and the rate constant for its decomposition. The rate constant for the decomposition can be derived from quantum theory and classical physics. The concentration of the TS is given by the relationship between the equilibrium constant and the difference in free energy($\Delta G^{\ddagger}$) according to equation 1 (Eyring 1935; Fersht 1999). The first order-rate constant for the decomposition of a reactant is thereby defined by equation 2.

$$k_1 = \left(\frac{k_b T}{h_p} \right) \times e^{\left(\frac{-\Delta G^{\ddagger}}{RT} \right)} \qquad (2)$$

where k_1 is the rate constant, k_b is the Boltzmann constant, h_p is the Planck constant and T is the temperature in Kelvin. It is thus possible to affect the rate of a reaction by manipulating the free energy of activation. Based on equations 1 and 2 and the mechanism for SEA autoproteolysis a model of the kinetics and energetics of the reaction is presented in figure 11. It involves the N→O acyl shift accelerated by conformational strain and subsequent hydrolysis of the ester. It describes the entire reaction starting with the folding equilibrium between unfolded protein and the cleavage competent strained precursor conformation. As folding kinetics from unfolded to monomer precursor state is fast the fraction present in monomer precursor state is defined by the partitioning constant (denoted $K_{precursor}$). Cleavage then proceeds via two consecutive reactions, the N→O acyl shift and the hydrolysis of the resulting ester. The rate constants for these reactions are denoted $k_{(N\rightarrow O)H}$ and $k_{(EH)H}$. The index H indicates that the reactions are dependent on pH. There is also an alternative reaction pathway, without the N→O acyl shift, but this reaction is slow at neutral pH and not included in this model (see section on

pH dependence of autoproteolysis). In figure 12(b) the energetics of the autoproteolysis is outlined. To illustrate the entire mechanism the figure involves the thermodynamics of both folding and the cleavage reaction. The free energy of the unfolded state compared to the monomer precursor state is defined as $\Delta G_{precursor}$. The free energy of activation for the conversion of the peptide bond to an ester is defined $\Delta G^{\ddagger}$ and it is considered to be the rate determining step of the autoproteolytic reaction. The difference in energy of the cleaved wt and the unfolded state is defined as $\Delta G_{D\text{-}N}$. Two factors contribute to the decrease in the energy of activation and thereby accelerate the reaction compared to uncatalyzed cleavage of peptide bonds. First, the strain imposed on the protein precursor (denoted ΔG_{strain}) defined as the difference in energy between the 4G mutant, where the strain is released (I, II, III), and the monomer precursor. This definition holds up as the 4G and wt are equally stable (III). The second factor leading to a decrease in $\Delta G^{\ddagger}$ is the structural prerequisites for the N→O acyl shift (denoted $\Delta G^{\ddagger}_{N\rightarrow O\ acyl\ shift}$).

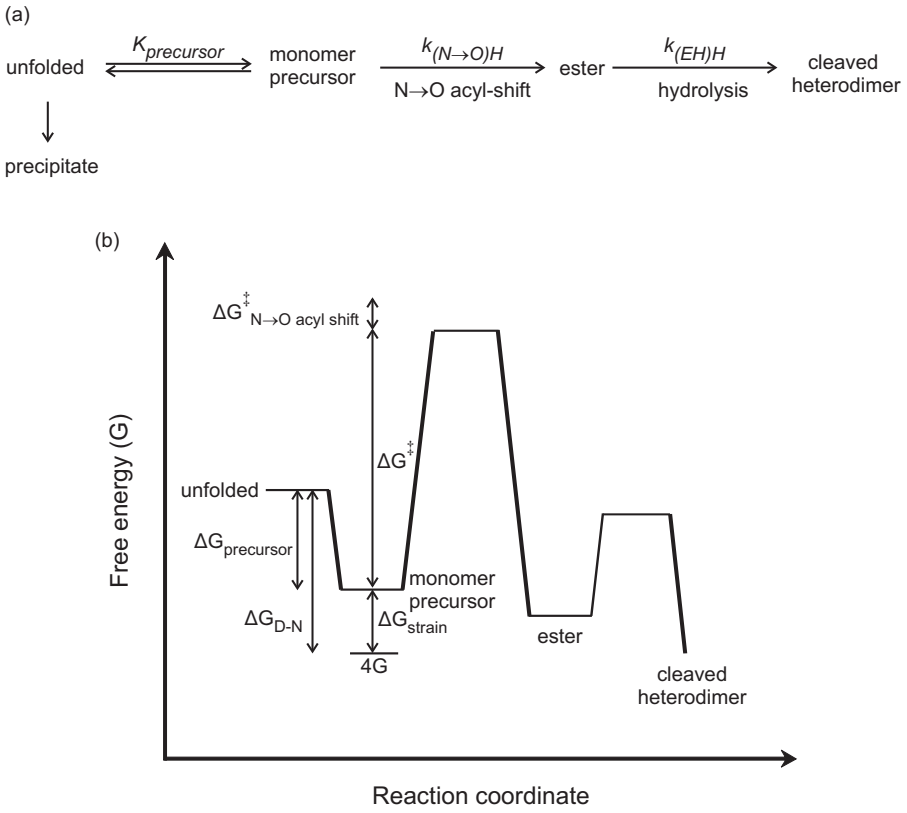

Figure 12. (a) Kinetic preequilibrium model of SEA domain cleavage at neutral pH involving adoption of a cleavage-competent monomer precursor and subsequent cleavage via N→O acyl shift and ester hydrolysis. (b) Outline of the energetics of SEA autoproteolysis.

The energy diagram in figure 12 illustrates a simplified reaction pathway but it is sufficient to give an overview of the steps involved and the possibilities of manipulating the rate of cleavage by affecting the energy levels of the reactants.

3.5.1 The apparent rate constant (k_{obs})

As outlined in figure 12(a) the formation of cleaved heterodimer at neutral pH is the result of two consecutive reactions with individual rate constants. The partitioning constant ($K_{\text{precursor}}$) is assumed to be large as is the rate of folding and these factors do not affect the rate of cleavage unless the stability of the precursor is manipulated. It is therefore possible to view formation of cleaved heterodimer as dependent on the rate constants of two consecutive reactions. This enables simulations of the reaction pathway since the concentrations of monomer precursor [MP], ester intermediate [E] and cleaved heterodimer [CH] over time in a reaction involving two consecutive reactions are given by equations 3 to 5 (Fersht 1999).

$$[MP] = [MP]_0 \times e^{(-k_{(N\rightarrow O)H})\times t)} \tag{3}$$

$$[E] = \frac{[MP]_0 k_{(N\rightarrow O)H}}{k_{(EH)H} - k_{(N\rightarrow O)H}} \times \left[e^{(-k_{(N\rightarrow O)H}\times t)} - e^{(-k_{(EH)H}\times t)}\right] \tag{4}$$

$$[CH] = [MP]_0 \left\{1 + \frac{1}{k_{(N\rightarrow O)H} - k_{(EH)H}} \times \left[k_{(EH)H} \times e^{(-k_{(N\rightarrow O)H}\times t)} - k_{(N\rightarrow O)H} \times e^{(-k_{(EH)H}\times t)}\right]\right\} \tag{5}$$

where $[MP]_0$ is the initial concentration of monomer precursor. In the kinetic experiments performed on SEA it was the increase in concentration of CH that was monitored. Unfortunately it is not possible to resolve the individual rate constants of the N→O acyl shift ($k_{(N\rightarrow O)H}$) and the ester hydrolysis ($k_{(EH)H}$) from these measurements. However, an apparent rate constant can be deduced from a mono-exponential fit to the rate of formation of cleaved heterodimer. This is illustrated in figure 13 where the concentrations of the reactants are plotted as a function of time according to equations 3, 4 and 5 for three different scenarios: (a) When $k_{(N\rightarrow O)H}$ and $k_{(EH)H}$ are of equal magnitude,

(b) when $k_{(N\rightarrow O)H}$ is ten times larger than $k_{(EH)H}$ and (c) when $k_{(N\rightarrow O)H}$ is ten times smaller than $k_{(EH)H}$.

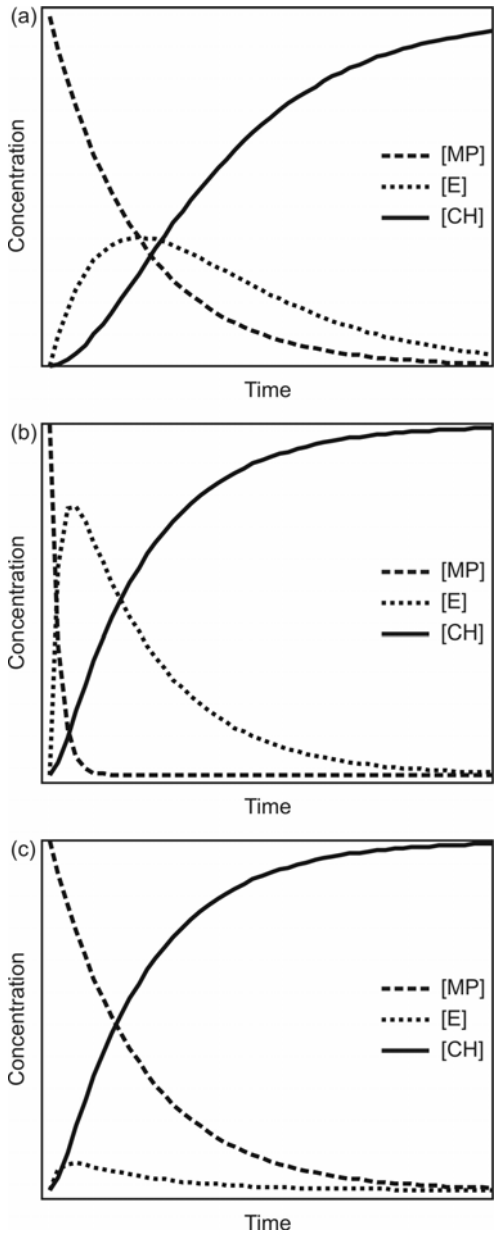

Figure 13. Illustration of the kinetics of SEA autoproteolysis. The reaction proceeds trough two consecutive reactions leading from monomer precursor (MP) to cleaved heterodimer (CH) via an ester intermediate (E). The concentrations of the reactants are plotted versus time according to equations 3, 4 and 5 for three different scenarios. In (a) the rate constants of the N→O acyl shift and the ester hydrolysis are equal. In (b) $k_{(N\rightarrow O)H}$ was set as ten times larger than $k_{(EH)H}$. In (c) $k_{(N\rightarrow O)H}$ was set as ten times smaller than $k_{(EH)H}$ which results in a situation where the rate of formation of CH is determined by the rate of the N→O acyl shift.

Except for the slow initial phase the formation of cleaved heterodimer has a mono-exponential appearance regardless of the magnitude of the individual rate constants. Since purified protein used in kinetic studies already is partly cleaved the initial phase can not be experimentally resolved. Thus, the rate of formation of cleaved heterodimer may be deduced, with good accuracy, from a mono-exponential fit. However, it is

important to note that the apparent rate constant obtained does not give any information on the individual rate constants. This is seen when comparing figures 13(b) and (c), where the rate of formation of cleaved heterodimer is equal although the rate limiting step of the reaction is shifted from the N→O acyl shift being ten times faster than the ester hydrolysis to the opposite situation where the ester hydrolysis is ten times faster than the N→O acyl shift.

3.6 The implications of a coupled folding and cleavage reaction

As outlined above it is possible to get an apparent rate constant for the cleavage reaction by monitoring the formation of cleaved heterodimer. This enabled us to address some of the implications of a coupled folding and cleavage mechanism. The most obvious implication is that the adoption of native fold, *i.e.* the conversion from monomer precursor conformation to cleaved heterodimer, should coincide with cleavage. This is experimentally difficult to show because of the rapid autoproteolysis in the wild type, where cleavage of soluble material proceeds to completion during expression and purification (III). However, the model in figure 12 also predicts that it should be possible to obtain slow cleaving mutants by manipulating either the extent of strain or the nucleophilic activity at the cleavage site, *i.e.* increase $\Delta G^{\ddagger}$ by decreasing ΔG_{strain} or $\Delta G^{\ddagger}_{N\rightarrow O\ acyl\ shift}$. The 4G and S1098A mutants presented above demonstrate that this is possible and further studies of a mutant where only one glycine was inserted (denoted 1G) proved to cleave at a rate suitable for kinetic studies (see below and article II). The 1G studies show that it is possible to fine tune the cleavage rate by a partial decrease of ΔG_{strain} leading to a partially increased $\Delta G^{\ddagger}$. The model also predicts a cleavage competent state that should involve both a strained peptide bond and close proximity between the Ser1098 hydroxyl and the carbonyl carbon of Gly1097. Finally it should be possible to affect the cleavage rate by manipulating the stability of the precursor, *i.e.* reduce the rate of the cleavage reaction by affecting the partitioning constant ($K_{precursor}$).

The implications listed above are addressed in articles II and III. The summary of the results given below are all in support of the proposed model involving acceleration of the reaction by strain imposed in the precursor. The destabilizing mutants also enabled

quantification of the amount of strain in the precursor calculated based on a model relating precursor stability to rate of cleavage (III).

3.6.1 Autoproteolysis and folding coincide

The first implication of a coupled folding and cleavage mechanism is that they should coincide. However, the rapid autoproteolysis in soluble wt SEA prevents reliable measurements of the cleavage rate. Instead measurements where done on a mutant where one glycine was inserted at the cleavage site (1G) (II). 1G proved to be slow-cleaving, not only enabling kinetic studies, but also further confirming the implication that the cleavage rate can be manipulated by decreasing the amount of strain imposed on the precursor. Cleavage of 1G was monitored by SDS-PAGE and adoption of native fold manifested as a change in CD signal at 249 nm. Purified 1G is already partially cleaved but the slow rate of autoproteolysis allows for measurements as cleavage proceeds to completion. As seen in figure 14 autoproteolysis indeed coincides with adoption of native fold.

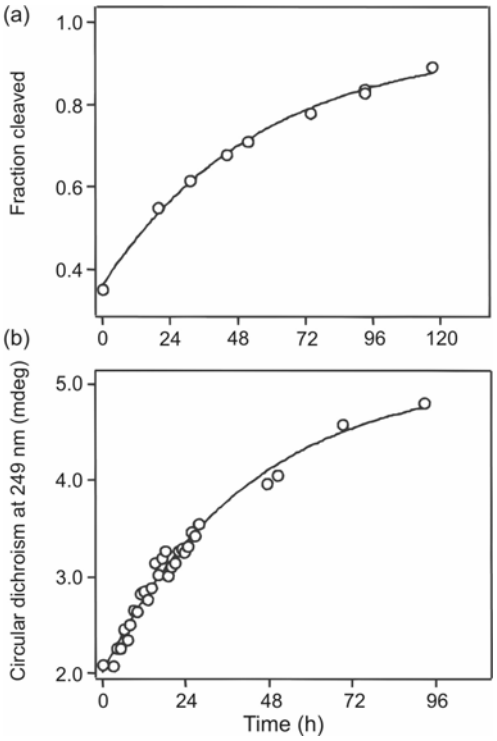

Figure 14. Comparision of kinetics of autoproteolysis and adoption of native fold in the slow-cleaving 1G mutant. (a) Autoproteolysis monitored by SDS-PAGE. $t_{1/2}$ = 39 ± 4h (b) Change in near-UV CD at 249 nm $t_{1/2}$ = 29 ± 3 h. Figure adapted from article II.

Autoproteolysis was also monitored for the S1098A mutant. This mutant is unable to undergo the N→O acyl shift, but cleavage through an alternative pathway can be accelerated by raising pH and elevate the temperature (see section on pH dependence). Parallel SDS-PAGE and NMR studies of S1098A revealed concomitant autoproteolysis and adoption of native fold (II). The line broadening and multiple peaks of the NMR spectra of S1098A (see figure 10) disappeared as cleavage proceeded and the spectra of fully cleaved S1098A proved to be very similar to wt. Also, the extent of line broadening and multiple peaks in the spectra of the uncleaved S1098A show that the entire structure of the SEA is affected upon cleavage. Taken together these results confirm the implications that it is possible to affect the rate of cleavage by manipulating the nucleophilic activity and that cleavage has a global effect on precursor structure (II). Furthermore, the fact that cleavage occurs specifically between the alanine and the glycine in the S1098A mutant corroborates the presence of a distorted peptide bond in the precursor structure.

3.6.2 The cleavage competent conformation

The model in figure 12 predicts that a cleavage competent monomer precursor conformation should exist. This conformation should include close proximity of the nucleophilic side chain of the conserved serine and the carbonyl carbon of the scissile peptide bond at the same time as the bond is distorted. *i.e.* $\Delta G^{\ddagger}$ is reduced by $\Delta G^{\ddagger}_{N\rightarrow O \text{ acyl shift}}$ and ΔG_{strain}. Such a cleavage competent conformation was revealed in molecular dynamics studies of the MUC1 SEA domain. Simulations were performed starting from the structure with a modeled peptide bond between Gly1097 and Ser1098 as outlined in figure 8. A conformation involving the prerequisites regarding distance and strain was reached several times during a 10 ns trajectory (II).

3.6.3 Destabilization of the precursor

So far, the connection between autoproteolysis and adoption of native fold has been thoroughly established by biochemical and structural analyses of the SEA domain. However, the energetics of the mechanism must also be accounted for. That is, the

reduction in free energy due to adoption of native fold must be large enough to compensate for the energetically unfavored distortion of the scissile peptide bond. This issue is addressed first in article I where it is suggested that conformational strain lowers the free energy of activation for autoproteolysis by 4.5 to 7 kcal mol^{-1} which is considered reasonable since the folding free energy typically is 5 to 15 kcal mol^{-1} for protein domains. In article III this is confirmed by measuring the folding free energy of wt MUC1 SEA to 16.6 ± 0.6 kcal mol^{-1}. It is also shown that the stability of the precursor affects the kinetics of the autoproteolysis. Analysis of relative cleavage rates related to the relative precursor stability allowed for quantification of the strain energy. It was concluded that 7 kcal mol^{-1} is partitioned as strain in the protein precursor of GPR116 SEA. The calculation was based on a pre-equilibrium model described in article III. In brief, the partitioning constant ($K_{precursor}$, see figure 12(a)) was manipulated by mutations in the SEA protein core or by addition of denaturant. Assumptions regarding the folding kinetics then allowed for measurement of the strain energy from a one parameter fit to equation 6 (figure 15).

$$\frac{k_{obs}}{k_0} = \frac{e^{((\Delta G_{D-N} - \Delta G_{strain})/RT)}}{e^{[((\Delta G_{D-N} - \Delta G_{strain})/RT)+1]}} \qquad (6)$$

Where k_{obs} is the experimentally determined rate constant for autoproteolysis in destabilized SEA, k_0 is the rate constant for wt autoproteolysis, R is the gas constant and T is the temperature in Kelvin. The difference in free energy for the unfolded and monomer precursor state is defined as the folding free energy for unstrained SEA ($\Delta G_{D\text{-}N}$) minus the energy partitioned as strain in the precursor (ΔG_{strain}) (see figure 12 and article III). The result was confirmed by the difference in free energy between wt MUC1 SEA and the S1098A mutant, stuck in the precursor conformation (III).

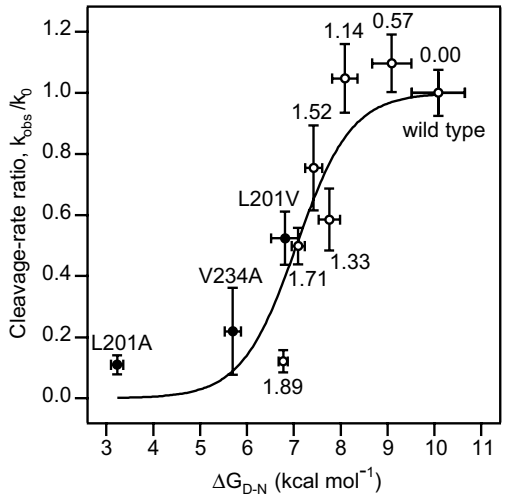

Figure 15. Determination of peptide strain during autoproteolysis. The cleavage rate ratio (k_{obs}/k_0) plotted versus the ΔG_{D-N} for the GPR116 SEA domain. White and black circles represent data obtained for wt protein at different urea concentrations and for the hydrophobic truncation mutants, as indicated. Solid line represents fit to a preequilibrium model results in 7 kcal mol^{-1} for the strain energy. Figure adapted from article III.

3.7 The pH dependence of SEA autoproteolysis

In all experimental setups above the pH was kept constant and all effects caused by an altered pH were thereby cancelled. However, an N→O acyl shift involves various protonation equilibria (Iwai and Ando 1967). Furthermore, the ester is assumed to be resolved by a nucleophilic attack of a water hydroxyl (Levitin, Stern *et al*. 2005 and article I). Hence, the autoproteolysis of SEA is expected to be pH dependent. The pH dependence of SEA cleavage was examined by monitoring the formation of cleaved 1G mutant (article IV). The apparent rate constant obtained shows a clear pH dependence where three kinetic pK_a values are imply three different pH-dependent reactions (figure 16).

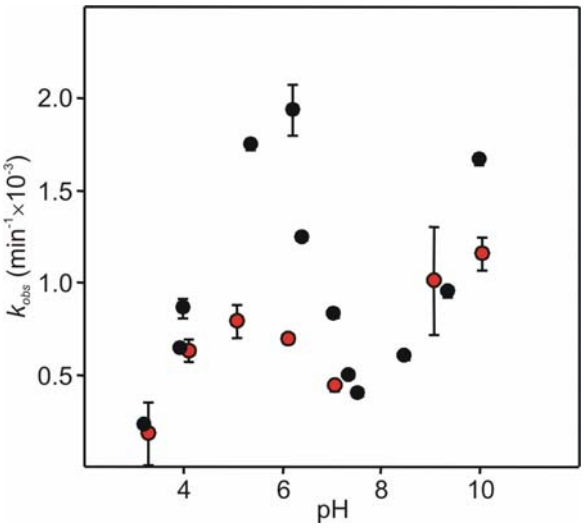

Figure 16. The pH dependence of SEA autoproteolysis. Plot of apparent rate constant (k_{obs}) versus pH obtained from CD measurements at 40 °C (solid black) or SDS-PAGE analysis at 30 °C (red circles). Figure adapted from article IV.

The increased cleavage rate at pH above 8 is in line with the observation of cleavage in the S1098A construct at elevated pH (II). Accordingly it can be ascribed to an alternative reaction pathway for autoproteolysis, without N→O acyl shift. The bell-shape of the pH dependence observed between pH 3 and pH 8 reveals the presence of at least two kinetic pK_a values reflecting different rate-determining ionization reactions. This can be accounted for by a detailed view of the acyl shift reaction pathway (figure 17). The increased cleavage rate between pH 3 and 5 is due to the hydroxyl-mediated ester hydrolysis being rate determining. The decrease between pH 5 and 7 is accounted to the N→O acyl shift, involving protonation of the amide nitrogen. The need for a nitrogen protonation in the N→O acyl shift is generally accepted. An initial protonation leads to stretching of the peptide bond due to loss of π-bond resonance and a consequent increase in carbon electrophilicity (Greenberg, Moore *et al.* 1996). In other words, peptide nitrogen protonation leaves the peptide carbonyl open for nucleophilic attack. Furthermore, experimental data show a correlation between amide pK_a and peptide geometry. Twisted amides have higher gas phase pK_a values than planar amides. Also, the solution pK_a for a distorted amide has been estimated to between 3.5 and 3.8 and it decreases towards the pK_a of a planar amide as the distortion decreases (Wang, Bennet *et al*. 1991). This provides a mechanistic link between the strain in the SEA precursor and the kinetics of cleavage. The strain induced in the precursor structure accelerates the reaction by pyramidalization of the amide nitrogen facilitating amide protonation. The shift in pK_a in the distorted amide allows to ascribe the decreased rate at pH between 5 and 7 to the N→O acyl shift (article IV). However, neither the biochemical data nor the solution structure of the SEA domain reveal any amino acid side chains within the domain involved in ionization reactions necessary for cleavage (I and II). Hence, the protonation and the resolution of the ester must involve other reactants. For the ester hydrolysis a water hydroxyl is proposed to be the attacking species. This is in line with previous assumptions regarding SEA autoproteolysis (Levitin, Stern *et al*. 2005) and supported by the fact that the hydrolysis leads to novel N- and C-termini. For the protonation of the amide in the N→O acyl shift we propose a water hydronium working as an acid and water forming hydronium working as a base. The initial protonation by a hydronium is in line with work on the C-terminal cleavage of inteins, based on quantum

mechanical calculations (Shemella, Pereira *et al*. 2007). The involvement of hydronium, water and hydroxyl is outlined in the detailed mechanism for SEA autoproteolysis via the N→O acyl shift shown in figure 17.

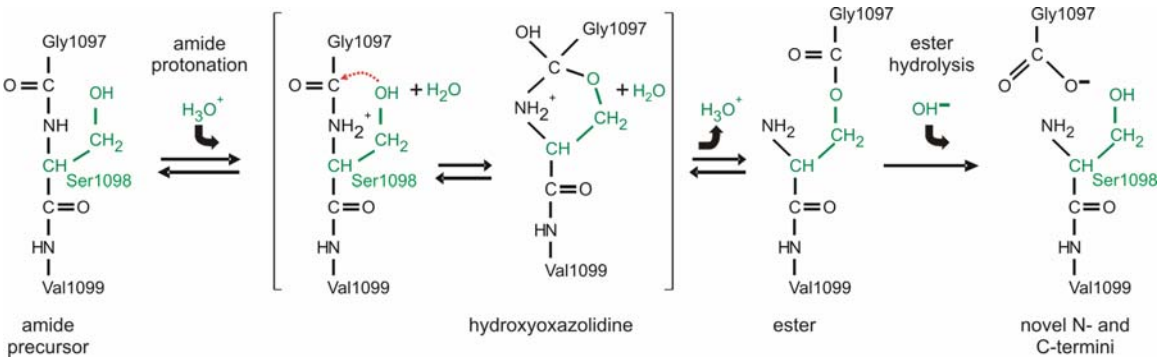

Figure 17. Mechanism SEA autoproteolysis involving an N→O acyl and subsequent hydrolysis of the ester. Figure adapted from article IV. A hydronium works as an acid in the protonation of the amide nitrogen and water works as a base, reforming hydronium in the conversion of the cyclic intermediate to an ester.

The mechanism in figure 17 is also supported by simulations of autoproteolysis performed in the range from pH 3 to pH 8 (IV). It may be claimed that the pH dependence could be due to altered charged groups in the domain leading to destabilization and thereby a decrease in cleavage rate, in analogy with the mutations in the protein core. However, no major effect was detected on tertiary structure, as measured by CD, in the range from pH 3 to pH 8 (unpublished results). Other possible explanations to the observed pH dependence are ruled out based on the biochemical and structural studies of the SEA domain (see article IV).

4 Discussion

The specific results, establishing that cleavage is an autoproteolytic event involving a strained precursor conformation are thoroughly discussed in the enclosed articles (I, II, III, and IV). The intention here is therefore to instead discuss the results in relation to other autoproteolytic systems and classical enzymatic mechanisms.

4.1 The precursor structure

The precursor structure has not been determined for any SEA domain but some information can be deduced from the NMR experiments performed on the S1098A mutant. The methyl region of this mutant shows a number of resonances with upfield chemical shifts. This is clearly an indication of a packed hydrophobic core. However, the shifts do not coincide with the cleaved heterodimer. This is also reflected in the near-UV CD spectra, where an increased signal at 249 nm is associated with the adoption of native fold. The far-UV spectra of S1098A and wt are indistinguishable. Hence, S1098A has identical secondary structure elements but a differently packed hydrophobic core and can not adopt native tertiary structure unless the peptide chain is fragmented. Instead the precursor is an interconverting ensamble of structures, giving rise to line broadening and multiple peaks in the NMR spectra (see figure 10). The existance of a cleavage competent conformation is supported by molecular dynamics simulations, as explained above. However, simulations also showed that a peptide bond in *cis* introduced between Gly1097 and Ser1098 is more compatible with the native structure than a peptide bond in *trans* (II). Notably, a peptide bond in *cis* is less stable than a bond in *trans* and agrees with the hypothesis of acceleration through substrate destabilization. Furthermore other autoproteolytic systems have been shown to have a scissile peptide bond in *cis* (Klabunde, Sharma *et al*. 1998). However, a *cis* bond is not likely to occur in SEA since attempts to undergo *trans*→ *cis* conversion ends up in a cleaved peptide.

4.2 Autoproteolysis in other proteins

As mentioned in the introduction there are a number of proteins undergoing autoproteolysis initiated by an N→O acyl shift. In addition to the SEA domain these include the pyruvoyl-dependent enzymes (van Poelje and Snell 1990), the intein domain in protein splicing (Noren, Wang *et al*. 2000), the hedgehog domain (Porter, von Kessler *et al*. 1995; Porter, Young *et al*. 1996), the N-terminal nucleophile (Ntn) hydrolases (Brannigan, Dodson *et al*. 1995), and the nucleoporins (Hodel, Hodel *et al*. 2002). Although the biological functions of these proteins are diverse there are striking similarities between them when it comes to the mechanism of autoproteolysis. Based on the available structural and biochemical data the main points supporting that strain is involved also in other autoproteolytical systems are summarized below.

4.2.1 Protein stability

One of the implications associated with strain induced cleavage is that the stability of the cleaved domain should be sufficient to compensate for the energy cost associated with a strained scissile bond in the precursor structure. The stability has not been determined for all the proteins in the families mentioned above, but there are indications that the penicillin acylase of the Ntn hydrolase family and the inteins are very stable (Lindsay and Pain 1990; Hiraga, Derbyshire *et al*. 2005). Furthermore, histidine decarboxylase, which is a pyruvoyl-dependent enzyme, is resistant towards thermal and chemical denaturation (Rosenthaler, Guirard *et al*. 1965; Yamagata and Snell 1979). Also, protein domains typically have stabilities in the order of 5 to 15 kcal mol^{-1}. Hence, all proteins mentioned above seem to fulfill the requirement regarding stability.

4.2.2 Indications of strain in the precursor structure

There are precursor structures available for some of the autoproteolytical proteins and it has been suggested that strain is involved in the autoproteolysis of these proteins, as discussed in (III). Mentionable examples are the pyruvoyl-dependent enzyme histidine

decarboxylase where conformational factors were suggested to increase the susceptibility of the scissile peptide bond to nucleophilic attack (Huynh and Snell 1986). There have also been suggestions of strain in the precursor structures of the aspartate decarboxylase and adenosylmethionine decarboxylase (Schmitzberger, Kilkenny *et al*. 2003; Tolbert, Zhang *et al*. 2003). Furthermore, precursor structures of the inteins GyrA and PI-SceI involve strain and it has been shown that the scissile peptide bond of GyrA is "unusual" (Klabunde, Sharma *et al*. 1998; Poland, Xu *et al*. 2000; Romanelli, Shekhtman *et al*. 2004). In addition there are indications of strain in the Nup98 precursor (Hodel, Hodel *et al*. 2002) and the Ntn-hydrolases, although there are conflicting results for the latter group (Ditzel, Huber *et al*. 1998; Xu, Buckley *et al*. 1999; Hewitt, Kasche *et al*. 2000; Kim, Kim *et al*. 2002; Kim, Yang *et al*. 2003; Kim, Yang *et al*. 2006). In light of these results cleavage accelerated by a strained precursor, and involving a distorted scissile peptide bond may well be a general mechanism for reaction pathways involving an N→O acyl shift.

4.2.3 Acceleration of cleavage through transition state stabilization and active deprotonation

We find that strain decreases the free energy of activation for the cleavage of SEA by ~7 kcal mol^{-1} (III). This is not enough to account for the acceleration of the reaction as compared to uncatalyzed cleavage of a peptide bond. It is therefore concluded that the total acceleration of SEA cleavage is due to the concerted action of the N→O acyl shift and the strain imposed on the scissile peptide bond (III). In the SEA domain there is no structural evidence for involvement of any other residues than the conserved serine in the cleavage reaction. Furthermore, alanine replacement mutagenesis and studies of the pH dependence suggest that no deprotonation of the attacking serine hydroxyl is necessary for the reaction (II and IV). This should be considered when discussing the generality of an intramolecular cleavage reaction accelerated by strain. In the examples of other protein families undergoing autoproteolysis there are possible additional factors contributing to the acceleration of the reaction. One of these is the existence of an oxyanion hole, leading to an accelerated reaction through stabilization of the TS. Work on the inteins (Klabunde, Sharma *et al*. 1998; Poland, Xu *et al*. 2000) and Ntn hydrolases

(Ditzel, Huber *et al*. 1998; Xu, Buckley *et al*. 1999; Hewitt, Kasche *et al*. 2000; Kim, Kim *et al*. 2002; Kim, Yang *et al*. 2003; Kim, Yang *et al*. 2006) do argue for an oxyanion hole. However, the observation that cleavage proceeds in a histidine decarboxylase mutant at the bond adjacent to the wt scissile bond (van Poelje and Snell 1990) argues against the necessity of an oxyanion hole since this mechanism probably proceeds trough an alternative pathway, in analogy with SEA S1098A mutant, thus generating a different TS (III and IV)

Yet another factor suggested to accelerate the cleavage reaction is an active deprotonation of the attacking nucleophile. It has been proposed for penicillin acylase (Hewitt, Kasche *et al*. 2000), cephalosporin (Kim, Kim *et al*. 2002; Kim, Yang *et al*. 2003; Kim, Yang *et al*. 2006) and Nup98 (Hodel, Hodel *et al*. 2002). In the pyruvoyl-dependent enzymes and Ntn hydrolases the reports of an active deprotonation are conflicting. Hence, an active deprotonation system seems to be necessary in some, but not all, autoproteolytic mechanisms.

In conclusion, there may be other accelerating factors involved in the autoproteolytic systems mentioned above. However, this does not exclude the involvement of strain in the precursor structures, as the enhancement caused by a strained scissile bond could work in concert with a TS stabilization or various ionizations to give a net acceleration of the reaction. On the other hand, this also implies that the effect of strain may be less pronounced compared to the impact it has in the SEA domain, thus making it harder to evaluate in other systems.

4.3 Autoproteolysis versus enzymatic reactions

At first glance the advantages of autoproteolytical cleavage seem obvious as an intramolecular reaction in general is energetically favored compared to an intermolecular reaction due to the relatively small decrease in entropy. This is because an intramolecular reaction does not involve the loss in entropy associated with the binding in a bi-molecular complex (Fersht 1999). Furthermore, an intramolecular reaction generates a heterodimer from a single gene, compared to a corresponding intermolecular reaction where the formation of a cleaved heterodimer is dependent on the expression of at least two genes.

The fundamental questions associated with enzymatic reactions have been discussed since the introduction of TS theory and the absolute rate of chemical reactions. In recent years the experimental work has been aided by extensive computational calculations. Still, the exact reasons for the amazing rate enhancing capacity of enzymatic reactions are under debate. The generally accepted mechanism is that enzymes lower the free energy of activation by stabilizing the TS. This is achieved in a number of different ways depending on the reaction in question (see Holliday, Almonacid *et al*. 2007). Perhaps one of the most studied mechanisms is that of the serine proteases, where a catalytic triad involving serine, asparagine and histidine, accounts for a rate acceleration of $\sim 2 \times 10^6$ in the hydrolysis of amide bonds (Carter and Wells 1988). This triad is widespread among enzymes and its stereochemistry is preserved despite different folding patterns of the enzymes (Dodson and Wlodawer 1998).

The concept of accelerating a reaction by destabilization of the substrate instead of stabilization of the TS has come up as a plausible alternative every now and then in the history of enzymology. However, it has also been judged as an improbable mechanism for bi-molecular enzymatic reactions based on the possibility of inhibiting reactions by adding TS analogs (Kraut 1988). Here it is brought back into the light and it is shown that the energetic aspects of the mechanism can be accounted for. When discussing substrate destabilization it terms of enzymatic reactions it is important to note that cleavage of the SEA domain displays fundamental differences compared to a classical bi-molecular enzyme reaction. First and foremost it does not involve the formation of an enzyme-substrate complex. Instead a cleavage competent state is formed upon folding of the domain. Since the cleavage competent conformation is energetically favored over an unfolded protein this equilibrium is in strong favor of the former. Hence, formation of a cleavage competent precursor is driven by the folding reaction while the formation of an enzyme substrate complex involves large changes in entropy for two separate proteins. Furthermore, specificity is “built-in” to the system in autoproteolytical reactions and in the case of the autoproteolytical proteins the catalytic efficiency is not that important since every domain only has to cleave once.

From an evolutionary point of view the diversity among enzymes containing a catalytic triad suggests parallel evolution within the group. Although dystroglycan has

been suggested as an ancient precursor of the SEA domains (Akhavan, Crivelli *et al*. 2007) the evolutionary origin is harder to sort out since SEA domain-like folds still are discovered in a diversity of proteins. However, one may suggest that there is an evolutionary connection between the classical enzymes and the SEA domain in analogy with the suggestion that inteins have their origin in primitive enzymes (Paulus 2001). Observations of the Ntn-hydrolases are supporting an evolutionary relationship as they seem to have evolved to combine the catalytic advantages of autoproteolysis and the catalytic triad (Dodson and Wlodawer 1998).

4.4 The novel mechanism

As outlined above the notion of strain in the precursor as an accelerating factor has been suggested previously. From an evolutionary point of view it has probably been around at least as long as the classical enzymes. However, none of the previously proposed mechanisms have included any clear connection between the conformational free energy and the accelerated reaction. Here we present a detailed view of a reaction where folding free energy works as an accelerating force by inducing strain on the scissile peptide bond leading to a pyramidalization of the peptide nitrogen and thereby facilitated protonation. The adoption of a cleavage competent conformation is the main driving force of the reaction. Hence, cleavage could be attained in any sequence provided that the structural prerequisites are fulfilled and that the conformational free energy is high enough to compensate for a distorted scissile peptide bond. In all other mechanisms suggested so far strain is depicted as a possible accelerator but the main catalytic effect has been accounted to side chains participating in protonation reactions and stabilization of the TS. The autoproteolysis of SEA is a novel mechanism where substrate destabilization is the major catalytic force. The discovery of this mechanism also leads to a new view of other autoproteolytic systems where the involvement of strain in the precursor perhaps has been underestimated.

5 Outlook

The results presented here give a complete view of the mechanistic and energetic aspects of SEA domain autoproteolysis. They also provide starting points for many interesting research projects within several fields of science. From the view of a structural biologist the next step could be to further investigate the interconversions in the SEA precursor structure to confirm the existence of a distorted scissile bond. Also, it is possible to further elucidate the various ionizations in the N→O acyl shift using NMR. In addition, the generality of the mechanism could be strengthened by additional studies of other autoproteolytical systems.

From a biological point of view the solution structure of the MUC1 SEA leaves some open questions regarding function. We propose that the MUC1 SEA domain is a key molecule in the regulation and control of the mucosal surface. It is likely that the SEA domain exist in a complex with a yet unidentified binding protein which is released upon shedding of the N-terminal. This may in turn function as a signal that shedding has occurred. Thus, Identification of a potential binding partner would increase the understanding of MUC1 function and perhaps the understanding of how the protective mucus layer is regulated.

Finally, the autoproteolytical process of the SEA domain may have implications in protein production. As the discovery of intein autoprocessing has led to development of protein engineering tools (Wood, Wu *et al*. 1999; Wood, Derbyshire *et al*. 2000) this may also bee the case for the SEA domain. The possibilities of using the entire domain as a "tag" is limited based on its size (see Terpe 2003) but the energetic and structural aspects of the cleavage may enable introduction of autoproteolytical sites where cleavage is accelerated solely by conformational strain in other domains and thereby contribute to the growing source of available tools in protein engineering.

6 Acknowledgements

I would like to thank:

My supervisor Torleif Härd for excellent guidance. Past and present members of the Härd research group, especially Anders Sandberg, Bertil Macao and Wolfgang Hoyer for helpful discussions and scientific guidance.

Past and present members of the Hansson and Enerbäck research groups, especially Martin Lidell. Fellow PhD students for creating a nice atmosphere, especially Joakim Strandberg, Kim Fejgin and Rasmus Jansson. I would also like to thank all other people who made this thesis possible, especially Anneli.

7 References

Abe, J., Fukuzawa, T. *et al*. (2002). "Cleavage of Ig-Hepta at a "SEA" module and at a conserved G protein-coupled receptor proteolytic site." *J. Biol. Chem.* **277**: 23391-23398.

Abe, J., Suzuki, H. *et al*. (1999). "Ig-hepta, a novel member of the G protein-coupled hepta-helical receptor (GPCR) family that has immunoglobulin-like repeats in a long N-terminal extracellular domain and defines a new subfamily of GPCRs." *J. Biol. Chem.* **274**: 19957-64.

Akhavan, A., Crivelli, S. N. *et al*. (2007). "SEA domain proteolysis determines the functional composition of dystroglycan." *Faseb J*. **22**: 1-11.

Bennet, A. J., Wang, Q. P. *et al*. (1990). "Relationship between amidic distortion and ease of hydrolysis in base - If amidic resonance does not exist, then what accounts for the accelerated hydrolysis of distorted amides." *J. Am. Chem. Soc.* **112**: 6383-6385.

Blackburn, G. M., Skaife, C. J. *et al*. (1980). "Strain effects in acyl transfer reactions. Part 5. The kinetics of hydrolysis of benzoquinuclidin-2-one: a torsionally distorted amide." *J. Chem. Research* **(S)**: 294-295.

Blair, W. S., Semler, B. L. (1991). "Self-cleaving proteases." *Curr. Opin. Cell Biol.* **3**: 1039-1045.

Bork, P., Patthy, L. (1995). "The SEA module: A new extracellular domain associated with *O*-glycosylation." *Prot. Science* **4**: 1421-1425.

Bowler, B. E. (2007). "Thermodynamics of protein denatured states." *Mol. Biosyst.* **3**: 88-99.

Brannigan, J. A., Dodson, G. *et al*. (1995). "A protein catalytic framework with an N-terminal nucleophile is capable of self-activation." *Nature* **378**: 416-419.

Carraway, K. L., Ramsauer, V. P. *et al*. (2003). "Cell signaling through membrane mucins." *Bioessays* **25**: 66-71.

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

Demchenko, A. P. (2001). "Concepts and misconcepts in the analysis of simple kinetics of protein folding." *Curr. Protein. Pept. Sci.* **2**: 73-98.

Ditzel, L., Huber, R. *et al*. (1998). "Conformational constraints for protein self-cleavage in the proteasome." *J. Mol. Biol.* **279**: 1187-1191.

Dodson, G., Wlodawer, A. (1998). "Catalytic triads and their relatives." *Trends Biochem. Sci.* **23**: 347-352.

Eyring, H. (1935). "The activated complex and the absolute rate of chemical reactions." *Chem. Rev.* **17**: 65-77.

Fersht, A. R. (1999). "Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding." W.H. Freeman, New York, NY.

Fersht, A. R. (2008). "From the first protein structures to our current knowledge of protein folding: delights and scepticisms." *Nat. Rev. Mol. Cell. Biol.* **9**: 650-654.

Gendler, S. J. (2001). "MUC1, the renaissance molecule." *J. Mammary Gland Biol. Neoplasia* **6**: 339-53.

Greenberg, A., Moore, D. T. *et al.* (1996). “Small and medium-sized bridgehead bicyclic lactams: a systematic ab initio molecular orbital study.” *J. Am. Chem. Soc.* **118**: 8658-8668

Greenfield, N. J. (2004). "Analysis of circular dichroism data." *Methods Enzymol.* **383**: 282-317.

Hattrup, C. L., Gendler, S. J. (2007). "Structure and function of the cell surface (tethered) mucins." *Annu. Rev. Physiol.* **70**: 431-457.

Hewitt, L., Kasche, V. *et al*. (2000). "Structure of a slow processing precursor penicillin acylase from *Escherichia coli* reveals the linker peptide blocking the active-site cleft." *J. Mol. Biol.* **302**: 887-898.

Hilkens, J., Buijs, F. (1988). "Biosynthesis of MAM-6, an epithelial sialomucin. Evidence for involvement of a rare proteolytic cleavage step in the endoplasmic reticulum." *J. Biol. Chem.* **263**: 4215-4222.

Hiraga, K., Derbyshire V. *et al*. (2005). "Minimization and stabilization of the *Mycobacterium tuberculosis rec*A intein." *J. Mol. Biol.* **354**: 916-26.

Hodel, A. E., Hodel M. R. *et al*. (2002). "The three-dimensional structure of the autoproteolytic, nuclear pore-targeting domain of the human nucleoporin Nup98." *Mol. Cell* **10**: 347-358.

Holley, R. W. (1953). "Steric inhibition of amide resonance and its possible significance in enzyme action." *Science* **117**: 23-25.

Holliday, G. L., Almonacid D. E. *et al*. (2007). "The chemistry of protein catalysis." *J. Mol. Biol.* **372**: 1261-1277.

Hollingsworth, M. A., Swanson, B. J. (2004). "Mucins in cancer: protection and control of the cell surface." *Nat. Rev. Cancer* **4**: 45-60.

Huynh, Q. K., Snell, E. E. (1986). "Histidine decarboxylase of *Lactobacillus 30a*. Hydroxylamine cleavage of the -seryl-seryl- bond at the activation site of prohistidine decarboxylase." *J. Biol. Chem.* **261**: 1521-1524.

Ishima, R., Torchia, D. A. (2000). "Protein dynamics from NMR." *Nat. Struct. Biol.* **7**: 740-743.

Iwai, K., Ando T. (1967). "N → O acyl rearrangement." *Methods Enzymol.* **11**: 263-282.

Keeler, J. (2005). "Understanding NMR spectroscopy." John Wiley & Sons Ltd, England.

Kelly, S. M., Jess, T. J. *et al*. (2005). "How to study proteins by circular dichroism." *Biochim. Biophys. Acta* **1751**: 119-139.

Kim, J. K., Yang, I. S. *et al*. (2003). "Crystal structures of glutaryl 7-aminocephalosporanic acid acylase: insight into autoproteolytic activation." *Biochemistry* **42**: 4084-4093.

Kim, J. K., Yang,I. S. *et al*. (2006). "Insight into autoproteolytic activation from the structure of cephalosporin acylase: a protein with two proteolytic chemistries." *Proc. Natl. Acad. Sci. USA* **103**: 1732-1737.

Kim, Y., Kim S. *et al*. (2002). "Precursor structure of cephalosporin acylase. Insights into autoproteolytic activation in a new N-terminal hydrolase family." *J. Biol. Chem.* **277**: 2823-2829.

Klabunde, T., Sharma S. *et al*. (1998). "Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing." *Nat. Struct. Biol.* **5**: 31-36.

Knowles, M. R., Boucher, R. C. (2002). "Mucus clearance as a primary innate defense mechanism for mammalian airways." *J. Clin. Invest.* **109**: 571-577.

Kraut, J. (1988). "How do enzymes work?" *Science* **242**: 533-540.

Lange, O. F., Lakomek, N. A. *et al*. (2008). "Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution." *Science* **320**: 1471-1475.

Levitin, F., Stern, O. *et al*. (2005). "The MUC1 SEA module is a self-cleaving domain." *J. Biol. Chem.* **280**: 33374-33386.

Ligtenberg, M. J., Kruijshaar, L. *et al*. (1992). "Cell-associated episialin is a complex containing two proteins derived from a common precursor." *J. Biol. Chem.* **267**: 6171-6177.

Lillehoj, E. P., Han, F. *et al*. (2003). "Mutagenesis of a Gly-Ser cleavage site in MUC1 inhibits ectodomain shedding." *Biochem. Biophys. Res. Commun.* **307**: 743-749.

Lindsay, C. D., Pain, R. H. (1990). "The folding and solution conformation of penicillin G acylase." *Eur. J. Biochem.* **192**: 133-141.

Lopez, X., Mujika, J. I. *et al*. (2003). "Alkaline hydrolysis of amide bonds: effect of bond twist and nitrogen pyramidalization." *J. Phys. Chem. A* **107**: 2304-2315.

Loria, J. P., Berlow, R. B. *et al*. (2008). "Characterization of enzyme motions by solution NMR relaxation dispersion." *Acc. Chem. Res.* **41**: 214-221.

Maeda, T., Inoue, M. *et al*. (2004). "Solution structure of the SEA domain from the murine homologue of ovarian cancer antigen CA125 (MUC16)." *J. Biol. Chem.* **279**: 13174-13182.

Milner-White, E. J. (1997). "The partial charge of the nitrogen atom in peptide bonds." *Protein Sci.* **6**: 2477-2482.

Nagi, A. D., Regan, L. (1997). "An inverse correlation between loop length and stability in a four-helix-bundle protein." *Folding & Design* **20**: 67-75.

Noren, C. J., Wang, J. *et al*. (2000). "Dissecting the chemistry of protein splicing and its applications." *Angew. Chem. Int. Ed.* **39**: 450-466.

Parry, S., Hanisch, F. G. *et al*. (2006). "N-glycosylation of the MUC1 mucin in epithelial cells and secretions." *Glycobiology* **16**:623-634.

Parry, S., Silverman, H. S. *et al*. (2001). "Identification of MUC1 proteolytic cleavage sites *in vivo*." *Biochem. Biophys. Res. Commun.* **283**: 715-720.

Pauling, L., Corey R. B. *et al*. (1951). "The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain." *Proc. Natl. Acad. Sci. USA* **37**: 205-211.

Paulus, H. (2000). "Protein splicing and related forms of protein autoprocessing." *Annu. Rev. Biochem.* **69**: 447-496.

Paulus, H. (2001). "Inteins as enzymes." *Bioorg. Chem.* **29**: 119-129.

Perler, F. B. (1998). "Breaking up is easy with esters." *Nat. Struct. Biol.* **5**: 249-252.

Perler, F. B., Xu, M. Q. *et al*. (1997). "Protein splicing and autoproteolysis mechanisms." *Curr. Opin. Chem. Biol.* **1**: 292-299.

Poland, B. W., Xu, M. Q. *et al*. (2000). "Structural insights into the protein splicing mechanism of PI-SceI." *J. Biol. Chem.* **275**: 16408-16413.

Porter, J. A., von Kessler, D. P. *et al*. (1995). "The product of hedgehog autoproteolytic cleavage active in local and long-range signalling." *Nature* **374**: 363-366.

Porter, J. A., Young K. E. *et al*. (1996). "Cholesterol modification of hedgehog signaling proteins in animal development." *Science* **274**: 255-259.

Radzicka, A., Wolfenden, R. (1996). "Rates of uncatalyzed peptide bond hydrolysis in neutral solution and the transition state affinities of proteases." *J. Am. Chem. Soc.* **118**: 6105-6109.

Romanelli, A., Shekhtman, A. *et al*. (2004). "Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the N-extein-intein junction." *Proc. Natl. Acad. Sci. USA* **101**: 6397-6402.

Rosenthaler, J., Guirard, B. M. *et al*. (1965). "Purification and properties of histidine decarboxylase from *Lactobacillus 30a*." *Proc. Natl. Acad. Sci. USA* **54**: 152-158.

Santoro, M. M., Bolen, D. W. (1988). "Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants." *Biochemistry* **27**: 8063-8068.

Schmitzberger, F., Kilkenny, M. L. *et al*. (2003). "Structural constraints on protein self-processing in L-aspartate-α-decarboxylase." *EMBO J.* **22**: 6193-6204.

Shemella, P., Pereira, B. *et al*. (2007). "Mechanism for intein C-terminal cleavage: a proposal from quantum mechanical calculations." *Biophys. J.* **92**: 847-853.

Somayaji, V., Brown, R. S. (1986). "Distorted amides as models for activated peptide N-C=O units produced during enzyme-catalyzed acyl transfer-reactions .1. The mechanism of hydrolysis of 3,4-dihydro-2-oxo-1,4-ethanoquinoline and 2,3,4,5-tetrahydro-2-oxo-1,5-ethanobenzazepine." *J. Org. Chem.* **51**: 2676-2686.

Tani, K., Stoltz, B. M. (2006). "Synthesis and structural analysis of 2-quinuclidonium tetrafluoroborate." *Nature* **441**: 731-734.

Taylor-Papadimitriou, J., Burchell, J. M. *et al*. (2002). "MUC1 and the immunobiology of cancer." *J. Mammary Gland Biol. Neoplasia* **7**: 209-221.

Terpe, K. (2003). "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems." *Appl. Microbiol. Biotechnol.* **60**: 523-533.

Terpe, K. (2006). "Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems." *Appl. Microbiol. Biotechnol.* **72**: 211-222.

Thathiah, A., Blobel, C. P. *et al*. (2003). "Tumor necrosis factor-α converting enzyme/ADAM 17 mediates MUC1 shedding." *J. Biol. Chem.* **278**: 3386-3394.

Tolbert, W. D., Zhang, Y. *et al*. (2003). "Mechanism of human *S*-adenosylmethionine decarboxylase proenzyme processing as revealed by the structure of the S68A mutant." *Biochemistry* **42**: 2386-2395.

Tsumoto, K., Ejima D. *et al*. (2003). "Practical considerations in refolding proteins from inclusion bodies." *Protein Expr. Purif.* **28**: 1-8.

van Poelje, P. D., Snell, E. E. (1990). "Pyruvoyl-dependent enzymes." *Annu. Rev. Biochem.* **59**: 29-59.

Wang, Q. P., Bennet, A. J. *et al.* (1991). "Distorted amides as models for activated peptide N-C(O) units 3. Synthesis, Hydrolytic profile, and molecular structure of 2,3,4,5-tetrahydro-2-oxo-1,5-propanobenzazepine." *J. Am. Chem. Soc.* **113**: 5757-5765

Wood, D. W., Derbyshire, V. *et al*. (2000). "Optimized single-step affinity purification with a self-cleaving intein applied to human acidic fibroblast growth factor." *Biotechnol. Prog.* **16**: 1055-1063.

Wood, D. W., Wu, W. *et al*. (1999). "A genetic system yields self-cleaving inteins for bioseparations." *Nat. Biotechnol.* **17**: 889-892.

Wreschner, D. H., McGuckin, M. A. *et al*. (2002). "Generation of ligand-receptor alliances by "SEA" module-mediated cleavage of membrane-associated mucin proteins." *Protein Sci.* **11**: 698-706.

Wüthrich, K. (1990). "Protein structure determination in solution by NMR spectroscopy." *J. Biol. Chem.* **265**: 22059-22062.

Xu, Q., Buckley, D. *et al*. (1999). "Structural insights into the mechanism of intramolecular proteolysis." *Cell* **98**: 651-661.

Yamagata, S., Snell, E. E. (1979). "Histidine decarboxylase from *Lactobacillus* 30a: reconstitution from separated subunits." *Biochemistry* **18**: 2964-2967.

Zhou, H. X. (2004). "Loops, linkages, rings, catenanes, cages, and crowders: entropy-based strategies for stabilizing proteins." *Acc. Chem. Res.* **37**: 123-130.