Understanding the Constraints of Sequence, Structure, and Energetics on Intramolecular Chaperone-Mediated Protein Folding

by

Ezhilkani Subbian

A DISSERTATION

Presented to the
Department of Biochemistry and Molecular Biology
and
The Oregon Health and Science University
School of Medicine

In partial fulfillment of the requirements for the degree
of
Doctor of Philosophy

March 9, 2005

School of Medicine
Oregon Health Sciences University

CERTIFICATE OF APPROVAL

This is certify that the Ph.D. thesis of

Ezhilkani Subbian

has been approved

[redacted]
Thesis mentor

[redacted]
Thesis committee chair

[redacted]
Member

[redacted]
Member

[redacted]
Member

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF TABLES

LIST OF ABBREVIATIONS

AAPF-pNA	Ala-Ala-Pro-Phe p-nitroanilide
AAPL-pNA	Ala-Ala-Pro-Leu p-nitroanilide
ANS	8-anilinonaphthalene-1-sulphonic acid
ATP	Adenosine triphosphate
BME	β-mercaptoethanol
C.elegans	*Caenorhabditis elegans*
CD	Circular Dichroism
CI2	Chymotrypsin Inhibitor 2
CPY	Carboxypeptidase Y
ΔCp	Change in heat capacity
ΔG	Change in free energy
ΔH	Enthalpy of a reaction
ΔS	Entropy of a reaction
E_a	Activation energy
EDTA	Ethylene diamine tetraacetic acid
EGTA	Ethyleneglycol-*bis*(β-aminoethyl)-N,N,N′,N′-tetraacetic acid
ER	Endoplasmic reticulum
ESP	Extracellular Serine Protease
GdnHCl	Guanidium hydrochloride
IMC	Intramolecular chaperone

ISP	Intracellular serine proteases
ISP1	Intracellular serine protease 1
ITC	Iso thermal calorimetry
K_a	Affinity constant
K_{eq}	Equilibrium constant
K_i	Inhibition constant
MMP	Matrix metalloproteases
NMR	Nuclear magenetic resonance
PAGE	Poly-acrylamide gel electrophoresis
PCs	Prohormone convertases
PMSF	Phenyl-methyl-sulphonyl-fluroide
PrA	Proteinase A
ProD	Denovo designed propeptide
Pro-PrA	Pro-Proteinase A
Pro-SbtE	Propeptide-Subtilisin E
ProWT	Wildtype propeptide of Subtilisin E
R	Universal gas constant
RDS	Rate-determining step
SbtBPN	Subtilisin BPN
SbtE	Subtilisin E
SDS	Sodium dodecyl sulphate
SSI	Streptomyces subtilisin inhibitor

T	Temperature
TCA	Trichloro acetic acid
TGN	Trans golgi network
TNP	Thermolysin-like neutral zinc metalloproteases
TPCK	L-(tosylamido-2-phenyl) ethyl chloromethyl ketone
UV	Ultraviolet
ω	Rate of non-synonymous to synonymous substitution

ACKNOWLEDGEMENTS

Two roads diverged in a wood, and I —

I took the one less traveled by,

And that has made all the difference.

The Road Not Taken

- Robert Frost

The following pages summarize the scientific backdrop for a very significant period in my life. It's been a very exciting, special and cherished chapter, thanks to a chosen few.

My greatest and most cherished find on this journey is my mentor Dr. Ujwal Shinde. An unbelievably wonderful human being and an excellent scientist, he has been so much more than a mentor. It is but beyond the language of words to describe his role in this work, or my gratitude for it. For all that it is worth – Thank you. I will uphold all that was taught!

During this time, I have had the chance to interact and learn from many excellent scientists. I would like to thank Yuki for his enormous scientific input and technical help. I am also grateful to everyone else who has helped me along: Vasanthi- for helping me get started in the lab, Martine- for technical support, Paul and Igor.

I have also been extremely fortunate with an excellent research advisory committee. My sincere thanks to Sveta (Dr. Lutsenko), Gary (Dr. Thomas), Dave (Dr. Farrens) and Hans Peter (Dr. Bächinger) for being amazing sources of help, ideas and inspiration. Thanks also to Tina, Guy and Kathleen for all their timely help.

While the science was cooking, I have learnt many a valuable lesson from my friends and family. Having just settled down among friends in a new country I had many a doubting thought when I moved to Portland. But am ecstatic that I did, becuase it has made all the difference! Cortny, Margie, Mihail, Abron, Lara, Niki, Martha J, Chris, Shirin, Aaron, Jan, Priscilla - have played their roles in their own unique ways. I am ever so grateful that my path crossed theirs. They have enormously reinforced my faith in human nature and how great it can be. I am also grateful that I got to discover graduate school together with a wonderful class of people, who have overwhelmed me with their friendliness. And, many a thanks to Arun, Archana, Malini, Divi and Sow for every little adventure that we went through and for all that is yet to come.

I was additionally fortunate to have found my second family, right here in Portland- Ujwal, Suneetha and Pushkar. Thanks for taking me into your beautiful circle of love.

But most importantly, none of this would be possible if not for four people - Appa, Amma, Akka and Priya. Thanks for letting go and yet being there, constantly, with your love. The intricacies of human connection are above and

beyond the most powerful of words. I only wish everyone experiences love and support the way I do!

The last five years have had their ups and their downs. But in the end it ended in a crest. Whoever (or whatever) pulls the strings, thanks a ton for working mine this way!!

ABSTRACT

Folding of polypeptides to their correct native states underlies efficient biological function. To ensure fidelity of the folding process, nature codes for chaperones that assist in the precise assembly of proteins. While most chaperones assist multiple substrates, several archaeal, prokaryotic, eukaryotic, and viral proteins, are synthesized with covalently attached propeptides that function as dedicated, single-turnover chaperones. Since the propeptides are indispensable for the folding of their cognate domains, and are part of the primary sequence of the polypeptide, such propeptides are referred to as "Intramolecular Chaperones" or IMCs. IMC-mediated folding pathways offer an ideal system to understand folding/unfolding mechanisms, to facilitate design of folding catalysts, and to explore how evolutionary selection has optimized the properties of a specific domain for folding advantages.

In this dissertation, the mechanism of IMC mediated protein folding was explored using Subtilisin E (SbtE), a classical bacterial serine protease, as a model. As a member of the ubiquitous family of proteases called subtilases, and being extremely well characterized both in structure and function, SbtE offers an ideal model to address the role of IMCs in protease folding, maturation, and regulation. Detailed analysis of SbtE and other subtilases demonstrates that the catalytic domains of the subtilase super-family are highly conserved, while their cognate IMCs vary significantly in their primary

sequences. Interestingly, most IMCs are also extremely charged, and appear to be unstructured or partially structured polypeptides. Further, biophysical studies demonstrate that the rate-determining step to protease maturation is not the IMC-mediated folding, but the release of the IMC from its inhibitory complex with the protease. Neither the significance of the above, nor the precise role of IMCs in protease maturation is completely understood. In this dissertation, some of the most fundamental questions about IMCs were addressed, namely; (i) How do IMCs facilitate efficient protease maturation? (ii) How do differences in charge, structure, and primary sequence within IMC relate to their chaperone and inhibitory functions? (iii) Why have specific proteins evolved dedicated IMCs to mediate folding pathways?

Our results demonstrate that the conformational dynamics of the IMC domain effectively regulate an inherently stochastic SbtE activation. Hence, apart from their role in folding, IMCs also appear to function as regulators of protease activation, and the absence of intrinsic structure within the isolated IMC-domain is critical for its function as a regulator. Furthermore, the chaperone function of the IMC is linked to its net charge and two conserved motifs, N1 and N2, while the overall $\alpha\beta$ scaffold of the IMC is critical for inhibition. Thus, IMC of SbtE is optimized to maintain the synergy of the protease maturation process and directs the protease to a kinetically trapped native state. Interestingly, Intracellular serine protease 1 (ISP1), an intracellular homologue of SbtE, is similar in sequence and structure, but folds to a thermodynamically stable state and independent of an IMC domain. Our results suggest that the thermodynamic versus kinetic constraints on the folding

pathways of the two homologues may have positively evolved to offer distinct functional advantages. Thus, it appears that not only the protein sequence and structure, but also its inherent folding pathway maybe optimized for functional advantages.

CHAPTER 1

INTRODUCTION

The question of how proteins find their unique native structures by using information encoded within their amino acid sequences lies at the heart of molecular biology. Understanding mechanisms of folding/unfolding are of immense scientific significance because of their potential in providing new approaches to combat protein-misfolding disorders and, in facilitating the design of novel proteins that catalyze desired reactions. In addition, the protein folding problem is ostensibly the most fundamental example of biological self-assembly, and symbolizes the first step towards tackling one of the most arduous questions addressable by contemporary science - *How have complex biological systems evolved the ability to self-assemble into robust and predictable assemblies?*

Based on the *in-vitro* refolding of ribonuclease (Anfinsen 1973), Anfinsen demonstrated that all information necessary for folding of a polypeptide to its native functional state is encoded within its primary sequence. While this largely holds true, polypeptides often have a tendency to mis-fold and aggregate, especially within the crowded macromolecular milieu *in vivo*. Interestingly, nature has evolved elaborate chaperone systems to ensure fidelity of protein

folding processes. Such chaperones assist the protein folding process from outside, and are not a part of the final folded state of the protein. Vast majority of them are non-specific and assist substrates to fold by preventing aggregation, through repetitive cycles of ATP dependent binding and release (Young, Agashe et al. 2004). In contrast to these non-specific chaperones, there are now several examples wherein certain polypeptides are synthesized with covalently attached propeptides that function as substrate-specific, single-turnover chaperones (Shinde and Inouye 1993; Shinde and Inouye 2000). How these unique propeptides facilitate folding, and why specific polypeptide sequences have co-evolved with dedicated chaperone domains, is a subject of intense research and the long-term goal for this thesis work.

1) INTRAMOLECULAR CHAPERONES (IMCs):

Inouye and coworkers, first demonstrated that SubtilisinE (SbtE), a bacterial alkaline serine protease from *Bacillus subtilis 168*, requires its propeptide to fold to its native functional state (Ikemura, Takagi et al. 1987). SbtE (275 residues) is an extracellular secreted protease that is synthesized as a pre-pro-protease with a 29-residue signal peptide and a 77-residue propeptide domain. Formation of active protease, in both *Eschericia coli* and *Bacillus subtilis,* is dependent on the presence of the propeptide domain (Ikemura, Takagi et al. 1987). While the denatured Propeptide-Subtilisin E (Pro-SbtE) effectively refolds *in-vitro* to produce active subtilisin, refolding of denatured SbtE in its isolated form failed to restore activity. Activity of SbtE was recovered only in the presence of the 77-residue propeptide or a Pro-SbtE inactive variant (Ikemura and Inouye 1988). These studies demonstrate that the presence of the propeptide is necessary and sufficient to obtain an active protease, and that the propeptide can assist folding both, *in cis* and *in trans*. Furthermore, the propeptide also functions as a slow-binding competitive inhibitor of SbtE (Li, Hu *et al.* 1995). Shorter fragments of the propeptide failed to inhibit protease activity and demonstrate that the propeptide requires the polypeptide beyond the substrate-binding pocket to function as a potent protease inhibitor. Interestingly, exogenously added serine protease inhibitors, such as Streptomyces Subtilisin Inhibitor (SSI) or Chymotrypsin inhibitor-2 (CI2) do not facilitate SbtE folding (Ikemura and Inouye 1988). This confirmed that, in addition to its role as an inhibitor, the propeptide actively functions as a dedicated chaperone for SbtE. Subsequent work has established that similar

propeptide mediated folding mechanisms exist within several different proteases including α-lytic protease (Silen and Agard 1989), carboxypeptidase Y (Sorenson, Winther et al. 1993), cathepsin L (Carmona, Dufour et al. 1996), aqualysin I (Lee, Miyata et al. 1991), proteinase A (van den Hazel, Kielland-Brandt et al. 1993), Kexin (Lesage, Prat et al. 2000), furin (Thomas 2002) and thermolysin (Marie-Claire, Roques et al. 1998). Since the propeptides are covalently attached to their cognate proteases and are essential for assisting folding of the catalytic domain, such propeptides are termed "Intramolecular Chaperones (IMCs)".

To distinguish propeptides from cellular chaperones, Inouye and coworkers, coined the term "Intramolecular chaperone (IMC)". Over the years, with the discovery of chaperone-like functions in the propeptides of various systems, "IMC" has become an accepted term. IMCs differ from classic cellular chaperones in a number of ways. Unlike the non-specific cellular chaperones, IMCs are highly substrate specific and are not regulated by ATP hydrolysis. While the cellular chaperones help refold polypeptides through iterative cycles of binding and refolding, IMCs are single-turnover catalysts that are degraded at the completion of folding. Further, most IMCs are effective inhibitors of the cognate catalytic domains and hence may additionally function downstream of the folding process as regulators of enzyme activity (Shinde and Inouye 2000; Young, Agashe et al. 2004).

Apart from IMCs that directly catalyze the folding process, there exist other propeptides that can indirectly assist in folding. For example, the propeptide of barnase interacts with the molecular chaperone GroEL and thus

ensures productive folding of barnase (Gray, Eder et al. 1993). Further, the propeptides in matrix metalloproteases (MMPs) contain a conserved cysteine residue the sulfhydryl group of which is coordinated by the catalytic Zn^{2+} ion, thus maintaining these proteases in a catalytically-inactive state. Proteolytic cleavage within the propeptide triggers a conformational change and releases shielding of the catalytic cleft in MMPs by interrupting the coordination between Zn^{2+} ions and cysteine residue (Morgunova, Tuuttila et al. 1999). Thus, based on their roles in protein folding, propeptides have been grouped into two major classes, Class I and II. Class I propeptides directly catalyze the folding, while Class II propeptides function in oligomerization, protein transport, localization, etc., and are indirectly involved in folding. The IMC of SbtE is a stereotypical Class I propeptide as it directly functions to catalyze the folding process (Shinde and Inouye 2000). In this dissertation the work refers to propeptides that directly function as IMCs, and the terms "IMC" and "propeptide" are used interchangeably.

A) EXAMPLES OF IMC-MEDIATED FOLDING PATHWAYS:

Initial studies demonstrated that IMC-mediated folding pathways exist in several secreted proteases of bacterial origin. However, subsequent work has established the existence of IMC-dependent folding in a variety of proteins that include both proteases and non-proteases from prokaryotes, eukaryotes, archaea and viruses. Some of the classic and emerging examples include the following:

1) α-lytic protease is a chymotrypsin-like serine protease that is secreted by the gram-negative soil-bacterium *Lysobacter enzymogenes* and serves to lyse and degrade microorganisms. α-lytic protease is secreted with a 166-residue propeptide and a 33-residue signal sequence. The 198-residue protease belongs to the same family as the mammalian digestive serine proteases, trypsin and chymotrypsin (Silen and Agard 1989). Several studies clearly establish that the propeptide functions as both a chaperone and an inhibitor of the protease, and ensures its folding to an active, secretion-competent, stable conformation. Interestingly, the eukaryotic homologues trypsin and chymotrypsin that display low sequence identity, but adopt similar three-dimensional scaffolds, can fold independent of an IMC domain (Baker, Sohl et al. 1992; Jaswal, Sohl et al. 2002).
2) Carboxypeptidase Y (CPY) from *Saccharomyces cerevisiae* is a serine carboxy-peptidase that is used extensively as a marker for protein transport and vacuolar sorting in yeasts. The protease, which is synthesized as a pre-pro-protease with a 91-residue propeptide, folds and cleaves its IMC in the endoplasmic reticulum (ER), resulting in an inhibited complex (Valls, Hunter et al. 1987). Upon its translocation to the yeast vacuole the enzyme is activated *in trans* by another serine protease, Proteinase A (Ammerer, Hunter et al. 1986; Ramos, Winther et al. 1994). Guanidine hydrochloride denatured pro-CPY can be rapidly and efficiently refolded by dilution into a suitable buffer. Under identical conditions, mature CPY, fails to refold to an enzymatically active form suggesting that the propeptide is required for correct folding of the mature protein (Winther and Sorensen 1991). Folding of mature CPY in the

absence of the propeptide results in the formation of a molten globule-like intermediate state, similar to that observed in the case of α-lytic protease and SbtE (Winther, Sorensen et al. 1994).

3) Proteinase A (PrA) from *Saccharomyces cerevisiae* is a vacuolar aspartic endo-proteinase (329 residues) that is vital for sporulation and viability during nitrogen starvation. The protease is secreted with a 54-residue propeptide that is proteolytically removed in the vacuole. Although the pro-PrA has been difficult to purify, indirect studies suggest that the propeptide functions to directly assist in protease folding (van den Hazel, Kielland-Brandt et al. 1993).
4) Procathepsin-L, a member of the large family of cysteine proteinases, was the first example of propeptide-assisted folding in this family of enzymes (Smith and Gottesman 1989). Studies demonstrate that loss of protease activity is directly proportional to truncations within the propeptide domain, and that the complete cognate propeptide is required for correctly folded cathepsin-L (Ogino, Kaji et al. 1999). Subsequently, it was established that cathepsin-S and cathepsin-B also required the presence of their cognate propeptides for productive folding (Wiederanders, Kaulmann et al. 2003). Folding of cathepsin-S under varying conditions of pH, time, redox state and ionic strength did not compensate for the loss of propeptide (Ogino, Kaji et al. 1999). Similar to the subtilisin family (Shinde, Liu et al. 1997), studies on cathepsins demonstrate that mutations in the propeptide directly affect protease function. This is highlighted by the hereditary disease pycnodysostosis, caused by "loss-of-function" mutations in the cathepsin-K

gene, one of which is directly localized in the propeptide domain (Hou, Bromme et al. 1999).

5) Elastase, an important virulence factor in the opportunistic pathogen *Pseudomonas aeruginosa,* is a thermolysin-like neutral zinc metalloprotease (TNP) that is synthesized with an amino terminal propeptide (174 residues). Elastase was the first TNP family member demonstrated to require it's propeptide for both folding and secretion (McIver, Kessler et al. 2004). Subsequent studies on other TNPs such as thermolysin, the prototype of this family of proteases, and pro-aminopeptidase processing protease (PA protease), demonstrated that their N-terminal propeptides function as IMCs as well (Tang, Nirasawa et al. 2003). Analysis of propeptides of TNPs demonstrates the presence of two conserved regions within the propeptide domains that may be critical for function. Mutations within the two conserved regions directly affect the chaperone function (McIver, Kessler et al. 2004). Interestingly, the propeptide of vibriolysin, another TNP, has been shown to chaperone the folding of PA-protease although they share only 36% sequence identity.

6) Apart from proteases, growth factors and neuropeptides such as transforming growth factor-β1, activin A (Gray, Eder et al. 1993), nerve growth factor and amphiregulin (Suter, Heymach et al. 1991; Thorne and Plowman 1994), hormones such as insulin (Steiner 2004), certain glycoproteins like von Willebrand factor (Voorberg, Fontijn et al. 1993), and bacterial pancreatic trypsin inhibitor (BPTI) (Weissman and Kim 1992) also depend on propeptides for folding assistance.

The above examples effectively highlight the wide scaffolds that fold using the assistance of IMC domains. To date, IMCs have been identified in all four major classes of proteases: serine, cysteine, aspartyl and metalloproteases. While the bacterial proteases are mostly secreted extracellularly, the eukaryotic proteases function mostly in sub-cellular compartments of extreme pH (Baker, Shiau *et al.* 1993; Shinde and Inouye 1993).

Why do specific sequences require IMCs to fold to their native states? Are there specific structural, functional, or sequence signatures that may necessitate the presence of IMC domains? Do IMCs employ a common mechanism to assist the folding of these varied scaffolds? Answering these questions would have enormous implications for the fundamental understanding of protein folding in general and regulation of cellular proteases in particular, and would also facilitate the rational design of protein specific chaperones. To address these fundamental questions in IMC-mediated folding, we use Subtilisin E, a member of a large super-family of subtilisin-like serine proteases (subtilases), as a model.

II) SUBTILASES - AS A FOLDING MODEL:

Subtilases are an ancient group of enzymes that are widely dispersed through evolution. Although functionally diverse, majority of these enzymes are synthesized as pre-pro-proteases, secreted extracellularly or within the secretory transport system, and are activated by cleavage of the propeptide. The catalytic domain of this super-family is characterized by an invariant Asp-, His-, Ser- catalytic triad, and a distinct α-β scaffold (Fig. 1.1 and Fig. 1.2). This subtilase super-family now includes more than 800 homologues encompassing archaea, prokaryota, eukaryota and even viruses (Siezen and Leunissen 1997; Rawlings, Tolle et al. 2004).

A) Subtilases are widespread in nature:

Subtilases constitute the S8 family in the Merops protease database and can be divided into two subfamilies: degenerative subtilases identified mainly in prokaryotes (S8A family; prototype: subtilisin BPN from *Bacillus amyloliquefaciens*) and processing proteases mainly from eukaryotes (S8B family; prototype: kexin)(Rawlings, Tolle et al. 2004). Nine subtilases have been characterized in mammals, seven of which belong to the S8B subfamily, the pro-protein convertases (PCs) (Steiner 1998; Seidah and Chretien 1999). More recently, two mammalian subtilases were identified within the S8A subfamily (Rawlings, Tolle et al. 2004).

A.1. Degradative Subtilisins:

A.1.1 Bacterial subtilisins:

These are normally secreted extracellularly and function under extreme conditions. They are implicated to function during bacterial sporulation and general protein turnover (Ferrari, Howard et al. 1986). The prototype of this family of enzymes is SbtBPN from *Bacillus amyloliquefaciens* (Bott, Ultsch et al. 1988). While the majority of bacterial subtilisins are secreted as pro-proteases (~100 residue propeptide-domain and 260-1700 residues catalytic domain), a few are devoid of distinct propeptides and are found to be intracellular (Strongin, Izotova et al. 1978; Siezen and Leunissen 1997). Due to their broad substrate specificity and the ability to function at alkaline pHs, bacterial subtilases have found wide applicability in detergents, leather industry and waste management (Gupta, Beg et al. 2002) (See Chapter1- Section II,C).

A.1.2. Plant subtilisins:

Plant subtilases, together with S1P and SKI-1, belong to the pyrolysin group within the S8A subfamily. Despite their prevalence, current understanding of subtilase function in plants is still very limited. Plant subtilases have been implicated in general protein turnover as well as highly specific regulation of plant development, or responses to environmental challenges (Berger and Altmann 2000; Janzik, Macheroux et al. 2000). The primary example for plant subtilases that function in protein degradation is cucumisin, which constitutes up to 10% of the soluble protein in melon fruit. The catalytic properties of this enzyme have been extensively documented and it was the first subtilase to be cloned from any plant species (Kaneda and Tominaga 1975; Yamagata,

Masuzawa et al. 1994). The tomato subtilase P69, initially identified as a pathogenesis-related protein, was shown to be one of the several subtilases that are specifically induced following pathogen infection (Tornero, Conejero et al. 1996; Meichtry, Amrhein et al. 1999). Since then, several other plant subtilisins have been identified and characterized, and most of these enzymes are highly abundant and exhibit broad substrate specificity (Rawlings, Tolle et al. 2004).

A.1.3. Subtilisins in parasites:

The sequencing of the malaria genome helped identify three subtilisin-like proteases that are crucial for malarial infection: *Plasmodium falciparum* SUB-1 (PfSUB-1), *Plasmodium falicparum* SUB-2 (PfSUB-2) and *Plasmodium falciparum* SUB-3 (PfSUB-3). Pf-SUB-1 has been extensively studied and appears to be processed to a propeptide-protease complex prior to activation, similar to other known subtilases. Further, the propeptide of this protease is a tight-binding inhibitor (K_i ~ 5.5 nM) (Sajid, Withers-Martinez et al. 2000; Jean, Hackett et al. 2003). Although the specific roles of Pf-SUB1, PfSUB-2, and PfSUB-3 in malarial parasite invasion are yet to be elucidated, these enzymes are currently being investigated as probable drug targets (Withers-Martinez, Jean et al. 2004).

Sensitivity of *Toxoplasma gondii* tachozyte invasion to serine protease inhibitors led to the identification of two subtilisin-like serine protease, Tg-SUB1 and Tg-SUB2 (Miller, Binder et al. 2001; Miller, Thathy et al. 2003). Furthermore, a subtilisin-like protease, NcSUB1, has been purified from

Neospora canium (Louie and Conrad 1999) and another unique member AhSUB has been identified in *Acanthamoeba healyi.* Interestingly, AhSUB that has been implicated to play a role in ameobic keratitis and encephalitis, displays a 67% sequence similarity with the thermostable subtilase aqualysin (Hong, Kong et al. 2000). Further studies on these proteases will help highlight the subtleties between the different subtilases, and may facilitate rationale design of potential drugs to target pathogens.

A.2. Regulatory Subtilisins:

A.2.1. Kexin family - Prohormone convertases (PCs):

The Kexin/Subtilisin-like eukaryotic proteases (prohormone convertases) are a family of subtilases that to date, have been identified in all eukaryotes. PCs are required for activation of a number of proteins including growth factors and their receptors, proneuropeptides, pro-hormones, adhesion molecules, plasma proteins, viral glycoproteins and bacterial toxins (Steiner 1998; Seidah and Prat 2002; Thomas 2002).

In mammals, seven members of the family have been identified: furin, PC1/3, PC2, PC4, PC5/6, PC7/8/LPC and PACE4. All members of the family are active within the secretory pathway and are transported into the ER through the presence of a signal peptide. Furin, PC5, PC-7, PC4 and PACE-4 are involved in the processing of proteins secreted via the constitutive pathway. In contrast, PC1 and PC2 are found in the core secretory granules and function in the regulated secretory pathway.

The N-terminal propeptides of PCs vary from 80-115 residues and are shown to function as chaperones and inhibitors of their cognate proteases (Rockwell and Thorner 2004). Similar to other subtilases, the propeptide is cleaved by a two-step maturation event. After the first autocatalytic cleavage, the propeptide remains attached to the protease as an inhibitor. The release of the propeptide occurs in the TGN where a decrease in the pH causes a second auto-proteolytic cleavage (Anderson, Molloy et al. 2002). The catalytic domain of PCs is a conserved subtilisin like scaffold with the Asp-His-Ser catalytic triad (Fig. 1.1) (Henrich, Cameron et al. 2003; Holyoak, Wilson et al. 2003). However, unlike the bacterial proteases, PCs preferentially cleave after dibasic residues or an R-X-X-K/R-R motif. As discussed later, specific residues that line the pockets of substrate binding sites S1 and S2 define the specificity (See Chapter1- Section II, B) (Henrich, Lindberg et al. 2005). Following the catalytic domain, the PCs have a conserved 150 residue P-domain (Fig. 1.1). The P-domain appears necessary for folding and maintaining the catalytic domain, and to regulate its calcium- and pH- dependence (Zhou, Martin et al. 1998). In addition, the middle of the P-domain in most PCs contains the cognate integrin-binding RGD sequence that may be required for intracellular compartmentalization and maintenance of enzyme stability within the ER (Rovere, Luis et al. 1999). The carboxy terminal tail provides uniqueness to each PC family member and is the least conserved region in all convertases (Rockwell and Thorner 2004). Of all known PCs, furin has been extensively studied in terms of structure, function, folding, and regulated activation (Thomas 2002). It has been established that the folding of furin in the absence of its

propeptide domain results in ER retention and inactivity (Anderson, Molloy et al. 2002).

The role of the PCs in processing of essential growth factors, adhesion molecules, MMPs and integrins makes them attractive targets for treatment of malignancies and tumors (Bassi, Lopez De Cicco et al. 2001; Khatib, Siegfried et al. 2002). In particular, the inhibitive properties of the propeptides have been tested for potential anti-cancer therapy (Khatib, Siegfried et al. 2002). The propeptide of furin and other PCs were shown to inhibit processing of growth factor precursors such as nerve growth factor and brain-derived neurotrophic growth factor, both of which are involved in cancer and metastasis. However, since the propeptides are not always specific for their cognate enzymes, it will be necessary to modify their structure to improve their selectivity and potency (Zhong, Munzer et al. 1999).

In yeast, there is a single gene that codes for kexin, the first PC to be characterized. Kexin is localized in the TGN and is required for the generation of mature α-factor and killer toxin (Julius, Brake et al. 1984). The recently solved crystal structure of furin and kexin offer interesting insights into the substrate specificity of this family of enzymes (Henrich, Lindberg et al. 2005).

The *C.elegans* genome contains a complement of four genes that comprise the subtilisin like PC family. However, these are predicted to produce at least 14 convertase proteins through alternative splicing events. They possess the overall PC like structure, but their precise functional roles are not known (Thacker and Rose 2000).

Subtilases similar to the mammalian PCs have also been isolated from *Drosophila melanogaster.* Dfur1 and Dfur2 are two furin-like genes that code for multiple subtilisin-like serine proteases. While they display similar catalytic specificities, these proteases differ in their biosynthetic fates from the mammalian homologues (De Bie, Savaria et al. 1995). A serine protease similar to PC2 was also isolated from the sheep blowfly, *Lucilia cuprina* (Mentrup, Londershausen et al. 1999).

B) Common fold and catalytic mechanism of subtilases:

Since the first crystal structure of Subtilisin BPN (SbtBPN) from *Bacillus amyloliquefaciens*, several well-resolved structures of subtilases (Bott, Ultsch et al. 1988), alone and with their peptide and protein inhibitors, have been solved (Siezen and Leunissen 1997). These structures establish a well-conserved catalytic domain in subtilases that is comprised of three layers, a seven-stranded β-sheet sandwiched between two layers of helices (Fig. 1.1). Between individual subtilisins, there are subtle differences in surface loops, disulphide bridges, and in calcium-binding sites. Structural analysis of the subtilase family shows the presence of up to five calcium-binding sites in most known subtilases. For example, SbtE has two calcium ions, while the thermostable subtilases Bacillus Ak.1 protease and sphericase have four and five calcium ions respectively (Smith, Toogood et al. 1999; Almog, Gonzalez et al. 2003). These calcium ions are often bound to surface loops and have been shown to enhance thermostability and proteolytic resistance (Bryan 2000). Apart from

higher number of bound metal ions (Fig. 1.1), the thermostable subtilisins have significantly more aromatic residues and lesser-exposed hydrophobic surface area (Smith, Toogood et al. 1999). In addition to their importance in maintaining structure and stability, calcium ions also appear to modulate enzyme specificity and activity. For example, the recently solved crystal structures of furin and kex 2 establishes that these eukaryotic subtilisin-like proteases have three bound calcium ions, one of which is bound at the base of the S1 site and may play a role in substrate recognition (Henrich, Cameron et al. 2003; Holyoak, Wilson et al. 2003).

Subtilisins follow a two-step reaction for hydrolysis involving the Asp-, His- and Ser- catalytic triad (Fig. 1.2; Fig. 1.3). The Ser provides the hydroxyl nucleophile for attack of the scissile bond, while His serves as the catalytic base for deprotonation of Ser, and the Asp serves to stabilize the positive charges on the His. Further, an Asn serves as an "oxyanion hole" and stabilizes the buildup of negative charges on the carbonyl oxygen. Once a suitable substrate is bound, the hydroxyl moiety of the catalytic Ser attacks the scissile bond. This acylation step results in the release of the peptide bond C-terminal to the cleavage site and the polypeptide remains bound to the Ser through an ester linkage. Subsequent deacylation mediated through a water molecule results in the release of the N-terminal product (A.Fersht 1985). Although all subtilases follow the same mechanism for substrate hydrolysis, the rate of acylation and deacylation may differ between individual enzymes (Philipp and Bender 1983; Rockwell and Fuller 2001).

Most known subtilases are broad specificity proteases and their specificity is largely defined by residues at S1 and S4 in the substrate-binding pocket. Whereas in the bacterial enzymes these are lined by large hydrophobic groups (Mei, Liaw et al. 1998), in the eukaryotic kexin and PCs, the S1 and S4 sites are lined by negatively charged residues resulting in specificity for cleavage after Lys/Arg residues (Henrich, Cameron et al. 2003). These enzymes have optimal activity in the mild-alkali pH and are inhibited by general serine peptidase inhibitors such as Diisopropyl-fluorophosphate (DFP) and Phenyl-methyl-sulphonyl-fluoride (PMSF) (Rawlings, Tolle et al. 2004). Additionally, since many members of the family bind calcium for stability, inhibition can also be seen with EDTA and EGTA (Yabuta, Subbian et al. 2002). Protein inhibitors of the subtilisin family include turkey ovomucoid third domain (Ardelt and Laskowski 1985), *Streptomyces* subtilisin inhibitor (Takeuchi, Satow et al. 1991), eglin C (Fei, Luo et al. 2001) and barley inhibitor CI-1A (Radisky, Kwan et al. 2004), many of which also inhibit chymotrypsin. Further, the subtilisin propeptide is a potent inhibitor of its cognate protease domain (Li, Hu et al. 1995).

C) Industrial applications of subtilases:

Microbial proteases constitute ~40% of the total enzyme sales in industrial sectors, the largest share of which has been held by subtilisin-like alkaline proteases (Gupta, Beg et al. 2002). Their high stability, broad substrate specificity, low molecular mass, and activity under alkaline conditions make subtilases attractive industrial targets (Kumar and Takagi 1999). For example,

the ability to facilitate the removal of proteinaceous stains makes subtilases useful additives in household detergents and cleaning solutions(Gupta, Beg et al. 2002). They are also widely used in the food and feed industry for preparation of protein hydrolysates of high nutritional value (Nekliudov, Ivankin et al. 2000). For example, *B.subtilis* proteases are used in the production of fish hydrolysates of high nutritional value, for cheese whey hydrolysates, and to convert lean meat waste into edible products (Perea, Ugalde et al. 1993; Kristinsson and Rasco 2000). Subtilases also play a major role in the waste management industry. They are used to lower the biological oxygen demand of aquatic systems, food-processing industries and poultry (Ichida, Krizova et al. 2001; Gupta, Beg et al. 2002). Further, subtilisins are extensively used for soaking, dehairing and bating stages of preparing skins and hides in the leather industry (Loperena, Ferrari et al. 1994; Pan, Huang et al. 2004), and for the treatment of burns, abscesses and wounds in the medical industry (Lobenko, Burov et al. 1991). Subtilases have also found applications in the photographic industry where they are extensively used for recovery of silver from photographic films (Masui, Yasuda et al. 2004). The widespread use of subtilases in industry hence makes them attractive targets for protein engineering. Rational design and evolutionary engineering approaches have enabled the creation of subtilisin variants with desired enzyme properties such has increased thermal stability (Takagi, Takahashi et al. 1990) and higher catalytic efficiency (Takagi, Morinaga et al. 1988). There is a wealth of information regarding protein sequences, tertiary structures and molecular functions of subtilases. Their ubiquitous distribution from viruses to mammals,

and the scope of their applications from waste management to cheese curing, makes subtilases attractive models to decipher relation between primary sequences, protein structures, functional evolution, and assisted versus unassisted folding pathways.

III) UNDERSTANDING THE FUNDAMENTALS OF IMC-MEDIATED MATURATION:

Over the past two decades, numerous genetic, biochemical, and structural analyses of subtilases have provided insights into the mechanism of IMC-mediated protein folding. While different systems have added complementary information towards deciphering the mechanism, SbtE and SbtBPN remain the classical subtilase models for these studies (Shinde and Inouye 1993). Additionally, extensive studies on α-lytic protease, a chymotrypsin-like serine protease from *Lysobacter enzymogenes* has provided valuable insights to understanding the overall mechanism (Jaswal, Sohl et al. 2002). Structural data on the above systems offer interesting snapshots into the gradual transition of the polypeptide from an unfolded state, through an inhibited complex, to an active protease. Complementary biophysical and biochemical studies have helped to elaborate reasons for non-productive folding of the isolated protease-domains, and to elucidate a general mechanism for how IMCs may function in this pathway.

A) Crystal structure of SbtE AND Pro-SbtE complex

Several high-resolution crystal structures of both SbtE and SbtE in complex with its IMC domain have been solved (Gallagher, Gilliland et al. 1995; Jain, Shinde et al. 1998; Radisky, King et al. 2003). The structure of SbtE (Fig. 1.3a) is comprised of three β-sheets and nine α-helices, with Asp32-His64-Ser221 forming the catalytic triad. The largest β sheet is comprised of seven, parallel, β-strands and is flanked on one side by three helices, and on the other

by two. The substrate-binding site (Fig. 1.3b) is a surface channel that accommodates 6 residues (P4-P2'). In SbtE, the substrate-binding pocket is large and hydrophobic made of main chain residues from Ser125-Leu126-Gly127, and main and side chain residues of Ala152-Ala153-Gly154 and Gly166. The Gly166 is at the bottom of the pocket for P1 and is critical for specificity. The P1-P4 substrate backbone forms the central β-strand in an anti-parallel β-sheet with the protease residues 100-102 and 125-127. Further, SbtE displays two calcium binding sites (Fig. 1.3a), a high affinity site that is well conserved (A-site) in most subtilases, and a weak affinity site that is less conserved (B-site) (Siezen and Leunissen 1997). Calcium at A-site is coordinated in pentagonal-bipyramidal geometry by the loop comprised of residues 75-83, Gln2 at the N-terminus, and an Asp at 41. The seven coordination distances range from 2.3-2.6 Å with the Asp being the closest (Fig. 1.3c). The second calcium makes contact with the main-chain carbonyl oxygen atoms of residues 169,171 and 174 in a shallow crevice near the surface of the molecule and is coordinated in a distorted pentagonal-bipyramid (Alexander, Ruan et al. 2001). These two calcium-binding sites together make SbtE an extremely stable protease in the absence of any cysteines, or stabilizing disulphides in its structure (Yabuta, Subbian et al. 2002). The lack of cysteine residues in SbtE is advantageous because it allows the probing of specific interactions during the folding process. This approach has been used to identify a precise non-native interaction by engineering two cysteine residues, which are distal in the native protein but are proximal during folding, and form a

specific intramolecular molecular disulfide bond under oxidative folding conditions (See Chapter1- Section III, B.2.2) (Inouye, Fu et al. 2001).

Engineering a $S_{221}C$ substitution at the active site of Pro-SbtE blocks the maturation process subsequent to the first proteolytic cleavage of the IMC. This variant facilitates the isolation of stable, cleaved stoichiometric Pro:S_{221}C-SbtE complex, whose X-ray structure (Fig. 1.4a) has been solved (Bryan, Wang et al. 1995; Gallagher, Gilliland et al. 1995; Jain, Shinde et al. 1998). The structure of the mature domain in the complex is super imposable with the structure of the isolated protease domain with a root mean square deviation of 0.46 Å, when the Cα atoms were compared. While the isolated IMC is largely unstructured, the IMC in complex with the protease folds into a single domain with a four stranded anti-parallel β sheet and two three-turn helices, forming an α+β plait. The structured, inhibitory, IMC packs against two surface helices (α1 and α2; Fig.1.3a) of the protease domain formed by residues Tyr104-Asn117 and Ser132-Ser145. Further, residues -1 to -7 from the IMC domain directly interact with the substrate-binding region to complete a three-stranded β-sheet with β-strands from the protease domain. In all, there are 27 hydrogen bonds (Fig. 1.4a) at the IMC-subtilisin interface. Interestingly, 24 hydrogen bonds stabilize the interaction of residues -1 to -9 with the active site and with the substrate-binding regions. These include the three hydrogen bonds that stabilize the backbone amide groups of Glu(-9) and Asp(-7), that form helix caps for the two SbtE interaction helices (α1 and α2). In contrast, the remaining 68 residues of the propeptide are stabilized by only three hydrogen bonds, between the backbone amides of residues -34 to -36 and the carboxylate group of the

Glu112 from the protease domain. The significance of this asymmetric distribution of hydrogen bonds at the IMC: protease interface is unknown. Another noteworthy point is that while the average B-factor of the side-chain and the main chain atoms is 15.0 and 13.6 $Å^2$ in the protease, it is 35.5 and 33.8 $Å^2$ respectively in the IMC (Jain, Shinde et al. 1998). This approximately two-fold increase in the B-factors within the IMC domain may be biologically significant and necessary for function.

An interesting insight from the structure of the Pro-S_{221}C-SbtE was that the N-terminus of the protease is more than 20 Å away from the active site. Earlier biochemical studies had clearly established that the processing of IMC was intramolecular (Li and Inouye 1996). Based on this, the structure of Pro-SbtE (Fig. 1.4b) just prior to cleavage was modeled with the N-terminus of the protease bound to the C-terminus of the IMC at the active site. The lack of a preferred conformation in N-terminal residues (Residues 1-6) of SbtE and the recovery of activity in a H_{64}A-SbtE active site mutant with a $E_{(-2)}$H substitution in IMC, through substrate assisted catalysis, substantiated the model (Shinde and Inouye 1995). A recent crystal structure of an active-site mutant of a subtilisin homologue, pro-kumamolysin, shows the propeptide bound to the protease prior to cleavage, and confirms this model proposed for the uncleaved precursor (Comellas-Bigler, Maskos et al. 2004).

B) Mechanism of IMC-mediated structural acquisition:

B.1. Why do polypeptide sequences require IMCs?

B.1.1 Kinetic and Thermodynamic characterization :

To establish how IMCs function in the folding pathway, folding of α-lytic protease and SbtBPN were compared in the presence and absence of their IMC (Baker, Sohl et al. 1992; Eder, Rheinnecker et al. 1993). Folding of SbtBPN in the absence of its IMC resulted in formation of a non-functional, structured state that is stable for weeks. This intermediate displayed a hydrodynamic volume intermediate between that of the fully folded and fully unfolded protease. Circular dichroism spectra of the intermediate in the far UV-region (190-250nm) corresponded to a well-defined secondary structure with a minimum at 208 and 222nm. However, in the near UV-region (250-320nm) the intermediate displayed no amplitude, suggesting a lack of well-formed tertiary packing. This was also confirmed through an NMR study that established a strongly reduced dispersion in the amide and methyl regions of the H^1 NMR spectrum compared with the fully folded protease. A noteworthy point, however, was that the intermediate appeared to bind calcium with a stoichiometry of 1, but with an affinity intermediate to affinities of A-site and B-site (Eder, Rheinnecker et al. 1993) (See Chapter1- Section III, A). Similar behavior was also observed with the intermediate state of α-lytic protease folded in absence of its IMC (Baker, Shiau et al. 1993). Although the intermediate state was extremely stable for weeks, addition of cognate IMCs yielded active native protease.

Hence, these studies suggested that in the absence of IMC, the protease folds to a kinetically-trapped state with properties of a classical "molten-globule" intermediate (Kuwajima 1989). A high-energy barrier between the molten-globule intermediate and transition state, limits the folding to a native state (Fig.

1.5). Addition of the IMC lowers this barrier to enable productive folding (Baker, Sohl et al. 1992; Bryan, Alexander et al. 1992; Eder, Rheinnecker et al. 1993; Eder, Rheinnecker et al. 1993). Based on these observations, it was established that *the IMC functions to overcome a kinetic barrier on the folding pathway, and that the observed intermediate is either on-pathway, or in equilibrium with a conformation on the folding pathway.*

To establish the relevance of the observed intermediate to the biological *in-cis* folding, refolding of full-length Pro-SbtE and Pro-SbtBPN were analyzed (Eder, Rheinnecker et al. 1993; Sohl, Jaswal et al. 1998). To avoid complications of proteolysis and to inhibit IMC processing, both studies were done using a Ser_{221}Ala active site variant that has six orders of magnitude lower proteolytic activity. This variant represents the propeptide:protease complex just prior to cleavage as discussed earlier (Fig. 1.4b; See Chapter1- Section III, A). The folded, but uncleaved, Pro-S_{221}A-SbtBPN binds calcium ions and adopts a compact conformation with an apparent molecular weight of 36kD. Equilibrium unfolding of fully folded Pro-S_{221}A-SbtBPN monitored through changes in circular dichroism, and fluorescence spectroscopy, followed a three-state unfolding curve. Most of the tertiary and a part of the secondary structure unfolded through an initial cooperative process while separated secondary structures followed a less cooperative second transition. Interestingly, the second unfolding transition was similar to the unfolding transition of the molten-globule intermediate formed in the absence of IMC. Additionally, at higher denaturant concentrations Pro-S_{221}A-SbtBPN displayed properties similar to the trapped molten-globule intermediate. These results, taken together, established

that the equilibrium unfolding of the Pro-S_{221}A-SbtBPN occurs via an intermediate that is similar to the kinetically trapped molten-globule intermediate. Further, three-state equilibrium-unfolding transition suggested that the polypeptide folds to a molten-globule like state and then transitions to the native state, through the assistance of the IMC. Thus it was argued that *the IMC functions only during the late stages of the folding pathway*. Another interesting finding was that the relative thermodynamic stability of the fully folded Pro-SbtBPN complex was only marginally higher than that of the SbtBPN intermediate (Eder, Rheinnecker et al. 1993; Sohl, Jaswal et al. 1998). The thermodynamic stability was also strongly dependent on salt concentrations, which was probably an effect of the highly charged propeptide and its effect on the electrostatic interactions of the solvent environment (Eder and Fersht 1995).

Most of the initial folding studies highlighted above employed slow dialysis for refolding and hence were not amenable to kinetic analysis (Ikemura and Inouye 1988). Subsequent optimization of a fast refolding that involves the rapid dilution of unfolded protein into denaturant free buffer, paved the way for kinetic studies of IMC-mediated folding. Using the technique of refolding by rapid dilution, Eder and Fersht (Eder, Rheinnecker et al. 1993) established the kinetics of refolding of Pro-S_{221}A-SbtBPN by monitoring the increase in intrinsic fluorescence. Pro-S_{221}A-SbtBPN follows two-state kinetics with a rapid phase (65% amplitude as native) and slow phase (35% amplitude as native). While the rapid phase reflects the acquisition of structure in the intermediate, the slow phase established the kinetics of folding of the intermediate in the presence of the IMC. This slow phase followed a rate constant of $0.0047s^{-1}$. Initial studies

had established a folding rate of $<1.4\times10^{-8}$ s^{-1} for SbtBPN refolded in absence of the IMC. Hence, taken together, this demonstrated that the IMC accelerates the kinetics of protease folding by atleast five orders of magnitude. Based on Eqn.1.1,

$$\Delta\Delta G^{*} = -RT\ln(\frac{k_{Pro\text{-}Sbt}}{k_{Sbt}}) \qquad \text{Equation (1.1)}$$

the difference in transition state energies for folding with and without the IMC was estimated to be >7.5 kcal/mol (Fig. 1.5) (Eder and Fersht 1995). Similar studies with α-lytic protease established that the propeptide lowers the kinetic barrier to folding by 18kcal/mol. With α-lytic protease, the relative thermodynamic stability of the native state relative to the unfolded and intermediate states was also established. Interestingly, these studies demonstrate that the native state is ~1kcal/mol less stable than the unfolded state (Baker, Sohl et al. 1992; Sohl, Jaswal et al. 1998). Thus the protease appears to be in a kinetically trapped, thermodynamically unstable native state. Hydrogen-exchange experiments establish that this state has extremely low conformational dynamics with >50% of the residues having a protection factor (P_f) $>10^4$ (Jaswal, Sohl et al. 2002).

B.1.2. Calcium deletion variant:

Both SbtE and SbtBPN have two well-defined calcium-binding sites, a high-affinity A-site and medium affinity B-site (See Chapter1-Section III, A; Fig. 1.3a). Studies establish that calcium binds to the A-site (Fig. 1.3c) with an affinity of $\sim10^7$ M^{-1} and contributes significantly to the thermodynamic stability of

the protease (Alexander, Ruan et al. 2001). What is particularly interesting is that the kinetic barrier to calcium dissociation is extremely high (~23kcal/mol) and is higher than the binding free energy (Alexander, Ruan et al. 2001). This suggests that the binding of calcium and the accompanied lowering of conformational entropy might limit IMC independent folding. To establish this, Bryan and coworkers created a variant ($SbtBPN_{\Delta cal}$) that had the calcium-binding loop (A-site: 75-83) removed. Upon refolding of $SbtBPN_{\Delta cal}$ under low ionic conditions, no activity was observed. However, folding of $SbtBPN_{\Delta cal}$ under high ionic conditions resulted in independent folding of $SbtBPN_{\Delta cal}$ to active, native state (Bryan, Alexander et al. 1992). Folded $SbtBPN_{\Delta cal}$ is unstable and further mutations, including a disulphide bond, have to be introduced to enhance stability. However, the new variant has a structure similar to the wildtype protease except in the region of the A-site (Almog, Gallagher et al. 1998). This suggests that the calcium binding A-site may be a critical factor that dictates the requirement for an IMC. However, it is interesting to note that productive folding is seen only under conditions of high ionic strength and that the calcium independent protease variant is extremely unstable.

Thus, a high-energy barrier that separates the unfolded and native states limits the spontaneous folding of specific proteases. IMCs assist in the folding pathway by lowering these barriers to enable productive folding. Selection of such kinetic barriers on the folding pathway may provide a mechanism for evolution of optimal functional properties. Since most IMC-dependent proteases appear to function in harsh protease-rich conditions the presence of the IMC may enhance longevity through high unfolding energy barriers.

B.2. How do IMCs assist in folding?

B.2.1. Stabilization of folding nucleus–"Side-on Model":

The 77-residue IMC domain is an intrinsically unstructured polypeptide that folds to an α-β conformation in the presence of the protease (Buevich, Shinde et al. 2001). Crystal structure of the inhibited Pro-S_{221}C-SbtE (Fig. 1.4a) demonstrated that the folded IMC interacts directly with two surface helices of the protease (Jain, Shinde et al. 1998). Based on this, it was proposed that the α-β-α binding interface of the IMC may represent the folding nucleation motif of the protease, and that stabilization of this sub-structure upon binding of the IMC may help to induce folding (Gallagher, Gilliland et al. 1995). To establish the nature of the interaction between the IMC and protease, Bryan et al analyzed the bimolecular folding of $SbtBPN_{\Delta cal}$.as given by Eqn.1.2,

$$\mathrm{Pro}^{U} + \mathrm{SbtBPN}^{U}_{\Delta cal} \leftrightarrow \mathrm{Pro}^{F}\text{–}\mathrm{SbtBPN}^{I}_{\Delta cal} \leftrightarrow \mathrm{Pro}^{F}\text{–}\mathrm{SbtBPN}^{F}_{\Delta cal} \rightarrow \mathrm{Pro}^{D} + \mathrm{SbtBPN}^{F}_{\Delta cal}$$

Equation (1.2)

Where Pro^{U} and $\mathrm{SbtBPN}^{U}_{\Delta cal}$ represent the unfolded propeptide and protease respectively, $\mathrm{Pro}^{F}\text{–}\mathrm{SbtBPN}^{I}_{\Delta cal}$ represents the partially structured intermediate, $\mathrm{Pro}^{F}\text{–}\mathrm{SbtBPN}^{F}_{\Delta cal}$ represents the folded complex, Pro^{D} represents the degraded propeptide and $\mathrm{SbtBPN}^{F}_{\Delta cal}$ represents the free active protease. Reaction rates were determined using the differences in tryptophan content between the IMC and protease domains. $SbtBPN_{\Delta cal}$, with three tryptophan residues in its primary sequence, shows a 1.7 fold increase in intrinsic

tryptophan fluorescence upon refolding. As the IMC has no tryptophan residues, any change in fluorescence reflects structural changes in the protease domain. Further, the binding of IMC to the protease increases the intrinsic fluorescence of the protease due to shielding of one of the tryptophan residues. By monitoring the rate of change in the intrinsic tryptophan fluorescence, the k_{on} and k_{off} for binding of IMC to $SbtBPN_{\Delta cal}$, and the rates of $SbtBPN_{\Delta cal}$ folding under increasing concentrations of the IMC were determined. These studies demonstrated that the formation of the initial complex ($\mathrm{Pro}^{F}-\mathrm{SbtBPN}^{I}_{\Delta cal}$) between IMC and $SbtBPN_{\Delta cal}$ was the rate-limiting step to folding. However, upon increasing residual structure in the isolated IMC domain the folding of $\mathrm{Pro}^{F}-\mathrm{SbtBPN}^{I}_{\Delta cal}$ to $\mathrm{Pro}^{F}-\mathrm{SbtBPN}^{F}_{\Delta cal}$ becomes rate limiting (Wang, Ruan et al. 1998; Ruan, Hoskins et al. 1999). This established that structural content within the IMC might have a direct effect on its chaperoning function. Simultaneous studies that showed a direct correlation between inhibition constants of isolated IMC mutants, and their chaperoning efficiency, strengthened this conclusion (Li, Hu et al. 1995; Ruvinov, Wang et al. 1997). Hence, it was proposed that the binding energy of the IMC contributes to stabilizing the $\alpha\beta\alpha$ structure either, by surmounting an entropic barrier through stabilization of native interactions or, by overcoming an enthalpic barrier through breaking of non-native interactions. While in the presence of an unstructured IMC the binding energy maybe diluted by its folding, the presence of a structured IMC ensures faster binding and folding.

Although this nucleation propagation mechanism of folding through a "side-on interaction" of the IMC with the protease seems possible, other studies

highlighted below question this hypothesis. Furthermore, the biological reaction is clearly unimolecular and evolution of covalently linked protease domains may itself be to enhance efficiency and economy of trans-chaperone mediated folding. Hence, while the bimolecular folding studies offer initial insights into propeptide mediated folding, establishing similar principles in unimolecular folding is fundamental to the above hypothesis.

B.2.2. Stabilization of folding nucleus–"Top-on Model":

Studies based on unimolecular folding highlight a similar, but slightly varied, mechanism. Random mutagenesis helped identify a number of mutations in the IMC that affected the secretion of active protease. Second site suppressor analysis for an $M_{(-60)}T$ mutation in the IMC identified a $S_{188}L$ substitution in SbtE that restored activity. Since, S188 and M-60 do not interact with each other and are ~47Å apart in the folded Pro-SbtE complex, it was suggested that these residues may interact during the folding process. One possible way in which these two residues could interact is via a "top-on interaction" (Fig. 1.6). To test this hypothesis, cysteine residues were introduced at positions –60 and 188 and the folding of this double-cysteine variant [$M_{(-60)}$C-Pro-S_{188}C-SbtE] was analyzed under oxidizing and reducing conditions. Interestingly, folding under oxidizing conditions results in the formation of a cross-linked intermediate with stable secondary structure. A noteworthy point is that upon prolonged incubation with small peptide substrates such as AAPL-pNA and AAPF-pNA, the trapped intermediate displays catalytic activity. Addition of a reducing agent to the cross-linked

intermediate triggers proteolysis of the IMC, and results in a wild type-like native state. Hence, isolation of the stable cross-linked intermediate suggests that the IMC interacts with the protease in a "top-on orientation" during folding, and further, that this interaction results in productive folding (Inouye, Fu et al. 2001).

Furthermore, the "side-on" interaction of the IMC with the protease is stabilized largely by three hydrogen bonds between E112 in the protease and IMC backbone amides (Fig. 1.4a). Disruption of these hydrogen bonds in E_{112}A-SbtE variant does not affect IMC-mediated folding of the protease. However, the K_i of IMC to E_{112}A-SbtE is lowered ~35-fold relative to that of SbtE. This suggests that the "side-on" interaction of the IMC is critical for inhibition, and further, that the inhibitory and chaperone functions of the IMC may not be obligatorily linked (Fu, Inouye et al. 2000). Additional evidence for the above comes from the fact that IMC of aqualysin is a ten-fold better inhibitor of SbtE, relative to the IMC of SbtE, but is unable to efficiently chaperone the folding of SbtE (Marie-Claire, Yabuta et al. 2001).

The above studies demonstrate that, in unimolecular folding, the "top-on" interaction may initiate folding while the "side-on" interaction is critical for inhibition. Thus the IMC may interact with the protease closer to the active site region before it transits to the "site-on interaction" seen in the crystal structure. This movement of the IMC may be coincident with its cleavage by the protease.

B.2.3. Changes coincident with IMC cleavage:

Subsequent to folding of the polypeptide to a structured state, the peptide bond between the IMC and protease domains is auto-proteolyzed (Li

and Inouye 1996). However the IMC remains bound to the protease as an inhibited complex. The solved crystal structure of the Pro-S_{221}C-SbtE (Fig. 1.4a) inhibited complex offers a structural snapshot of the polypeptide folding at the completion of autoprocessing (Jain, Shinde et al. 1998).

To establish changes coincident with autoprocessing of the IMC, structural properties of the complex before and after autoprocessing were characterized (Shinde and Inouye 1995). ANS is a fluorescent dye that binds to exposed hydrophobic surfaces on structured proteins (Shastry and Udgaonkar 1995). ANS displayed a higher intensity and a shift towards a lower wavelength in the presence of the unautoprocessed complex (Pro-S_{221}A-SbtE) relative to processed complex (Pro-S_{221}C-SbtE). This suggested that there is a large decrease is exposed hydrophobic surface coincident with autoprocessing. Further, in the crystal structure of Pro-S_{221}C-SbtE, the N-terminus of the protease is at least 20Å away from the active site. Based on this, it was proposed that upon autoprocessing, the N-terminus of the protease folds back to form the N-terminal α-helix that contributes Gln2 to calcium binding at the A-site (Fig. 1.4) (Shinde and Inouye 1995). Recent studies establish that while the rate of folding and autoprocessing are independent of calcium, the stability of the autoprocessed complex and mature subtilisin show a strong dependence on calcium. Further, the autoprocessed complex has much higher thermal, thermodynamic, and proteolytic stability, relative to the unautoprocessed complex (Yabuta, Subbian et al. 2002). This demonstrated that the A-site is indeed formed subsequent to cleavage, and that the IMC regulates its formation. The processing of the IMC triggers calcium binding, and induces

structural changes in the protease, that serve to lock the protease in a more stable conformation.

The release and degradation of the IMC from the stable inhibited complex is required for release of active protease (Yabuta, Takagi et al. 2001). While this appears to be mediated through an autocatalytic activation, the precise mechanism of activation is unknown.

B.2.4. Alternative approaches to SbtE folding:

While the above offer direct analyses of IMC-mediated folding, an alternative approach was to identify specific conditions that may allow folding of proteases in absence of their IMCs. Studies demonstrated that folding low concentrations of the protease in 2M potassium acetate at pH 6.5, or refolding the protease from an acid denatured state, allowed recovery of IMC-independent activity. With potassium acetate, while the yield of active protease was initially low immobilization of the protease on a solid support increased the folding efficiency. These experiments elucidate an interesting effect of electrostatics on the IMC-mediated folding process. In the case of the acid-denatured proteins, CD studies of the denatured protein established residual structure in the protease, even after denaturation. This further substantiates the hypothesis that stabilization of a folding nucleus may enable productive folding.

Hence, the IMC appears to initiate folding by stabilizing a sub-structure within the protease. This may involve the interaction of the IMC with the protease in a "top-on" orientation (Fig. 1.6). Upon acquisition of structure within the protease and formation of the catalytic triad, the IMC is cleaved in an

autoprocessing reaction. This is coincident with the movement of the IMC to the "side-on" inhibitory orientation and the formation of a calcium-binding site (Fig. 1.4). Release and degradation of the IMC from this inhibited complex releases active native protease (Fig. 1.7).

B.3. Kinetics of IMC-mediated maturation:

Based on structural and biochemical studies, the overall IMC-mediated maturation pathway (Fig. 1.7) can be studied in terms of three distinct stages. 1) Folding of the polypeptide to a structured state, 2) Autoprocessing of the peptide bond between the propeptide and protease domains resulting in an IMC:protease inhibited complex, and 3) Release and degradation of propeptide from the protease resulting in active protease. As discussed above (See Chapter1- Section III, B.1), most of the early studies focused on the kinetics of folding. Utilizing the optimization of folding by rapid dilution, and the isolation of folding mutants, Yabuta et.al carried out a thorough characterization of the kinetics of all the sub-stages of the maturation pathway. These studies established that while *in-vitro* folding and autoprocessing are rapid and reach completion in 30 mins, the activation of the protease is not seen untill ~240 mins. Hence, activation and not protein folding, is rate limiting to IMC-mediated maturation (Yabuta, Takagi et al. 2001). As evident from the crystal structure, the protease is fully folded as an inhibited complex (Fig. 1.4a). However it appears that the release of the IMC from this complex is extremely slow. Establishing the reasons for this slow release and the mechanism of release would give further insights into IMC-mediated maturation. Further, during the

maturation process the IMC switches from a chaperone to an inhibitor and eventually to a proteolytic substrate. Establishing the energetics of each of these stages will give further insights into the high-energy kinetic barrier and how the IMC functions to modulate this barrier.

C) Similarities and differences in IMCs:

Concurrent studies on subtilases, alpha-lytic protease and carboxypeptidase established that productive folding mediated by the IMC is a kinetically driven process (Baker and Agard 1994). This conservation across unrelated protease families suggests that IMCs have evolved in multiple parallel pathways and may share a common mechanism of action (Eder and Fersht 1995). However, while the protease domains of the subtilases are well conserved, the IMCs show very low sequence similarity. Since IMCs perform functions different from their cognate catalytic domains, they may be subjected to different mutational frequencies due to different functional constraints (Miyata and Yasunaga 1980). Nevertheless, sequence analyses of the IMCs highlight some unique characteristics that may be critical for function. Alignment of known IMC sequences helped identify two small hydrophobic motifs, N1 and N2 that appears to be conserved within such propeptides (Shinde and Inouye 1994). Interestingly, folding mutants identified through earlier random mutagenesis experiments are mostly localized to these motifs, suggesting that these may be necessary for chaperone function (Kobayashi and Inouye 1992; Li, Hu et al. 1995). Further, the IMCs are extremely charged polypeptides. While the mature domain of SbtE has 12% charged residues, the IMC

sequence has 36% charged residues (Shinde and Inouye 1993). Also, the charge on the IMC of SbtE directly complements a pocket around the substrate-binding site (Jain, Shinde *et al.* 1998). Establishing why IMCs have evolved to be charged, and if the charge on IMCs was selected with their chaperone function would give further insights into the nature of kinetic barriers on folding pathways. Another interesting feature of the IMCs is their intrinsic instability. While the IMC of SbtE is largely unstructured when isolated, a number of other IMCs have unstable or molten-globule like characteristics (Buevich, Shinde et al. 2001). However the relevance of the structure of the IMC to its function is yet to be elucidated.

Thus, the goals of this dissertation are to explore,

1) How IMCs facilitate efficient protease maturation
2) How differences in charge, structure, and primary sequence within IMC relate to their chaperone and inhibitory functions, and
3) Why specific proteins have evolved dedicated IMCs to mediate their folding.

IV) FIGURES

Fig. 1.1: Structural conservation among subtilases – All subtilases have a well-conserved subtilisin-like catalytic domain and are stabilized by bound metal ions. Yellow spheres depict calcium ion, while the orange sphere represents sodium ion. **(a)** Crystal structure of Subtilisin E (1SCJ), a bacterial alkaline serine protease, **(b)** Crystal structure of furin (1P8J), a eukaryotic subtilisin-like protease, with the P-domain and an inhibitor (magenta) **(c)** Structure of a Ak.1 protease (1DBI), a thermostable subtilase from *Bacillus*.

Fig. 1.1

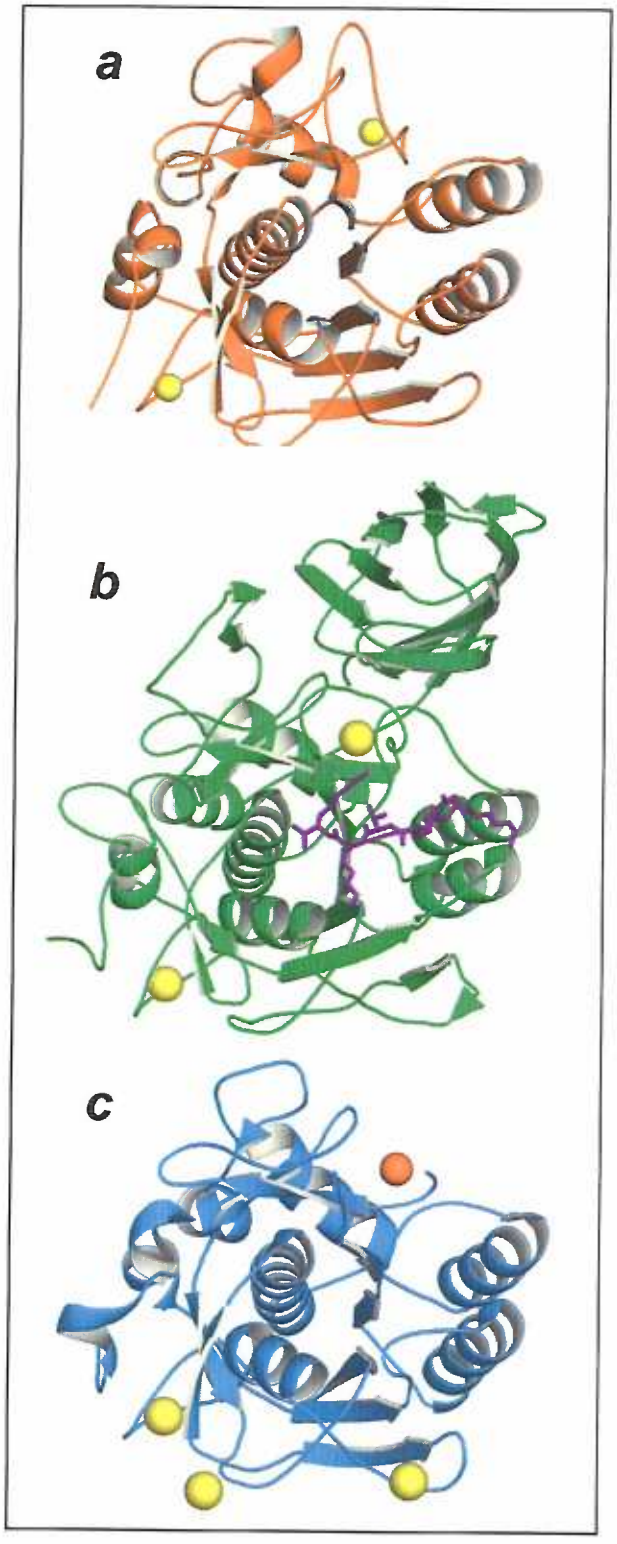


Fig. 1.2: Catalytic Mechanism of Serine Proteases – Subtilases have a conserved -Asp, -His,-Ser catalytic triad that catalyzes proteolysis through two steps: acylation and a subsequent deacylation. A well-conserved Asn residue functions as an oxyanion hole during catalysis (Adapted from Rockwell and Thorner 2004).

Fig. 1.2

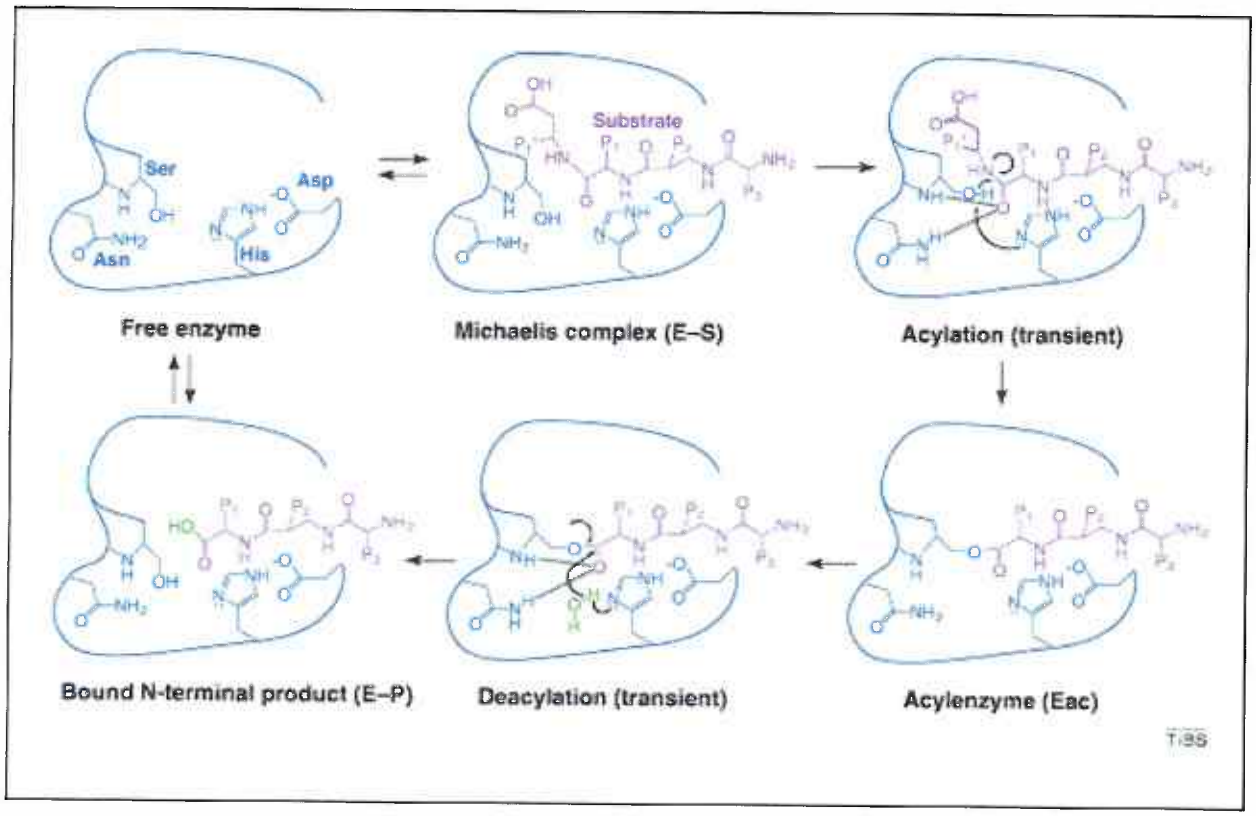

Ser
Asp
Asn
His
Substrate
Free enzyme
Michaelis complex (E–S)
Acylation (transient)
Bound N-terminal product (E–P)
Deacylation (transient)
Acylenzyme (Eac)
TiBS

Fig. 1.3: Structural characterization of SbtE – (a) Structure of SbtE is depicted together with the substrate-binding site (b), and the calcium (orange) binding A-site (c). SbtE also has a second calcium binding site (B-site) of medium affinity. The catalytic residues are highlighted in yellow. **(b)** The substrate binding site is highlighted with an inhibitor (magenta) bound in the S1-S4 pocket.Residues lining the substrate binding pocket are highlighted in blue. **(c)** Calcium binding at the A-site is coordinated by residues from a loop comprised of residues 75-83, an N-terminal Gln, and a Asp.

Fig. 1.3

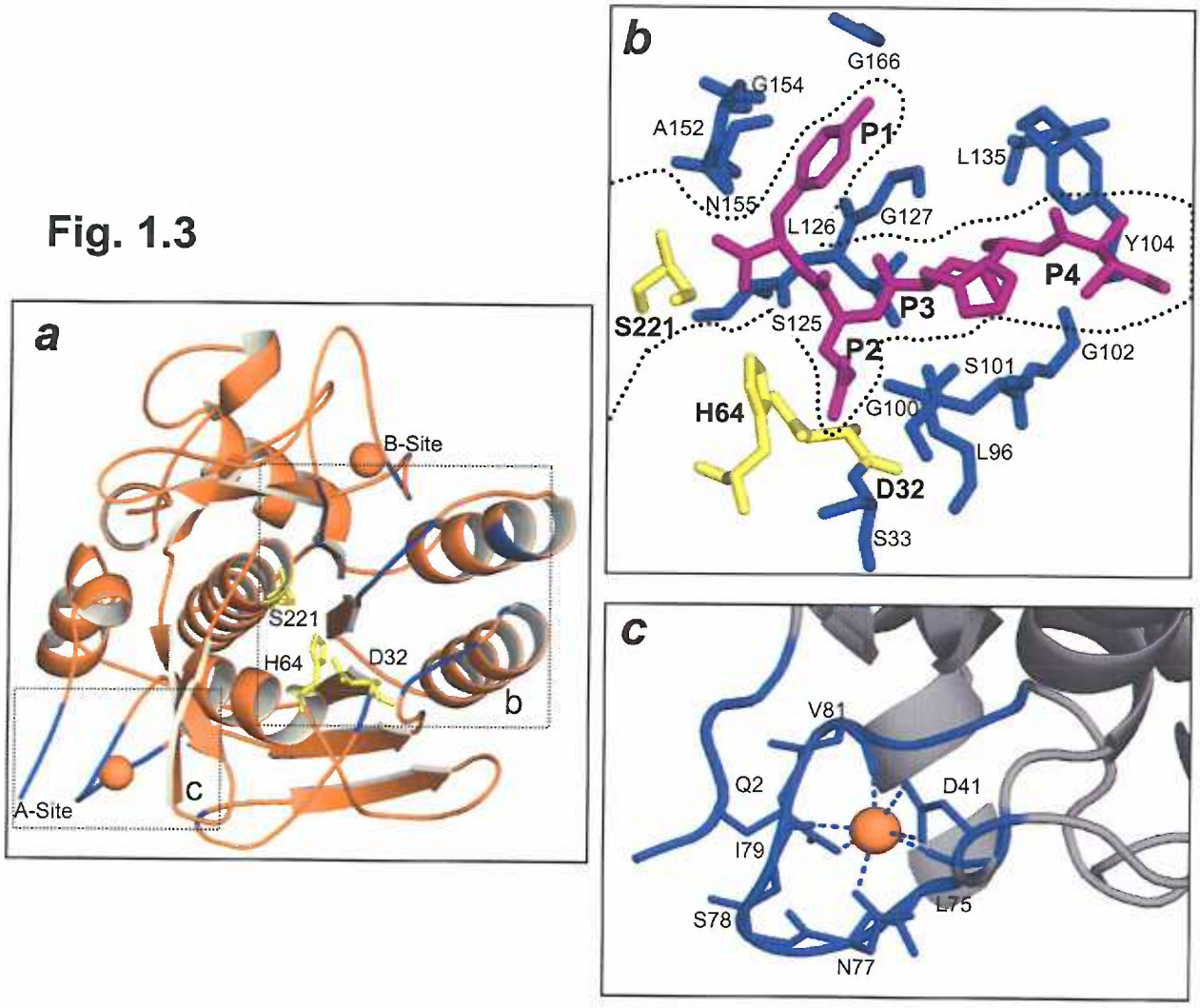
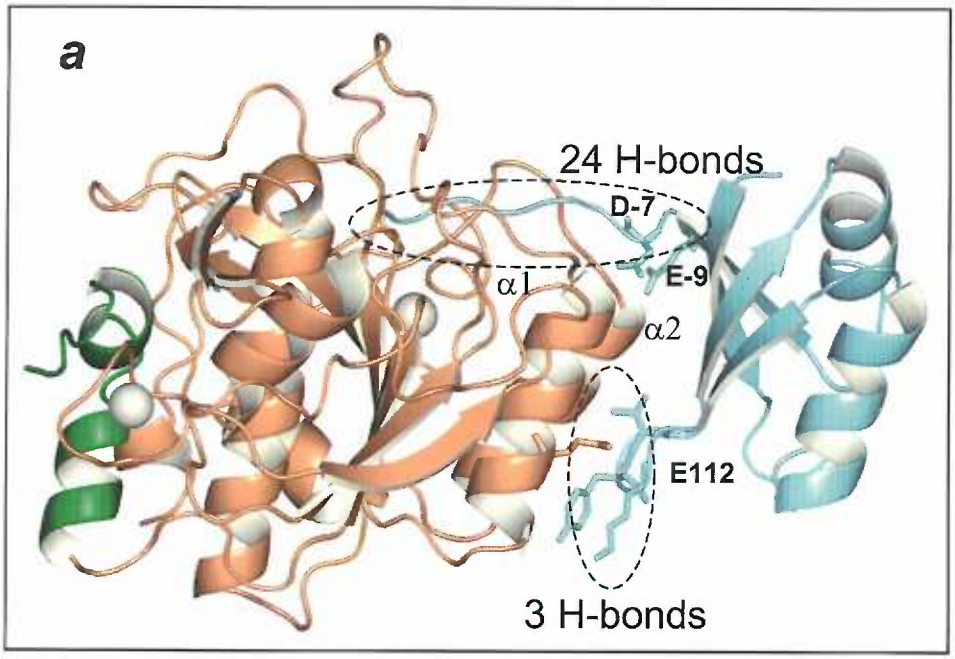

a
B-Site
S221
H64
D32
b
c
A-Site
b
G166
G154
A152
P1
L135
N155
L126
G127
Y104
P4
S221
S125
P3
P2
S101
G102
H64
G100
L96
D32
S33
c
V81
Q2
D41
I79
L75
S78
N77

Fig. 1.4: Structural characterization of Pro-SbtE complex – **(a)** Crystal structure of the autoprocessed, inhibited complex (1SCJ). The IMC (blue) docks against two helices (α1 and α2) in the protease (orange) and occludes the substrate-binding site. The binding interface is stabilized by 27 hydrogen bonds in a asymmetric distribution, of which three are contributed by E112 in protease. D-7 and E-9 form helix caps for the two helices (α1 and α2). The calcium ions (white) and the N-terminal helix (green) are also highlighted. **(b)** Modeled structure of the propeptide-protease complex prior to cleavage. The N-terminus (green) of the protease (orange) is bound to the IMC (blue) at the active site and, hence the calcium-binding A-site is not fully formed.

Fig. 1.4



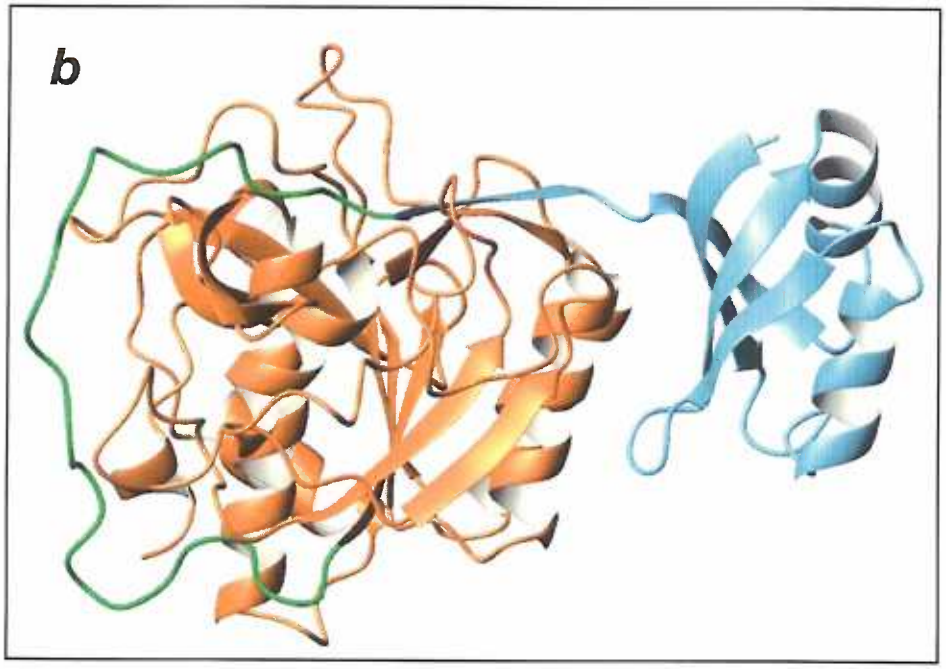


Fig. 1.5: IMCs lower kinetic barriers on the folding pathway – (a) Unfolded Pro-Ser221CysSbtBPN (U) folds to the native state (F) through a stable molten-globule intermediate (I). Kinetic studies establish that the high activation energy barrier to protease folding (broken line) is lowered by >7.5kcal/mol in the presence of the IMC (solid black line).

Fig. 1.5

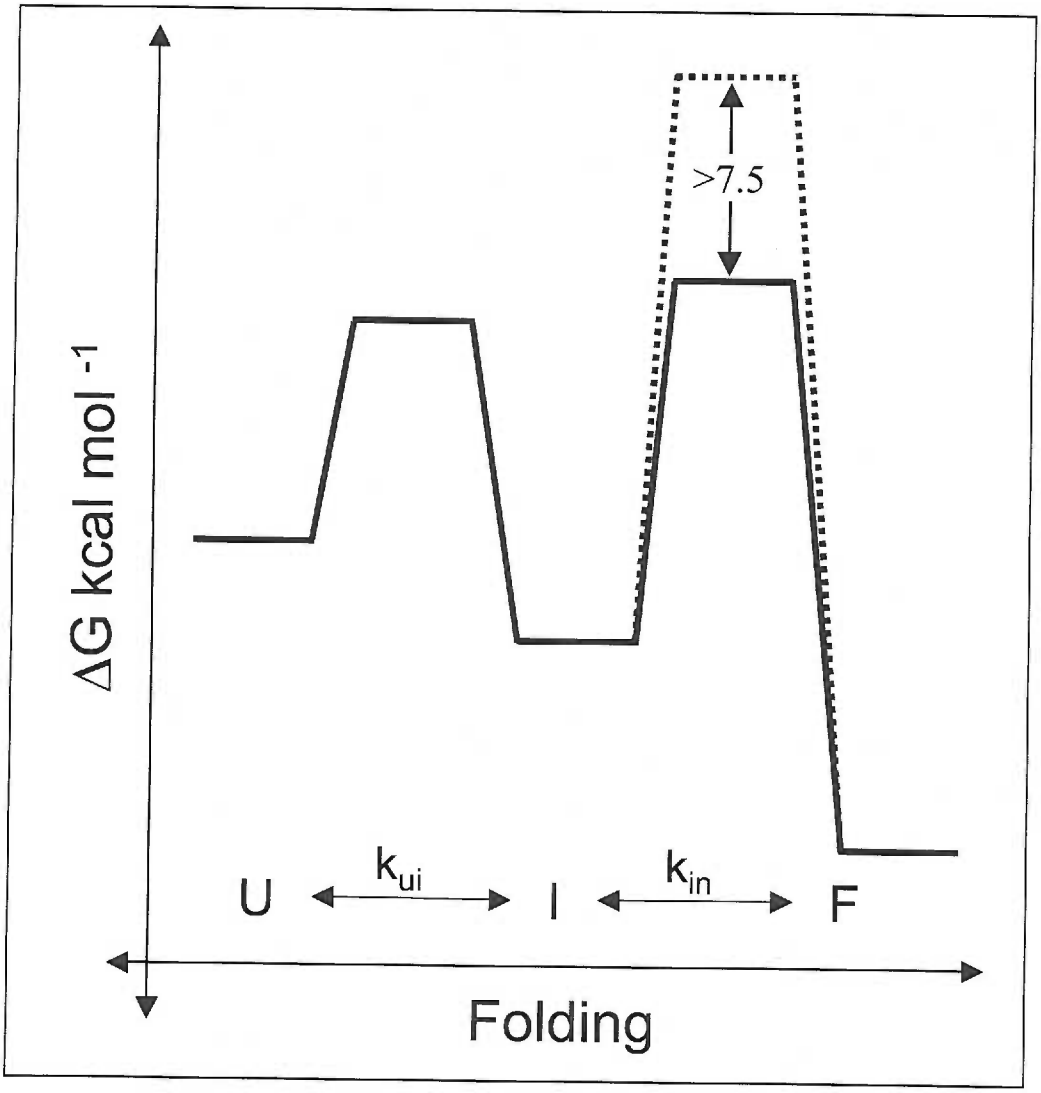


Fig. 1.6: IMC interacts in a "top-on" orientation – (a) M-60 in the IMC (blue) and S188 in the protease (orange), identified through second-site suppressor analysis are, 47Å apart in the inhibited complex. **(b)** Cross-linking studies with a double cysteine mutant helped isolate a stable cross-linked intermediate, based on which the "top-on" interaction between the propeptide (blue) and protease (orange) was modeled.

Fig. 1.6

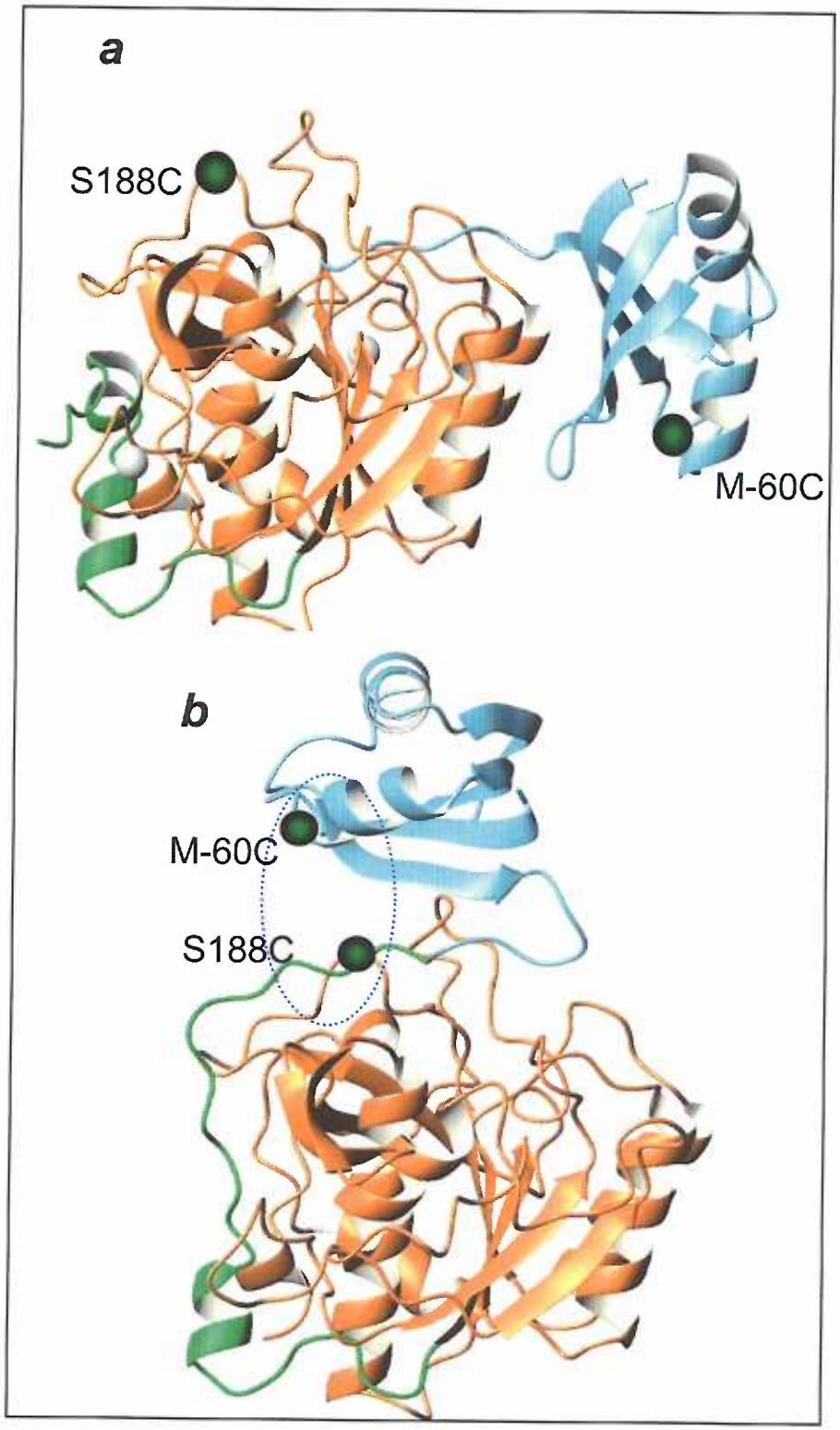
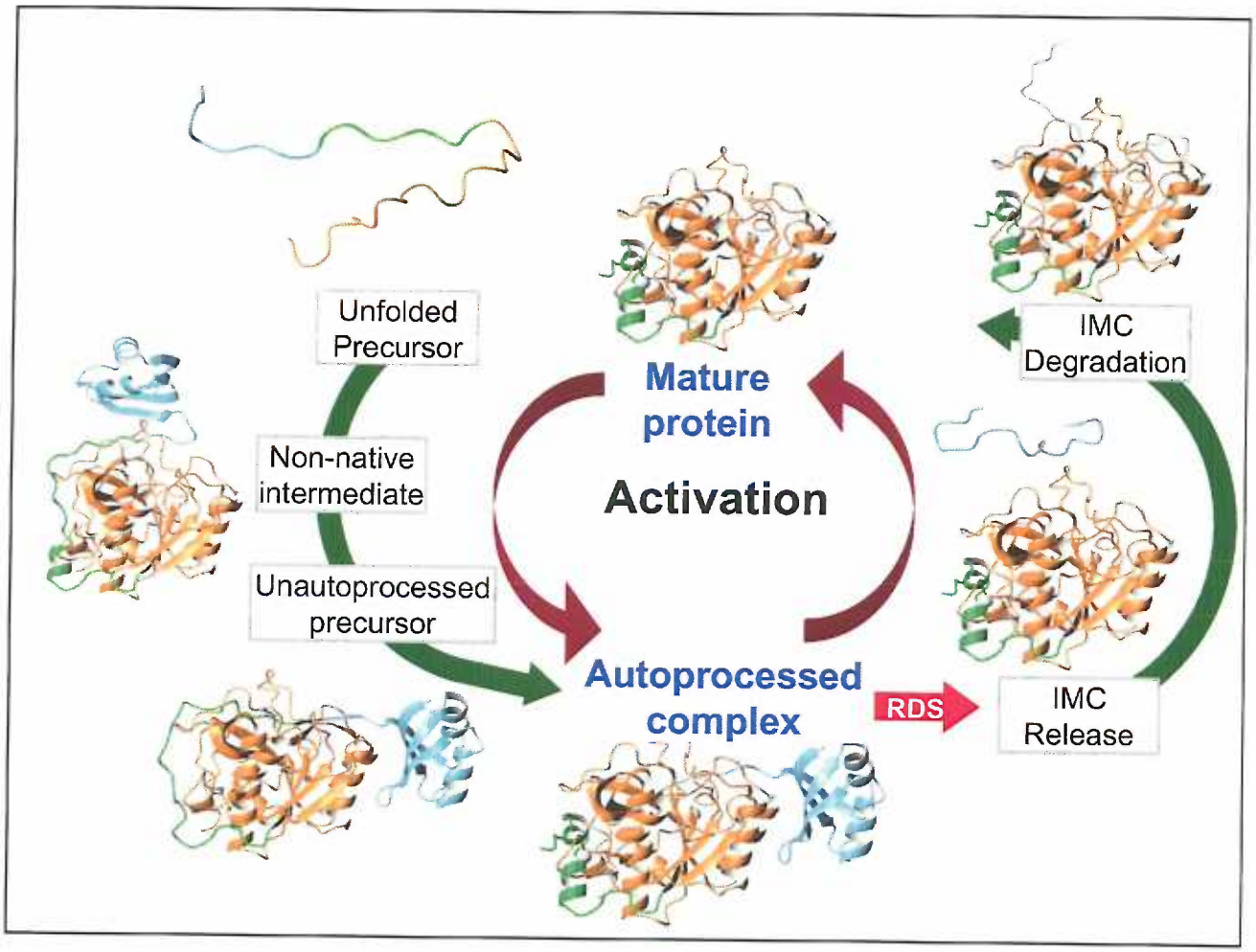


Fig. 1.7: IMC mediated SbtE maturation – N terminal helix of protease (green); calcium ion (white). Maturation of Pro-SbtE occurs in three stages: 1) A non-native top-on interaction between the protease (orange) and IMC (blue), punctuates the transition of the precursor from the unfolded state, to a structured state (unautoprocessed precursor). 2) Once the active site is formed, the precursor autoproteolyzes to an inhibited, autoprocessed Pro:SbtE complex and 3) The release and degradation of the IMC from the complex release active protease that can subsequently trans-activate other proteases. Activation is the rate-limiting step to maturation.

Fig. 1.7


Unfolded
Precursor
Mature
protein
IMC
Degradation
Non-native
intermediate
Activation
Unautoprocessed
precursor
Autoprocessed
complex
RDS
IMC
Release

CHAPTER 2

FOLDING PATHWAY MEDIATED BY AN INTRAMOLECULAR CHAPERONE: INTRINSICALLY UNSTRUCTURED PROPEPTIDE MODULATES STOCHASTIC ACTIVATION OF SUBTILISIN

Ezhilkani Subbian, Yukihiro Yabuta* and Ujwal Shinde

Department of Biochemistry and Molecular Biology,

Oregon Health and Science University,

3181 SW Sam Jackson Park Road, Portland, OR 97239, USA

[1]Current Address: RIKEN Center for Developmental Biology, Kobe, Japan

Journal of Molecular Biology
Vol. 347(2), 367-83; Mar 25, 2005

I) ABSTRACT:

Several secreted proteases are synthesized with N-terminal propeptides that function as Intramolecular Chaperones (IMCs) and direct the folding of proteases to their native functional states. Using Subtilisin E as our model system, we had earlier established that (i) release and degradation of the IMC from its complex with the protease, at the completion of folding, is the rate-determining step to protease maturation and (ii) IMC of SbtE is an intrinsically unstructured, extremely charged polypeptide that adopts an $\alpha-\beta$ structure only in the presence of the protease. Here, we explore the mechanism of IMC release and the intricate relationship between IMC structure and protease activation. We establish that the release of the first IMC from its protease domain is a non-deterministic event that subsequently triggers an activation cascade through trans-proteolysis. By *in-silico* simulation of the protease maturation pathway through application of stochastic algorithms, we further analyze the sub-stages of the release step. Our work shows that modulating the structure of the IMC domain through external solvent conditions can vary both the time and randomness of protease activation. This behavior of the protease can be correlated to varying the release-rebinding equilibrium of IMC, through simulation. Thus, a delicate balance underlies IMC structure, release and protease activation. Proteases are ubiquitous enzymes crucial for fundamental cellular processes and require deterministic activation mechanisms. Our work on SbtE establishes that through selection of an intrinsically unstructured IMC domain, nature appears to have selected for a viable deterministic handle that controls a fundamentally random event. While this outlines an important

mechanism for regulation of protease activation, it also provides a unique approach to maintain industrially viable subtilisins in extremely stable states that can be activated at will.

II) INTRODUCTION:

Proteases are essential for biology and play vital roles in several cellular processes such as; development, immune response, apoptosis, cell cycle and protein turn over. Almost all proteases are secreted as inactive zymogens that are activated by specific cellular factors at precise times and in definite cellular locations (Khan and James 1998). Several zymogens are secreted with N-terminal propeptides that aid in protease inhibition, protein targeting, stabilization and/or regulation (Li, Hu et al. 1995; Takeshima, Sakaguchi et al. 1995; Ruvinov, Wang et al. 1997; Cui, Hackenmiller et al. 2001). However, N-terminal propeptides of specific, secreted proteases have evolved to assist the folding of their cognate protease domain and are termed as Intramolecular chaperones (IMCs) (Shinde and Inouye 1993; Shinde and Inouye 2000). IMCs are covalently attached to the proteins they help to fold and are essential to obtain active proteases both *in vivo* and *in vitro* (Ikemura, Takagi et al. 1987; Silen and Agard 1989; Zhu, Ohta et al. 1989; Shinde and Inouye 2000). IMC dependent folding is observed in several prokaryotic and eukaryotic proteases that include serine, cysteine, aspartyl and metalloproteases (Shinde and Inouye 1993; Shinde and Inouye 2000). Since IMCs are also good inhibitors of their cognate proteases (Li, Hu et al. 1995; Sohl, Shiau et al. 1997; Khan and James 1998), release and degradation of the IMC subsequent to folding, signals the completion of maturation and releases active protease (Yabuta, Subbian et al. 2002). Degradation of IMCs uncouples folding and unfolding pathways, and increases unfolding energy barriers of native proteins (Yabuta, Takagi et al. 2001) (Sohl, Jaswal et al. 1998; Jaswal, Sohl et al. 2002). This ensures

proteolytic stability of the native state and helps secreted proteases to function in harsh extra-cellular environments (Jaswal, Sohl et al. 2002).

Subtilisin (Bryan, Alexander et al. 1992; Shinde and Inouye 1993; Eder and Fersht 1995; Ruan, Hoskins et al. 1999; Yabuta, Takagi et al. 2001; Bryan 2002; Yabuta, Subbian et al. 2002) and α-lytic protease (Baker, Sohl et al. 1992; Sohl, Jaswal et al. 1998; Jaswal, Sohl et al. 2002; Cunningham and Agard 2003; Fuhrmann, Kelch et al. 2004; Truhlar, Cunningham et al. 2004) constitute the best-studied examples of IMC-mediated protein folding. Maturation of IMC-subtilisin involves a series of distinct steps (Yabuta, Takagi et al. 2001; Bryan 2002; Yabuta, Subbian et al. 2002) (Fig. 2.1a): (i) Folding of the protease mediated by its cognate IMC (ii) Autoprocessing of the IMC at its primary cleavage site to form a stable IMC:Subtilisin inhibition complex (Jain, Shinde et al. 1998) and, (iii) The release and degradation of the now inhibitory IMC-domain from the cleaved complex resulting in activation. We earlier established that while folding and autoprocessing are rapid *in vitro*, the rate-determining release and degradation of the IMC is extremely slow (Yabuta, Takagi et al. 2001). However, the mechanism of release and the significance of the delay are unknown. Further, while the X-ray crystallographic structure (Jain, Shinde et al. 1998) demonstrates that the IMC-domain adopts a well-defined α–β conformation as an inhibition complex with subtilisin, circular dichroism and NMR spectroscopy studies establish that the IMC-domain is unstructured in its isolated form (Buevich, Shinde et al. 2001; Yabuta, Subbian et al. 2003). IMC of SbtE is also an extremely charged polypeptide (36% charged) (Shinde and Inouye 1993) that exhibits properties similar to an

emerging family of intrinsically unstructured proteins (Dyson and Wright 2002; Dyson and Wright 2002; Tompa 2002). This unstructured 77-residue IMC-domain is nonetheless capable of functioning as both a chaperone and potent inhibitor. Interestingly IMC-mutants with increased independent secondary structure are invariably better chaperones and are more potent inhibitors (Eder, Rheinnecker et al. 1993; Li, Hu et al. 1995; Ruvinov, Wang et al. 1997; Ruan, Hoskins et al. 1999; Bryan 2002; Yabuta, Subbian et al. 2003). Understanding why IMCs have evolved to be intrinsically unstructured will help establish the significance of this and also determine the intricacies of the maturation process.

Here we explore the mechanism for release of the IMC domain, and analyze the relation between IMC-structure and precursor activation. Our results demonstrate that the release of the IMC domain, which results in the first active protease molecule, is a probabilistic event. The first free protease molecule thus formed then triggers rapid activation through trans-proteolysis. Specific solvent conditions that modulate structure within the isolated IMC domain can significantly affect protease activation and its associated randomness. To better understand the randomness associated with activation, we mathematically modeled the entire maturation process using Cellware (Dhar, Meng et al. 2004). Our simulations concur with a stochastic release step and suggest that the IMC domain modulates protease activation by altering the binding equilibrium of IMC. Furthermore, the activation of SbtE coincides with its transition from a thermodynamically stable inhibited complex to a kinetically trapped native state. The presence of an intrinsically unstructured IMC aids this transition. Hence IMCs appear to be not optimized for the individual chaperone

or inhibition functions but rather selected to be intrinsically unstructured to maintain synergy, precision and control of the protease maturation pathway. Because of their high stability, subtilisins are widely used in industrial applications (Wolff, Showell et al. 1996; Gupta, Beg et al. 2002) as non-specific proteases. Thus, the ability to control and effect precise temporal activation of subtilisins has wide applicability in industry and would provide a unique approach to improve the shelf life of proteases.

III) RESULTS:

Maturation Pathway of IMC-subtilisin:

The maturation of IMC-subtilisin involves the three distinct stages of folding, autoprocessing and, IMC release and degradation (Shinde and Inouye 1995; Yabuta, Takagi et al. 2001) (Fig. 2.1a). Release and degradation involves the release of the first active protease molecule and subsequent trans-proteolysis resulting in complete protease activation. Folding kinetics can be monitored by changes in secondary structure measured using circular dichroism spectroscopy (Marie-Claire, Yabuta et al. 2001; Yabuta, Takagi et al. 2001) and is found to be complete within 5min of initiation (Fig. 2.1c, inset). To monitor the kinetics of autoprocessing aliquots were taken over time and the amount of IMC:Subtilisin complex formed was monitored using SDS PAGE and quantitative gel scanning densitometry. Fig. 2.1b depicts aliquots taken as a function of time from a single maturation reaction. The kinetics of formation of precursor, mature and IMC fragments, analyzed by gel-scanning densitometry (Fu, Inouye et al. 2000; Yabuta, Takagi et al. 2001; Yabuta, Subbian et al. 2002), establish that autoprocessing is complete within 30 min of folding initiation (Fig. 2.1c). The steps of release and degradation of the IMC-domain, which together constitute IMC-subtilisin activation, can be monitored through the release of *p*-nitroaniline, following degradation of a chromogenic substrate (Yabuta, Takagi et al. 2001; Yabuta, Subbian et al. 2003) (Materials and Methods). Activation on an average, takes ~240 min, that is ~210 min after completion of autoprocessing and represents the rate-determining step of maturation (Yabuta, Takagi et al. 2001). Further there appears to be randomness

in the activation time between experiments. In order to decipher the significance of the delay and randomness in activation time, and the precise mechanism of IMC release, we critically analyzed the activation of subtilisin.

Stochastic Activation of IMC-subtilisin:

Folding was initiated at room temperature (23°C) through a rapid dilution of denatured protein into a large volume as described in Materials and Methods. Aliquots of 200 μl of the folded protein were taken on a microplate after 30 min of folding initiation, and the emergence of activity in these aliquots was observed. Since earlier studies have established that the folding and autoprocessing reaction of IMC-subtilisin is saturated within 30 min of folding initiation (Yabuta, Takagi et al. 2001), these two stages are essentially complete when aliquots of the protein are taken onto the microplate. The activity of subtilisin can be monitored using a chromogenic substrate N-succ AAPFpNA. The synthetic substrate has a 10^5-fold lower affinity and ~20 fold lower proteolytic sites than the 77-residue IMC-domain (Li, Hu et al. 1995). Hence the emergence of yellow color in the aliquots indicates the activation of the protease after complete degradation of the inhibitory IMC-domain. Fig. 2.2a & 2.2b establish that the individual aliquots of the folded protein undergo 'none-to-all' activation at different times, although they represent the *same* folding reaction. The activation time of these aliquots can be calculated from the X-intercept of the transition curve, corresponding to the sudden burst of activity. Interestingly, a plot of number of wells active versus time of activation displays a Poisson like distribution around a mean of 235 min (Fig. 2.2c). This suggests that there exists an inherent randomness in the activation of IMC-subtilisin. Since

aliquots were removed from a single rapid dilution reaction after 30mins, the observed variations in activation time may not occur due to subtle differences in folding or autoprocessing.

It may be argued that the distribution in the lag time in the mechanism is not surprising because it is located at the end of a sequential reaction. Therefore small fluctuations in the individual steps of folding and autoprocessing may add up to result in a notable distribution of the lag time. We argue that cumulative fluctuations in the steps of folding and autoprocessing are not a source of this observed variation in activation times because the reaction mixture was aliquoted onto the microplate 30 min after folding initiation, as described in the Materials and Methods. Therefore, the folding and autoprocessing reactions are completed in a homogenous solution prior to removing aliquots onto the microplate. Moreover, the rates of folding and the extent of autoprocessing are highly deterministic events (Yabuta, Takagi et al. 2001). Hence, variation in activation times should result from differences in release and degradation of the IMC-domain. To establish this conclusively, differences between enzymatically active and inactive aliquots were monitored through SDS-PAGE. IMC-subtilisin maturation was terminated by trichloroacetic acid (TCA) precipitation (Yabuta, Takagi et al. 2001; Yabuta, Subbian et al. 2002; Yabuta, Subbian et al. 2003), after 20 - 30 % of the wells displayed protease activity. An SDS-PAGE analysis (Fig. 2.2b, inset) establishes that, while the amount of mature protease is similar, the inhibitory IMC-domain is absent from enzymatically active aliquots. Observed variation in activation time is also evident when the folding reaction is aliquoted into tubes made from different

materials, such as glass, polyethylene, polypropylene etc., establishing that it is not an effect of the material used (data not shown).

To further establish that the observed variation in the activation times was not an artifact of different enzyme concentration, the amount of active subtilisin in individual wells of the microplate was estimated. IMC-subtilisin was folded at 23 °C through rapid dilution into buffer that does not contain the synthetic substrate (see Materials and Methods). After aliquoting on to a microplate the reaction was allowed to proceed for 24 hrs, at the desired temperature. This prolonged incubation allows complete maturation of IMC-subtilisin. Subsequent to this incubation, synthetic substrate was added to a final concentration of 0.5 mM and the velocity of substrate hydrolysis in each well was measured. Under these conditions, velocity of substrate cleavage was proportional to enzyme concentration and 80 nM of mature subtilisin E gave ~20 mOD min^{-1}. The yield of protease varied less than 1% between individual aliquots and was linearly dependent upon the precursor concentration used for folding (Fig. 2.2d). These data suggest that the variation in activation time does not arise due to aggregation or extent of autoprocessing, since this would be expected to diminish yields of active subtilisin. Lack of aggregation was further confirmed using gel filtration chromatography (data not shown).

Effect of Concentration on Release of IMC:

Since an autocatalytic activation cascade could be affected by concentration, we next established the effect of IMC-subtilisin concentration upon the time of activation. In these sets of experiments the folding reaction was initiated for

different concentrations and the observed variations in activation times were studied. While these variations were observed at the different precursor concentrations (data not shown), Fig. 2.2d demonstrates that the average time of activation increases exponentially at low concentrations (<50 nM) and is consistent with an auto-catalyzed bimolecular (trans-proteolysis) reaction. At higher concentrations, activation time saturates out, probably limited by turnover of the inhibitory IMC-domain. It is important to note that under different precursor concentrations, the extent of autoprocessing remains constant (Inset to Fig. 2.2d). This suggests that folding/autoprocessing are concentration independent (unimolecular) while activation is concentration dependent (bimolecular).

Proteolysis of IMC domain:

Release and degradation of IMC from its complex with the protease involves two known progressive sub-stages represented by Eqn 2.1.

$$\underset{\text{I}}{IMC_{folded}:SbtE \leftrightarrow} \quad \underset{\text{II}}{IMC_{unfolded} + SbtE \rightarrow IMC_{degraded} + SbtE^*} \qquad \rightarrow \text{Equation (2.1)}$$

The variation in time of activation may result from a rate-determining release (I), or may be limited by rate of proteolysis (II). Since release and degradation of IMC are closely coupled reactions we monitored the rate of proteolysis of IMC indirectly. A 20-fold excess of free IMC was added to mature subtilisin to maintain an IMC:Subtilisin complex with saturating amounts of IMC. Aliquots of the reaction were removed at different time intervals, the reaction was stopped using TCA, and the samples were resolved using Tricine-SDS-

PAGE and quantitated using gel-scanning densitometry (Fig. 2.3a, inset). Approximately 9-fold of IMC was degraded within 15s. This may represent the time necessary for IMC to fold and inhibit subtilisin activity. Once bound, mature subtilisin requires 30min to proteolyze the remaining 11-fold excess of the IMC (Fig. 2.3a). The reaction was also monitored using the synthetic chromogenic substrate. In this experiment 100μl of the reaction mixture (active subtilisin with 20 fold excess of IMC) was rapidly mixed with 100μl of the enzyme assay buffer that contains 0.5mM chromogenic substrate. The release of p-nitroaniline was monitored as described (Materials and Methods). The time of p-nitroaniline release coincides with complete degradation of added IMC (Fig. 2.3a). As the IMC has a ~10^5 fold higher affinity relative to the synthetic substrate, the release of p-nitroaniline was seen only after the complete degradation of the IMC domain. Thus the above establishes that IMC is degraded at ~7.5 and 0.05pmol/s during the rapid and slow phases respectively. An ~20 fold excess of IMC is completely degraded by mature protease in 30min which is ~8 fold lower than the activation time of 240 min. Hence, this suggests, that release of IMC (I) and not its proteolysis (II) is rate-limiting to the overall maturation.

Effect of free subtilisin on activation time:

Our data suggests a bimolecular activation cascade that is driven by a stochastic release. Since a bimolecular activation would be reactive to the presence of free protease, we established the extent to which exogenously added subtilisin modulates the time of activation. IMC-Subtilisin was folded as described (Materials and Methods) and different amounts of active subtilisin

were added to the folding mixture at 30min. Aliquots were removed from the folding reaction at fixed time intervals and subtilisin activity was measured as described (Materials and Methods). IMC-Subtilisin, folded in the absence of exogenously added subtilisin, and free mature subtilisin were used as positive and negative controls (Fig. 2.3b, inset). Exogenously added subtilisin significantly reduces the time of activation with activation time reduced by ~50% with addition of even ~20% active subtilisin (Fig. 2.3b, inset). Fig. 2.3b establishes that activation time reduces as an exponential function of exogenously added active subtilisin. Thus the release of the first free protease is inherently random but is closely dependent on the presence of free active protease.

Taken together, the above data suggests that maturation of IMC-subtilisin is inherently stochastic. The randomness is not due to folding or autoprocessing but rather due to a stochastic activation step (Fig. 2.2). This activation is driven by random release of the inhibitory IMC. Once the first free protease is formed, an autocatalytic activation drives maturation to completion. Hence, activation is closely modulated by the concentration of the precursor (Fig. 2.2d) and by the presence of free active protease (Fig. 2.3b).

Modeling and Simulation of IMC-Subtilisin maturation:

To establish and analyze the nature of the stochastic interactions driving protease activation we modeled the maturation pathway of IMC-subtilisin using Cellware (Dhar, Meng et al. 2004), a recently developed simulation package.

Cellware offers a multi-algorithmic environment for modeling and simulation of kinetic networks using both deterministic and stochastic algorithms (Dhar, Meng et al. 2004; Meng, Somani et al. 2004). Elementary molecular interactions are modeled in terms of rate equations and the temporal changes in molecular species or their stationary state values are studied. Through the use of stochastic algorithms, such models allow us to analyze stochastic effects on the evolution of molecular species, and to study how it affects the behavior of the system. The main advantage offered by such models is that it allows us to evaluate how modulating individual interactions affects the overall process outcome (Meng, Somani et al. 2004; Zhu, Huang et al. 2004). Using Cellware 1.2 we modeled the complete maturation pathway of IMC-Subtilisin. Unfolded IMC-Subtilisin, Folded IMC-Subtilisin, Autoprocessed IMC:Subtilisin complex, Free folded IMC, Degraded IMC and Free active Subtilisin were represented as individual molecular species. Interactions between them were defined using law of mass action with experimentally derived rate constants (Ruvinov, Wang et al. 1997; Yabuta, Takagi et al. 2001; Yabuta, Subbian et al. 2002). Fig. 2.4a depicts the complete pathway together with the experimental rate constants that were used for simulation. For an initial particle size of 10000 unfolded molecules, the evolution of individual molecular species along the maturation pathway was simulated using the Gillespie algorithm (Materials and Methods). Behavior of folded, autoprocessed and mature subtilisin over simulated time concurs with experimental data (Fig. 2.4b). Folded molecules rapidly decelerate over time into autoprocessed complex. Rapid decrease in accumulated

autoprocessed particles coincides with a none-to-all increase in mature species indicative of rapid, autocatalytic activation.

We next monitored the behavior of activation time between iterations simulated for an initial unfolded particle size of 10000. The model allows us to evaluate the synergy and differences, in formation of each species, between iterations. This can be correlated to the behavior of individual aliquots taken from the same maturation reaction. For the different iterations, we compared the evolution of folded, autoprocessed and active protease. The formation of folded (Fig. 2.5a) and autoprocessed (Fig. 2.5b) species is uniform between the iterations. However release of free protease is extremely random (Fig. 2.5c). A plot of number of active samples versus time of activation for 1000 sample iterations follows a Poisson's distribution similar to experimental results (Fig. 2.5d). Thus our model corroborates that folding and autoprocessing are deterministic, and randomness in maturation of subtilisin is a result of the stochasticity in release and degradation of the IMC.

Our experimental results clearly established an autocatalytic activation that is dependent on initial precursor concentration. To further test our theoretical model we simulated the maturation pathway for varying particle sizes. The mean time of activation for 1000 iterations was determined for each initial particle size. The simulated activation time follows an exponential dependence on particle size with the time of activation increasing exponentially with decrease in initial particle size. Similar to our experimental data, at higher particle sizes the activation time saturates out. Fig. 2.5e depicts the normalized, simulation data, together with the experimental data. Next we simulated the

effect of exogenously introduced active protease on time of activation. To our maturation model, active protease was introduced as an activator molecule with a rate of reaction similar to mature subtilisin. For varying ratios of activators, the maturation was modeled and the variation in activation time was observed. Exogenous addition of active protease to the maturation pathway greatly reduces the time of activation and follows an exponential curve similar to our experimental results (Fig. 2.5f). Further the intrinsic noise or degree of randomness in activation time is closely modulated by the amount of free protease.

Thus, we fit a minimalist mathematical model to the protease maturation pathway (Fig. 2.4a). Fitting our model to a stochastic algorithm clearly correlates our experimental results (Fig.5). The theoretical model substantiates our experimental data that subtilisin activation is indeed stochastic. This stochasticity results from a fundamentally random release event, which then results in rapid protease activation through autocatalysis.

Time specific activation of IMC-subtilisin:

It is widely accepted that stochastic events underlie deterministic biological processes such as gene-expression and signal transduction (Elowitz, Levine et al. 2002; Ozbudak, Thattai et al. 2002; Swain, Elowitz et al. 2002). Although functional interactions between macromolecules are inherently "noisy", the noise is nature's way to allow small changes in macromolecular interactions to signal specific transitions between states. The intrinsic noise allows for easier regulation and adaptation to change (Elowitz, Levine et al.

2002; Ozbudak, Thattai et al. 2002; Swain, Elowitz et al. 2002). Hence, our interest was to identify specific conditions that regulate the inherent randomness, and hence the activation of subtilisin.

A refolding reaction was initiated as described earlier and structural pertubants that could expedite precursor activation were identified. Since subtilisin is a secreted enzyme, free protease and pH/salt changes can represent naturally occurring conditions (Siezen 1996; Siezen and Leunissen 1997). Fig. 2.6a demonstrates that both pH/salt changes and a small amount of active subtilisin can expedite precursor activation. Interestingly, sodium dodecyl sulfate (SDS) an ionic detergent can also rapidly activate the precursor (Fig. 2.6a), while a non-ionic detergent such as DM (n-decyl-maltopyranoside) is ineffective (data not shown). However, the efficacy of each activator is different, with 0.01% SDS being the most effective causing instantaneous activation (Fig. 2.6a). Upon the addition of the above activators, activation is rapid and deterministic (Fig. 2.6a). It is important to note that mature subtilisin is stable in 0.01% SDS. To understand how these activators function, aliquots from the precursor maturation reaction were removed at different time intervals and subjected to SDS-PAGE. Fig. 2.6b establishes that all activators promote specific degradation of the precursor and cleaved IMC-domain (Fig. 2.6b). When activation was expedited by adding specific solvents at different times after the completion of the first cleavage, the yield of mature subtilisin was found to be independent of both, the types of activators and the time they were added (Fig. 2.6c). This confirms that precursor maturation is complete within 30 min of folding initiation (Yabuta, Takagi et al. 2001) and the cleaved complex is

in an active competent state. However, if activators were added prior to the completion of folding or autoprocessing, the yields of the protease are lower (data not shown). Earlier studies on subtilisin suggest that the delay seen in the time of activation may be necessary to increase protease yields and ensure efficiency of the folding process (Zhu, Ohta et al. 1989; Yabuta, Takagi et al. 2001; Yabuta, Subbian et al. 2002). Our results establish that the inhibitory function of the IMC-domain is necessary only until the completion of autoprocessing and, prolonged inhibition (after 30 min) does not increase protease yield. However IMC-domain may function as a regulator, by blocking enzymatic activity until a suitable 'activator' is available. We establish here that active protease, pH, and SDS, can function as *in vitro* regulators. Osmotic, temperature and pH shock may suffice as physiological activators.

In order to function as regulators, the IMC-domains should also be able to prolong the release of enzyme activity beyond its normal activation time. We next attempted to identify solvent conditions that prolong precursor activation. It has been established that the isolated IMC-domain is unstructured but adopts significant secondary structure in glycerol and trifluoroethanol (Eder, Rheinnecker et al. 1993; Bryan, Wang et al. 1995). Moreover, enhancing the independent stability of the IMC-domain through site-specific mutations increases its affinity for its cognate protease, as well as its chaperoning activity (Kojima, Minagawa et al. 1998; Ruan, Hoskins et al. 1999; Yabuta, Subbian et al. 2003). Here we examined whether such solvents can prolong precursor activation. To the individual aliquots taken from a refolding mixture, 100 μl of the different additives or 'stabilizers' were added and the protease activation was

monitored. Fig. 2.7a demonstrates that the presence of glycerol (10%) increases the activation time and its associated variations by approximately three fold. Interestingly, the addition of SDS (0.01%) rapidly activates the complex under both normal and stabilizing conditions by expediting IMC-degradation (Fig. 2.7a & 2.7b), although the activation is marginally delayed in glycerol. High salt [1M $(NH_4)_2SO_4$] also prolongs activation and shows a similar but more enhanced effect. In the case of high salt the intrinsic activation of the protease was not observed till 24 hrs after folding initiation. We also identified a high salt-glycerol combination as a stabilizer that prolongs activation. While the glycerol-salt condition stabilizes the inhibitory complex for several weeks (data not shown), SDS addition always triggers activation (Fig. 2.7b & 2.7c). We propose that this prolonged activation results specifically through the modulation of IMC release and its subsequent degradation, because the stabilizing solvents were added to the reaction only after the completion of folding and autoprocessing (30 min after folding initiation). It is important to note that while glycerol (10%) induces structure into the isolated IMC-domain (Fig. 2.8b), it alters neither subtilisin activity (data not shown), nor the secondary structure of the inhibition complex and mature subtilisin (Fig. 2.8a). This suggests that glycerol (10%) shifts the equilibrium towards formation of the inhibition complex by inducing structure in the isolated IMC-domain (Eqn.2.1).

IMC structure and protease activation:

The X-ray-structure of the inhibition complex establishes that the B-factor of the IMC in complex, is twice that of subtilisin, and implies substantially larger

backbone dynamics within the IMC-domain (Jain, Shinde et al. 1998). We therefore examined whether glycerol (10%) affects the structure of the IMC-domain in the inhibition complex, to decrease its proteolytic susceptibility. The inactive IMC:Subtilisin complex was isolated and incubated under normal and stabilizing conditions with TPCK-treated trypsin (1/10 the amount of complex). Proteolytic susceptibility of the IMC-domain from the complex was measured using a quantitative gel-scanning densitometry as described in Materials and Methods. Fig. 2.8c shows that trans-proteolysis of the IMC within the inhibition complex, by TPCK-treated trypsin, is substantially reduced in glycerol, suggesting a significantly compacted folded state for the IMC domain (Jaswal, Sohl et al. 2002). Using our maturation model we additionally simulated the trans-proteolysis of IMC. Our model was varied to include the presence of a activator molecule whose activity was defined by the proteolytic activity of trypsin. Using an initial particle size of 1000 we simulated the trans-degradation of IMC in the presence of trypsin as an activator. Fig. 2.8d depicts the simulated trans-proteolysis curve. Altering the binding affinity of IMC for the protease by 10-fold closely overlaps the trans-degradation in the presence of glycerol (Fig. 2.8d). Thus inducing structure in the IMC domain, reduces its conformational dynamics and shifts the equilibrium (IMC_U ⟵⟶ IMC_F) of the IMC-domain towards the bound state ($^{*}IMC_F$:S), as seen from Eqn 2.1, resulting in a tighter inhibition complex.

To establish this conclusively, we next determined the affinity of IMC for the protease in the presence and absence of 10% glycerol. The affinity constant of IMC for subtilisin E was measured using isothermal titration calorimetry (Fig.

2.9*a, lower panel*). The K_a for this interaction was $7.34x10^6\ M^{-1}$ and is similar to that observed through previous independent binding studies (Li, Hu et al. 1995; Bryan 2002). In the presence of 10% Glycerol the affinity of IMC is increased ~10 fold to $7.2x10^7\ M^{-1}$ (Fig. 2.9*a, upper panel*). Thus, inducing structure in the IMC, enhances its binding affinity, and modulates release and associated randomness (Fig. 2.7a). We next simulated the maturation pathway for different values of K_a and analyzed the distribution in time of activation (Fig. 2.9b & c). For each value of K_a and an initial particle size of 10000, the mean time of activation for 1000 iterations was estimated. At low K_a, activation is fairly uniform and rapid, whereas as affinity is increased it results in a corresponding increase in both the time of activation (Fig. 2.9b) and the intrinsic noise in activation time (Fig. 2.9c).

Thus, the above data establishes that intrinsic noise in activation of the protease is modulated by the addition of specific solvents that appear to function by modulating the structure of IMC without altering the structure, activity, or stability of the protease (Fig. 2.8). This variation in the folded state of the IMC directly affects the binding equilibrium of IMC and hence its stochastic release (Fig. 2.7 & 2.9). It is noteworthy that such modulation in protease activation is effected without any mutations in the protease or IMC domain. Hence, the solvents outline conditions that specifically modulate the time and randomness of activation, without affecting the earlier stages of the maturation pathway.

Energetics of precursor activation:

To investigate causes for the probabilistic release of the IMC domain and the variation in the time of activation, we established the activation energy, E_a for release (Segel 1975; Sohl, Jaswal et al. 1998) (Materials and Methods). The mean time of activation of subtilisin as a function of temperature was established. The temperature dependence of the activation rate (Fig. 2.10a) fits well to non-linear Eyring equation as well as to a linear function. The parameters computed from the two functions are similar (ΔG: 17.39kcal/mol; ΔS: -45.6cal/molK; ΔH: 3.88kcal/mol; ΔCp: -0.123 kcal/molK). The ΔCp for this reaction is relatively low suggesting a small exposure in hydrophobic surface in going from the inhibited complex to the transition state. The activation energies computed from the equations for release and degradation of the IMC is ~17.3 ± 0.3 kcal mol^{-1}. Unlike molecular chaperones, IMCs can bind tightly with their cognate proteases and the affinity is similar to the best-known inhibitors (Shinde, Li et al. 1993; Li, Hu et al. 1995). To complete the free-energy diagram for precursor activation we utilized the affinity constant of IMC for Subtilisin E that was established by isothermal titration calorimetry (Fig. 2.9*a*). The K_a for this interaction was $7.34 \times 10^6\ M^{-1}$. By using the thermodynamic relation shown in Eqn. 2.2

$$\Delta G_{release} = - RT \ln (K_{eq}) \qquad \text{Equation (2.2)}$$

where $K_{eq} = k_{off}/k_{on} = 1/K_a$

we estimate that IMC binding with subtilisin stabilizes the inhibition complex by ~9.4 kcal mol^{-1} (Fig. 2.10b). This computed free energy is in close agreement with k_{on} and k_{off} rates determined using fluorescence spectroscopy for a

subtilisin homologue (Bryan, Wang et al. 1995; Ruan, Hoskins et al. 1999; Bryan 2002). It is interesting to note that the free-energy diagram for precursor maturation favors formation of a thermodynamically stable inhibition complex (Fig. 2.10b) over the active protease. Although the measured E_a signifies the net limiting energy barrier for release and degradation, since proteolysis is rapid, it may be safely assumed that the energy barrier for the stochastic release is equivalent or higher. Interestingly, the binding affinity of IMC for subtilisin increases 10-fold in the presence of glycerol (Fig. 2.9a). This in turn would further stabilize the inhibition complex there by prolonging IMC-release and degradation as evident in Fig 7a-c. Based on results, we propose that the variation in activation time arises due to the non-spontaneous nature of the release step.

IV) DISCUSSION:

Stochasticity of IMC-Release and its Significance:

Stochastic effects in gene expression play crucial roles in biological processes such as development and growth, where the initial asymmetries, amplified by feedback mechanisms, determine cell fates (Elowitz, Levine et al. 2002). The inherent noise in interactions between macromolecules allows small changes in macromolecular states to signal specific transitions. Here, using the bacterial subtilisin system we demonstrate that protease activation may be driven by a fundamentally stochastic event (Shinde and Inouye 1993). Our results show that although folding and subsequent autoprocessing of IMC-subtilisin are extremely deterministic, the release and degradation of the IMC domain is a probabilistic event that causes variations in the time of activation. Interestingly, the release process is easily stabilized or destabilized under specific conditions that affect the binding equilibrium of the IMC domain. Controlled spatial and temporal activation is well established in many proenzymes. For example cathepsins are specifically activated under the low pH conditions of the lysozyme (Jerala, Zerovnik et al. 1998; Wiederanders, Kaulmann et al. 2003). It has been suggested that in human cathepsin L and cathepsin S, this may be mediated through loss of tertiary structure and affinity in the propeptide (Maubach, Schilling et al. 1997; Jerala, Zerovnik et al. 1998). Further, furin, a eukaryotic homolog of subtilisin that folds and auto processes in the endoplasmic reticulum, becomes active in the trans golgi network upon pH shock (Anderson, Molloy et al. 2002). Using SubtilisinE, here we

demonstrate that this activation maybe inherently random. Through selection of IMCs, with specific structural characteristics, which regulate the inherent noise, nature may have selected for a system that effectively responds to external conditions. Modeling and simulation of the maturation pathway allows us to analyze the intricacies of this single molecular event using stochastic models. Since release of IMC and the subsequent trans-activation of the protease are tightly coupled events, the mathematical model allows us to monitor in detail the behavior of each event and to further evaluate hypotheses from our experimental results. Interestingly, this allows for analyses without introducing mutations that may perturb the system.

The analysis of the energetics associated with activation process establishes that the release of the inhibitory IMC-domain from the thermodynamically stable complex (RDS) is energetically unfavorable. The release of the first free protease forces the equilibrium towards activation because the IMC-domain in the inhibition complex is an excellent substrate for the free protease. This transition of the protease although energetically unfavorable, helps to ensure longevity of the protease under harsh, environmental conditions by increasing the energy barrier for unfolding (Jaswal, Sohl et al. 2002). It is noteworthy that deterministic activation of proteases observed in earlier studies of IMC-mediated folding have been facilitated by the addition of external activators such as trypsin (Sohl, Jaswal et al. 1998; Bryan 2000; Yabuta, Subbian et al. 2003).

Relation Between IMC-structure and Protease Activation:

A number of recent studies (Dyson and Wright 2002; Dyson and Wright 2002; Tompa 2002) have shown the importance of naturally occurring intrinsically unstructured proteins (IUPs) that offer distinct advantages in cellular functions, such as DNA and RNA binding (Spolar and Record 1994; Mogridge, Legault et al. 1998), transcriptional regulation (Kussie, Gorina et al. 1996; Bowers, Schaufler et al. 1999; Parker, Rivera et al. 1999), signal transduction (Kim, Kakalis et al. 2000; Huber and Weis 2001) and cell-cycle regulation (Kriwacki, Hengst et al. 1996; Pavletich 1999). Their increased flexibility enables binding to multiple targets, better affinity and rapid responses in cellular signaling. IMC of Subtilisin appears to behave as an intrinsically unstructured protein. Further many propeptides of proteases, characterized in the literature, are either completely unfolded or exhibit limited secondary structure (Jerala, Zerovnik et al. 1998; Marie-Claire, Yabuta et al. 2001; Wiederanders, Kaulmann et al. 2003; Yabuta, Subbian et al. 2003). We propose that the intrinsic disorder in IMC structure plays an important role in ensuring the synergy of the maturation pathway and enables IMCs to effectively switch from a chaperone to an inhibitor and subsequently to a protease substrate. Release of IMC coincides with the transition of the protease from a thermodynamically stable inhibited state to a kinetically trapped native state. The intrinsically unstructured IMC may aid this transition. Hence, while the kinetically trapped native states ensure proteolytic stability under harsh extracellular conditions (Jaswal, Sohl et al. 2002), having intrinsically unstructured IMC domains that effectively

modulate protease-activation provides a mechanism for proteases to ensure synergy and precision of their activation.

Industrial Importance:

The role of IMCs in regulating the protease activation can be exploited in industrial applications of subtilisin. A number of industries, including detergent, leather, food, sewage treatment, silk, and industrial waste management utilize subtilisins as classical non-specific proteases (Cheng, Hu et al. 1995; Wolff, Showell et al. 1996; Kumar and Takagi 1999; Nekliudov, Ivankin et al. 2000; Gupta, Beg et al. 2002). A common problem faced in industrial applications is ensuring the shelf life of the protease. Our finding that modulation of the conformational dynamics of the IMC can effectively control the temporal activation of subtilisins provides a novel approach to increasing shelf life of these enzymes. The isolation of specific solvent conditions, to super-stabilize the inhibited complex (Fig. 2.7*a-c*) offers a distinct method to keep the protease in an inactive state for prolonged periods. Activators that destabilize the IMC-domain can be employed to trigger activation at will.

V) MATERIALS AND METHODS:

Rapid-dilution folding technique:

Proteins were expressed and purified as described earlier (Shinde and Inouye 1995; Jain, Shinde et al. 1998; Yabuta, Takagi et al. 2001). 20 μl of denatured prosubtilsin (100 μM in 6M GdnHCl, pH 4.8) was rapidly mixed in 19980 μl of the refolding buffer (50 mM MES-NaOH, pH 6.5, 0.5 M $(NH_4)_2SO_4$, 1mM $CaCl_2$) at 23°C with 0.5 mM synthetic substrate (N-succ-AAPF-pNA). After 15 min the stirring was stopped and 200 μl aliquots of the reaction mixture were transferred to a 96-well microplate, after 30 min of folding initiation. Subtilisin activity was monitored by release of *p*-nitroanilide measured at 405 nm in a microplate reader (Yabuta, Takagi et al. 2001; Yabuta, Subbian et al. 2002; Yabuta, Subbian et al. 2003) maintained at the desired temperature. The time of activation was calculated from the X-axis intercept of the 'none-to-all' transition curve as described earlier (Fu, Inouye et al. 2000; Inouye, Fu et al. 2001).

Amount of active subtilisin:

Precursors were folded at 23 °C as described above, however without the synthetic substrate. The reaction was allowed to proceed on the microplate at the desired temperature. After 24 hrs of incubation, synthetic substrate was added to a final concentration of 0.5 mM and the velocity of substrate hydrolysis in each well was measured. Care was taken to ensure that substrate concentration was not rate limiting during velocity measurements. Under these

conditions velocity of substrate cleavage was proportional to enzyme concentration and 80 nM of mature subtilisin E gave ~20 mOD min^{-1}.

Proteolysis of IMC:

IMC (23μM) was added to 120nM of subtilisin to maintain a complex with excess IMC. 100μl aliquots were removed at various times, and the reaction was stopped using trichloroacetic acid precipitation. The samples were subjected to SDS-PAGE and analyzed using densitometry. Additionally the activity of subtilisin was monitored over time as described above.

Mathematical Modeling:

Mathematical modeling of the maturation pathway was done using Cellware version 1.2. Cellware is a multi-platform modeling software with a convenient user interface and extensive algorithms library (Dhar, Meng et al. 2004). The model used for our simulation is represented in Fig.4a. Known intermediates along the maturation pathway were represented as individual molecular species. Reactions between the species were represented as reversible/irreversible reactions defined at the rates denoted. All rates used for simulation were based on experimental results. The models were simulated using the Gillespie stochastic algorithm available in the package. The package allows for varying the particle size in a defined initial volume. Due to the computational complexity of stochastic simulations, the models were simulated using higher initial concentrations such that activation occurs in a feasible time scale. To establish the evolution of folded, autoprocessed and mature species

between iterations, the simulations were terminated at specific times, the data on the individual species were parsed and analyzed using Sigma-Plot. To establish the effect of exogenously added protease, the pathway was modified to include the presence of a activator molecule. The model was simulated under similar conditions as above but with varying ratios of the activator molecule to IMC:Subtilisin complex. We established the effect of concentration by varying the initial particle size (U) from 100 to 50,000. The time of activation under each condition was parsed and data plotted using Sigma Plot. Altering the k_{on} and k_{off} rates established the effect of IMC binding equilibrium on the time of activation. For graphs that depict both the experimental and simulation data, the concentration and time of activation of the simulated data was normalized to experimental scales.

Co-Solvent induced modulation of precursor activation:

Precursors (200 nM) were folded (as described above) and aliquoted (100 μl each well) onto the microplate. Activators (2X conc.) were prepared in the folding buffer and 100 μl were added to each well at the desired time and mixed. The final concentration of the activators that expedite activation are SDS - 0.01 %; salt-shock-0.25 M $(NH_4)_2SO_4$; active subtilisin - 2.5 nM, pH-shock - Tris-HCl (50 mM, pH 8.5). While 1 μl of SDS (1 %) in each well was sufficient to activate the precursor, a 100 μl volume was chosen to enable complete mixing. The aliquots with the added activators were incubated on a microplate reader at the desired temperature and the emergence of activity as a function of time was monitored at 405 nm.

100 μl of glycerol (20 %), $(NH_4)_2SO_4$ (2.0 M) and the combination of glycerol- $(NH_4)_2SO_4$ serve as stabilizers that prolong activation. Stabilizers were added 30 min after folding initiation to allow complete folding and autoprocessing (Yabuta, Takagi et al. 2001; Yabuta, Subbian et al. 2002).The aliquots taken on the microplate were incubated at the desired temperature and the activation of pro-subtilsin was monitored. 1 μl of 2 % SDS was added as an activator at different time points to induce uniform activation.

Circular dichroism measurements:

CD measurements were performed on an automated AVIV 215 spectrophotometer maintained at 25°C and spectra were taken between 190 to 260 nm as described earlier (Shinde, Li et al. 1993; Shinde, Liu et al. 1997; Fu, Inouye et al. 2000; Yabuta, Subbian et al. 2002; Yabuta, Subbian et al. 2003). Protein concentrations were maintained between 0.25 mg to 0.4 mg ml^{-1}. A 1 mm path-length cuvette was used to measure spectra except in case of the IMC-domain, where a 0.5 mm path-length cuvette was used. The spectra depicted in Fig. 2.8*a* represent averages of 3 independent scans.

Proteolytic stability of the precursor:

A cleaved but proteolytically inactive complex (Jain, Shinde et al. 1998) (~1 μM) was prepared as described elsewhere (Hu, Haghjoo et al. 1996). The sample was maintained under normal and stabilized conditions (10% glycerol) and 1/10 the amount of TPCK-treated trypsin was added. 50μl of the complex were removed at fixed time intervals and the reaction was stopped by TCA

precipitation (Yabuta, Takagi et al. 2001; Yabuta, Subbian et al. 2002). The samples were analyzed using SDS-PAGE and the extent of IMC-degradation was quantitated using gel-scanning densitometry (Yabuta, Takagi et al. 2001; Yabuta, Subbian et al. 2002) .

Activation Energy:

The activation rate at different temperatures was calculated using a fixed amount of precursor (200 nM) and the mean time of activation at different temperatures was obtained from the none-to-all activation curves as described above. The temperature dependence of activation rates were fit to linear Arrhenius relation (Segel 1975; Sohl, Jaswal et al. 1998) for temperature dependence as well as to non-linear kinetics (van Nuland, Chiti et al. 1998) to obtain Ea for activation.

Affinity Constant by Isothermal Titration Calorimetry (ITC):

ITC experiments were performed with a Microcal VP-ITC titration calorimeter. For ITC measurements, inactive $Ser_{221}Cys$ active site variant of subtilisin was used (Jain, Shinde et al. 1998; Yabuta, Takagi et al. 2001; Yabuta, Subbian et al. 2003). Both the inactive mature domain and the IMC were dialyzed against buffer containing 50mM MES (pH 6.5), 0.5M Ammonium Sulphate, 1mM $CaCl_2$ and 2mM BME. A 5μM solution of Mat $S_{221}C$ was taken in the sample cell and a 28μM solution of IMC was loaded in the injection syringe. The sample cell was maintained at 23 °C. For measuring the heat of binding, 5μl injections of IMC were diluted into the sample cell with 250s spacing. For

control titrations, the propeptide was diluted into dialysis buffer. Titration data was analyzed using the Origin 5.0 software and K_a was estimated. For the binding constant in the presence of glycerol, the above experiment was repeated with 10% glycerol added to the above buffer.

VI) FIGURES

Fig. 2.1. Maturation pathway of IMC-subtilisin. **(a)** Schematic representation of IMC-subtilisin maturation. The rate-limiting step (RDS) is shown in red and trans-autocatalytic activation is shown in green **(b)** SDS-PAGE of aliquots taken over time from a maturation reaction. The precursor (36-kDa) undergoes autoprocessing to give IMC (8-kDa) and mature subtilisin (28-kDa) fragments that exist as a complex. **(c)** The amounts of precursor (filled circles), mature (open circles) and IMC (open diamond) were estimated by densitometry. Folding of the precursor (inset) saturates in ~5mins and autoprocessing is complete in ~30mins.

Fig. 2.1

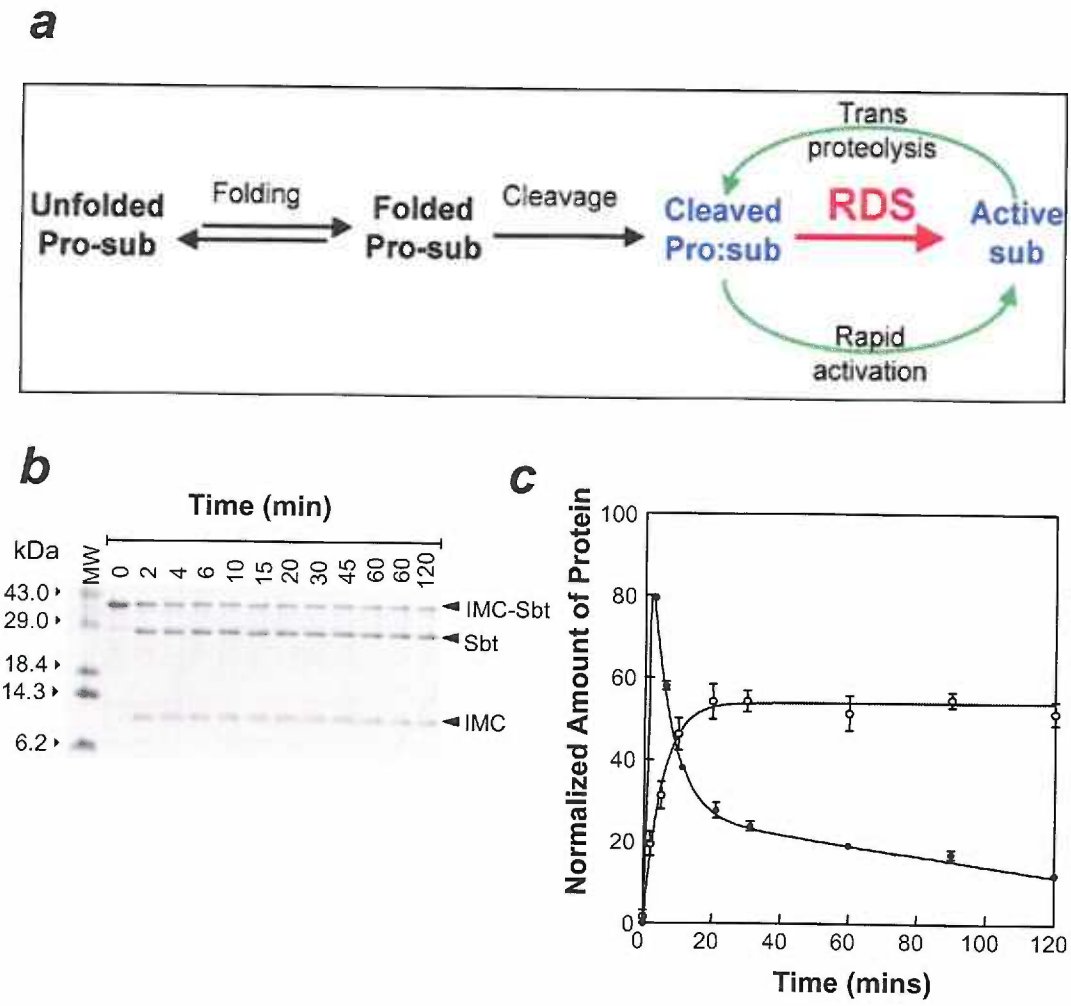

a
Unfolded Pro-sub
Folding
Folded Pro-sub
Cleavage
Cleaved Pro:sub
RDS
Active sub
Trans proteolysis
Rapid activation
b
Time (min)
kDa
MW
0 2 4 6 10 15 20 30 45 60 60 120
43.0
29.0
18.4
14.3
6.2
IMC-Sbt
Sbt
IMC
c
Normalized Amount of Protein
0 20 40 60 80 100
0 20 40 60 80 100 120
Time (mins)

Fig. 2.2. Variation in the time of activation of IMC-subtilisin. **(a)** Non-deterministic activation of multiple aliquots from a single folding reaction captured at different time intervals using a chromogenic substrate. **(b)** Enzyme activity as a function of time, for 12 randomly selected samples from a maturation reaction. Inset depicts SDS-PAGE of IMC-subtilisin maturation with active (+) and inactive (-) aliquots. Inactive aliquots differ by the presence of the IMC domain. **(c)** Frequency distribution of the number of active aliquots in a microplate as a function of time. Activation follows a poisson's distribution with a mean of ~235min **(d)** Subtilisin yield (open circles, Left Axis) and average activation time (filled squares, Right Axis) as a function of precursor concentration. The inset shows an SDS-PAGE gel that depicts the extent of autoprocessing for increasing precursor concentrations monitored 30mins after folding initiation. The ratio of mature/precursor is constant throughout the concentrations studied establishing that autoprocessing is concentration dependent.

Fig. 2.2

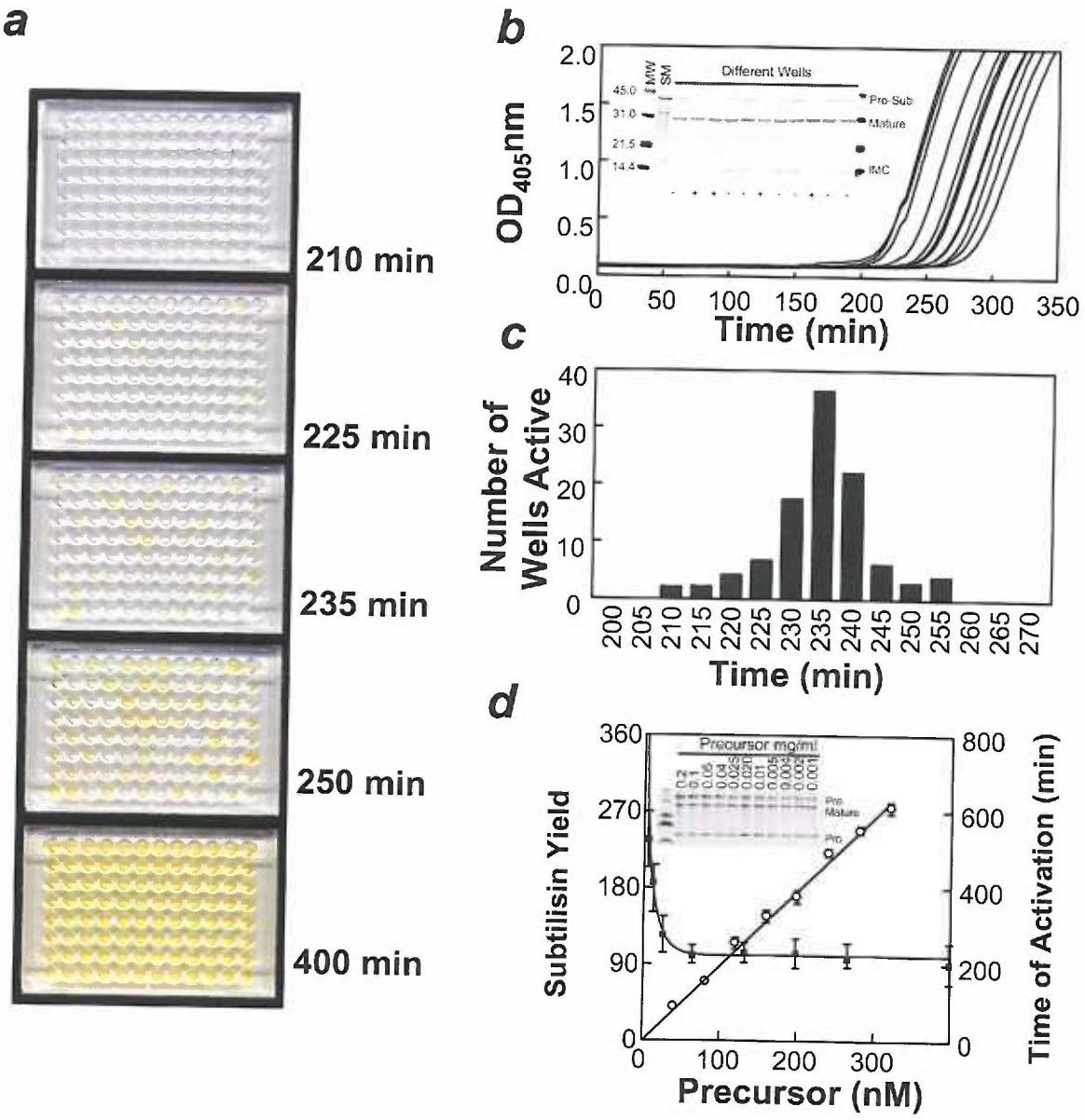

a
210 min
225 min
235 min
250 min
400 min
b
OD_{405}nm
Time (min)
MW
SM
Different Wells
Pro-Sub
Mature
IMC
45.0
31.0
21.5
14.4
c
Number of Wells Active
Time (min)
d
Subtilisin Yield
Time of Activation (min)
Precursor (nM)
Precursor mg/ml
Pro
Mature

Fig. 2.3. Rate of proteolysis of the IMC in trans. (a) Degradation of a 1:20 complex of subtilisin:IMC was followed by both Tricine-SDS PAGE analysis (*inset* and filled circles in graph) and activity (broken line). Degradation of IMC occurs in two distinct phases. The first rapid phase denotes the degradation of free IMC and the second slow phase corresponds to the degradation of IMC in complex. Increase in protease activity measured using a chromogenic substrate coincides with complete degradation of IMC. **(b)** Modulation of activation time by exogenously added active protease. Addition of active protease exponentially reduces the time of activation. In controls without exogenous protease, activation occurs at ~240 mins (*inset,* filled circles). Activation time is reduced by ~50% (*inset,* open circles) by addition of 20% active protease (*inset,* filled triangles).

Fig. 2.3

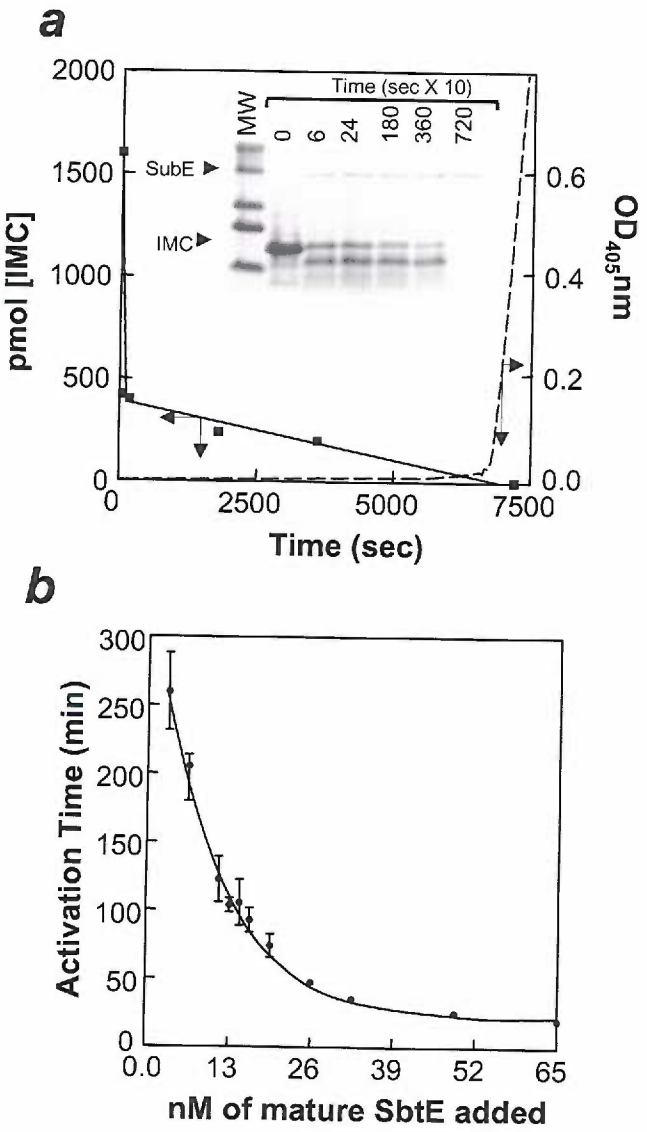

a
2000
1500
1000
500
0
pmol [IMC]
Time (sec X 10)
MW
0
6
24
180
360
720
SubE
IMC
0.6
0.4
0.2
0.0
OD405nm
0
2500
5000
7500
Time (sec)
b
300
250
200
150
100
50
0
Activation Time (min)
0.0
13
26
39
52
65
nM of mature SbtE added

Fig. 2.4. Mathematical modeling of IMC-Subtilisin maturation. (a) Maturation pathway used for simulation. **U**-Unfolded IMC-Subtilisin; **F**-Folded but unautoprocessed IMC-Subtilisin; **M**-Mature Subtilisin; **Pf**-Folded IMC; **Pu**-Unfolded IMC; **D**-Degraded products; $\mathbf{K_f}$-rate of folding; $\mathbf{K_{auto}}$-rate of autoprocessing; $\mathbf{K_a}$-binding affinity; $\mathbf{K_{deg1}}$-rate of proteolysis of IMC; $\mathbf{K_{deg2}}$-rate of trans-proteolysis of complex by free M; $\mathbf{K_{deg3}}$-rate of degradation of F by free M; $\mathbf{K_{deg4}}$-rate of degradation of U by free M. Folding and binding equilibrium of IMC were represented as reversible reactions. Folding and Autoprocessing are deterministic events (depicted in blue). Release and degradation (depicted in red) of IMC from its complex with the protease (M:Pf) results in active protease (M). Active, free protease (M) can proteolyze IMC bound to another protease (M:Pf), unfolded (U) and folded but unautoprocessed (F) species (depicted in green). **(b)** *In silico* evolution of unfolded (U, red), folded (F, blue), autoprocessed (M:Pf, green) and mature (M, black) species.

Fig. 2.4

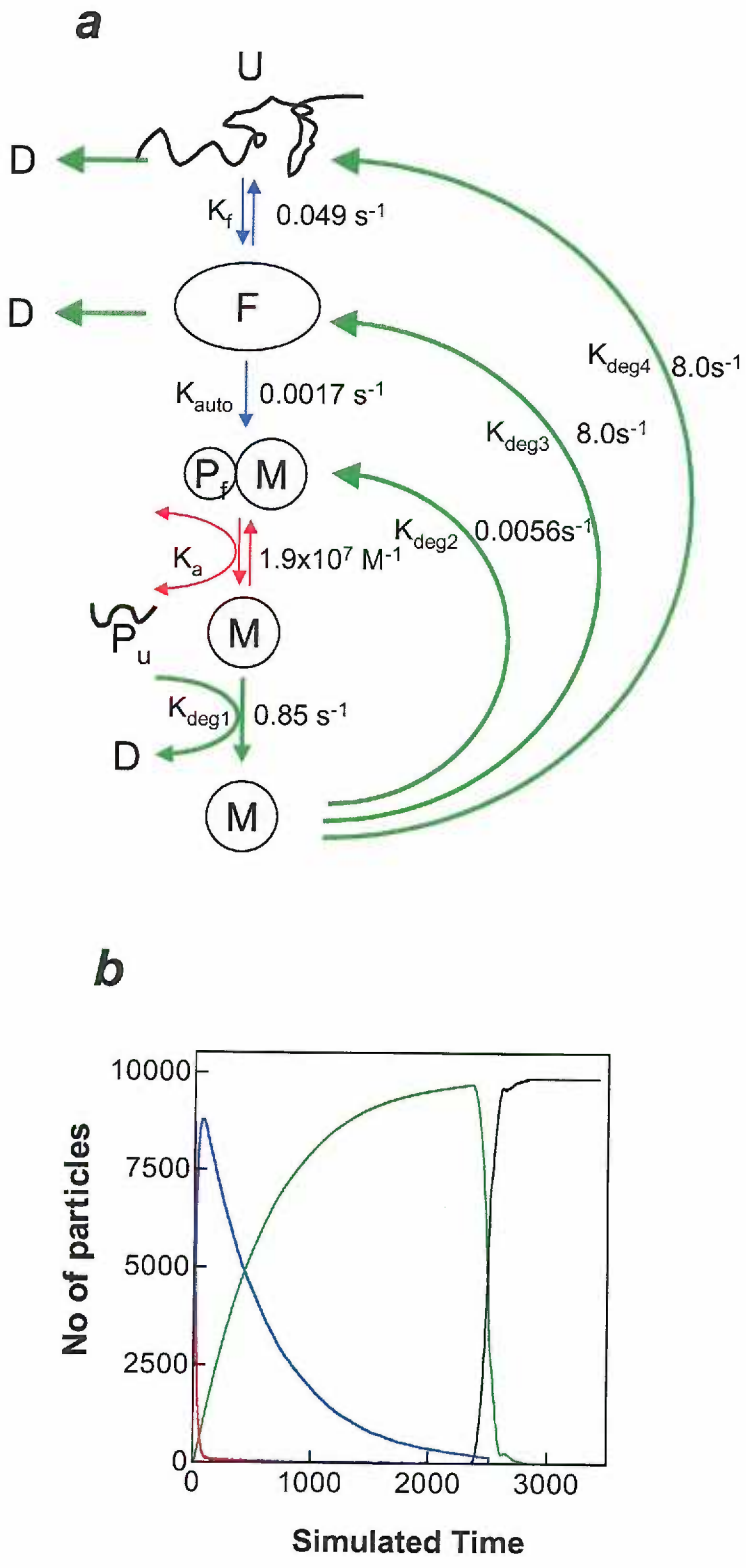

a
U
D
K_f 0.049 s^{-1}
D
F
K_{deg4} 8.0s^{-1}
K_{auto} 0.0017 s^{-1}
K_{deg3} 8.0s^{-1}
P_f M
K_{deg2} 0.0056s^{-1}
K_a 1.9x10^7 M^{-1}
P_u M
K_{deg1} 0.85 s^{-1}
D
M
b
No of particles
10000
7500
5000
2500
0
0
1000
2000
3000
Simulated Time

Fig. 2.5. Simulation of IMC-Subtilisin maturation using stochastic algorithm. (a,b &) Representative curves for evolution of folded (a), autoprocessed (b) and mature (c) for 20 randomly selected sample iterations. Folding and autoprocessing are deterministic. The release of mature protease is stochastic. **(d)** Normalized time of activation for 1000 iterations simulated using stochastic algorithm for an initial particle size (U=10000) follows a Poisson's distribution. **(e)** Variation of time of activation with initial particle size. The experimental data (filled circles) together with simulated activation time (open circles) normalized for differences in concentration. **(f)** Variation of activation time with exogenous addition of active protease. The simulated data (open circles) normalized for differences in concentration follows identical behavior as the experimental results (filled circles).

Fig. 2.5

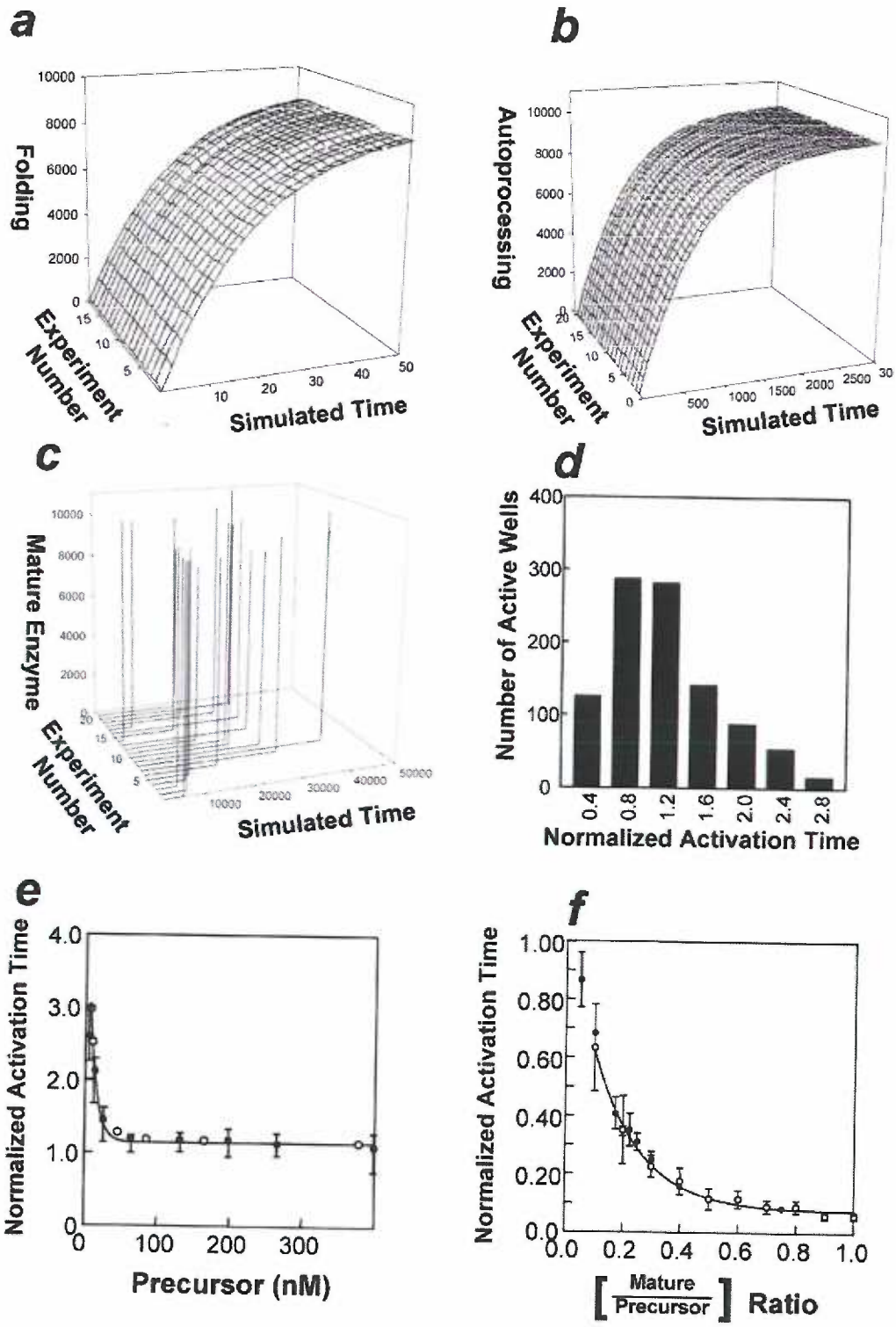

a
Folding
Experiment Number
Simulated Time
b
Autoprocessing
Experiment Number
Simulated Time
c
Mature Enzyme
Experiment Number
Simulated Time
d
Number of Active Wells
Normalized Activation Time
e
Normalized Activation Time
Precursor (nM)
f
Normalized Activation Time
[Mature / Precursor] Ratio

Fig.2.6. Activator induced IMC-subtilisin release. (a) Deterministic activation induced by external additives (activators). The leftmost panel depicts the control samples that show intrinsic variation in activation times, while the other panels depict the immediate increase in absorbance seen due to activators added at different times - Solid line:0.01% SDS; dotted line:2.5nM active subtilisin; dashed line: 50mM Tris-HCl, pH 8.5. The solid black arrow indicates the time of addition (60,120,180 mins). **(b)** Effect of activators monitored using SDS-PAGE. Lanes 7-10 represents aliquots taken from the control at indicated time points. The O/N sample shows the activation of Prosubtilisin as seen by complete degradation of the IMC and precursor. Lane 2 depicts the starting material. Lanes 4-6 depict aliquots that were taken 1hr after folding initiation and to which activators (subtilisin:lane 4; Tris-HCl:lane 5; SDS:lane 6) were added. Addition of activators induces immediate degradation of precursor and IMC-domains. Lane 3 represents the amount of subtilisin added as an activator to lane 4. **(c)** Subtilisin yield when activators (SDS:solid circles; subtilisin:open circles; Tris-HCl:gray circles) are added at different times. Amount of active subtilisin produced is independent of the activator and the time of addition.

Fig. 2.6

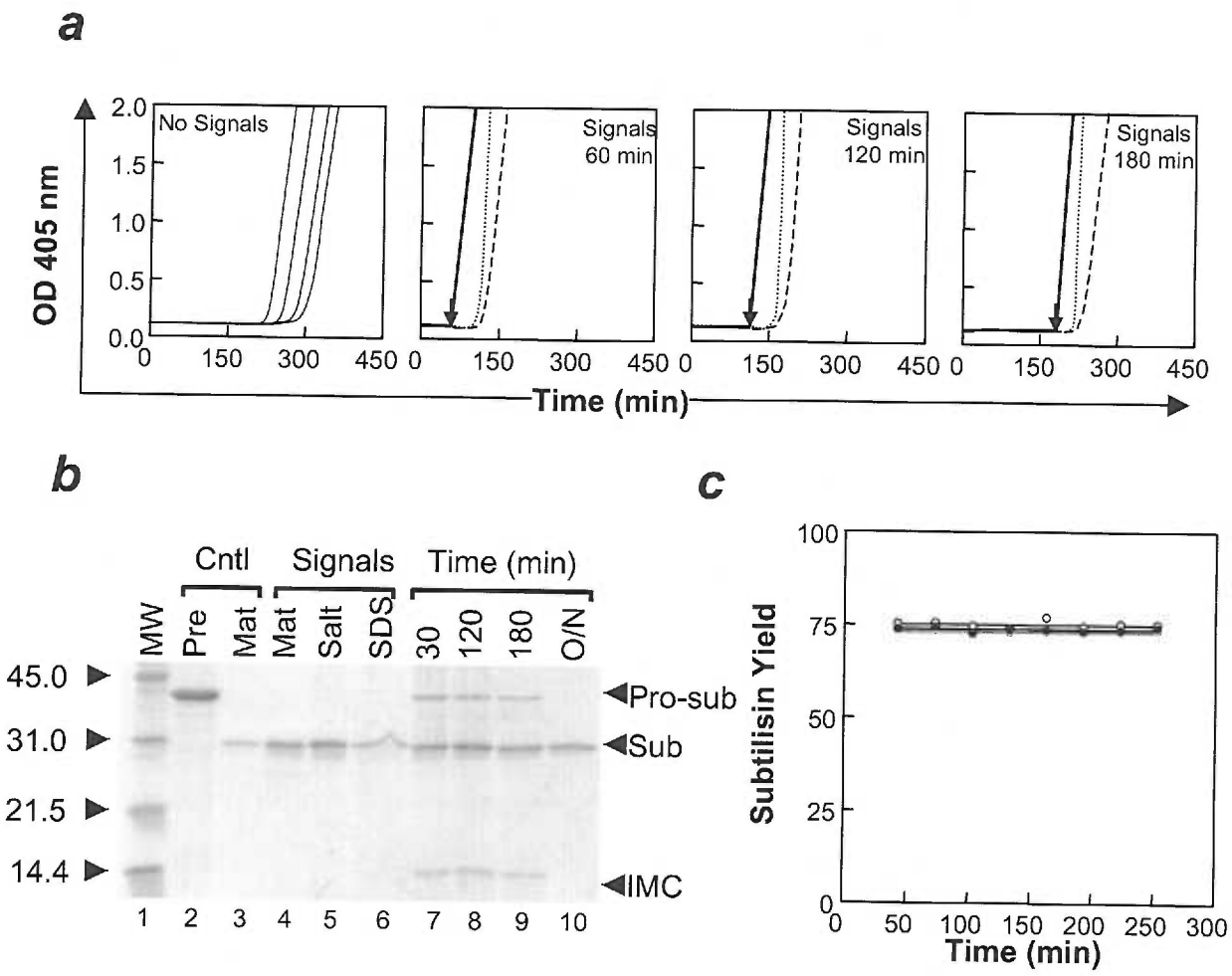

a
OD 405 nm
2.0
1.5
1.0
0.5
0.0
No Signals
Signals 60 min
Signals 120 min
Signals 180 min
0
150
300
450
Time (min)
b
Cntl
Signals
Time (min)
MW
Pre
Mat
Mat
Salt
SDS
30
120
180
O/N
45.0
31.0
21.5
14.4
Pro-sub
Sub
IMC
1 2 3 4 5 6 7 8 9 10
c
Subtilisin Yield
100
75
50
25
0
0 50 100 150 200 250 300
Time (min)

Fig. 2.7. Solvent-induced conformational stabilization of IMC. (a) Frequency distribution of the number of active aliquots (black bars) as a function of time, under normal and glycerol conditions. The mean time of activation and variations in time of activation increase in the presence of glycerol. Addition of 0.01% SDS at 120mins to both normal and glycerol conditions instantaneously activates all wells (Grey bar). **(b)** Effect of stabilizing conditions monitored using SDS-PAGE. Lanes 2-5 depict aliquots taken at 2hr, and Lanes 7-10 depict aliquots taken at 8hr, after folding initiation. Under normal conditions intrinsic activation is seen at 8hr (Lane 7). With glycerol the IMC remains bound at both 2 and 8hr (Lane3 and 8). However activation is triggered under both conditions when the IMC is destabilized through the addition of SDS (Lanes 4,5,9,10) ***c)*** Averaged activation under normal and stabilized conditions (glycerol, salt and glycerol-salt). Intrinsic (dotted lines) and induced (solid lines) activation due to activators (0.01% SDS) added at times indicated.

Fig. 2.7

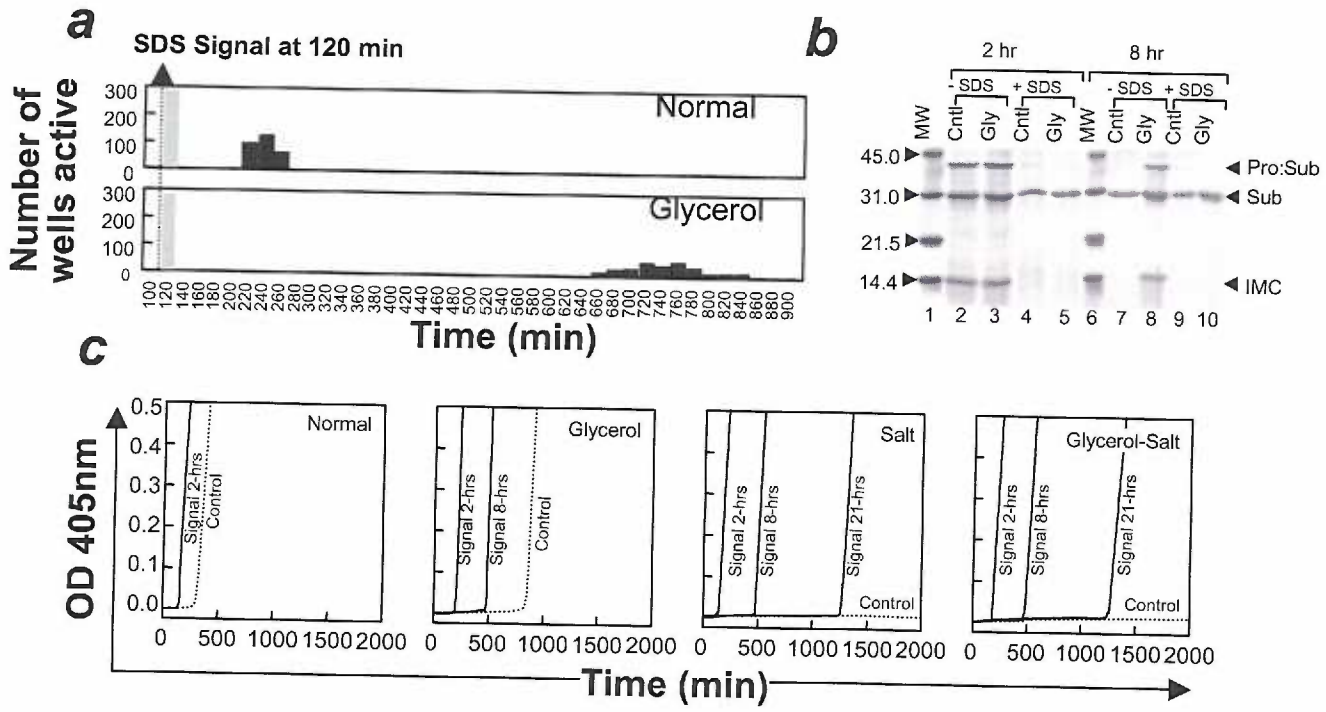

a
SDS Signal at 120 min
Number of wells active
Normal
Glycerol
Time (min)
b
2 hr
8 hr
- SDS
+ SDS
MW
Cntl
Gly
45.0
31.0
21.5
14.4
Pro:Sub
Sub
IMC
1 2 3 4 5 6 7 8 9 10
c
OD 405nm
Normal
Glycerol
Salt
Glycerol-Salt
Signal 2-hrs
Signal 8-hrs
Signal 21-hrs
Control
Time (min)

Fig. 2.8. Effect of glycerol on IMC structure and dynamics (a) Circular dichroism spectra for inhibition complex (solid line) and mature (dotted line), and differential spectra of IMC- in complex (broken line), under normal (grey) and stabilized (black) conditions. The spectra for the IMC-domain in complex were obtained from the difference between the inhibition complex and mature subtilisin. **(b)** Circular dichroism spectra for isolated IMC domain (solid line) under normal (grey) and stabilized (black) conditions. **(c)** Proteolytic stability of the IMC-domain in the inhibition complex under normal (filled circles) and stabilized (open circles) conditions against TPCK-treated trypsin. **(d)** Simulated curves for trans-proteolysis of the IMC domain in normal (filled triangles) and glycerol (open triangles) conditions.

Fig. 2.8

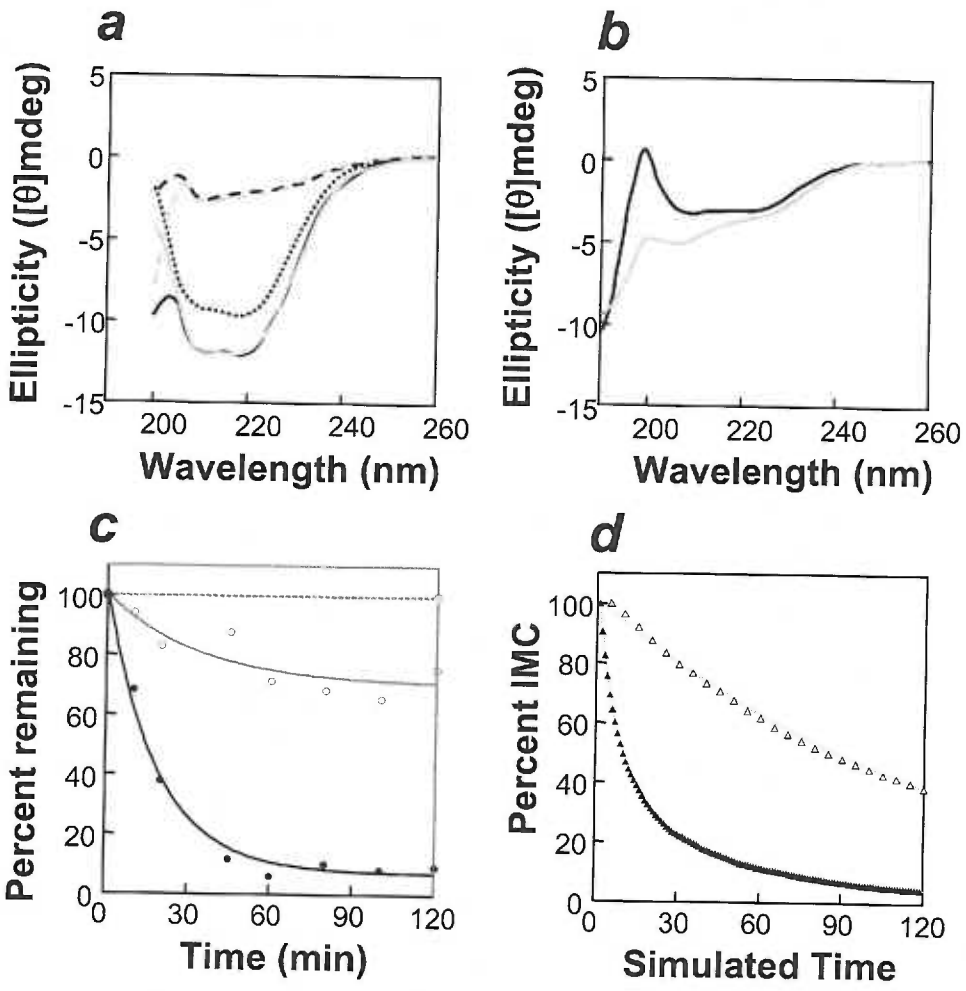

a
Ellipticity ([θ]mdeg)
5
0
-5
-10
-15
200
220
240
260
Wavelength (nm)
b
Ellipticity ([θ]mdeg)
5
0
-5
-10
-15
200
220
240
260
Wavelength (nm)
c
Percent remaining
100
80
60
40
20
0
0
30
60
90
120
Time (min)
d
Percent IMC
100
80
60
40
20
0
0
30
60
90
120
Simulated Time

Fig. 2.9. Binding equilibrium regulates protease activation. (a) Titration curves for binding of IMC, and IMC in 10% glycerol to subtilisin mature domain. The heats of binding for successive additions of IMC are plotted against the IMC/subtilisin molar ratio. The estimated K_a for each is depicted. **(b)** Dependence of simulated activation time on the K_a of IMC for its protease. **(c)** Distribution of intrinsic noise in simulated activation time with K_a.

Fig. 2.9

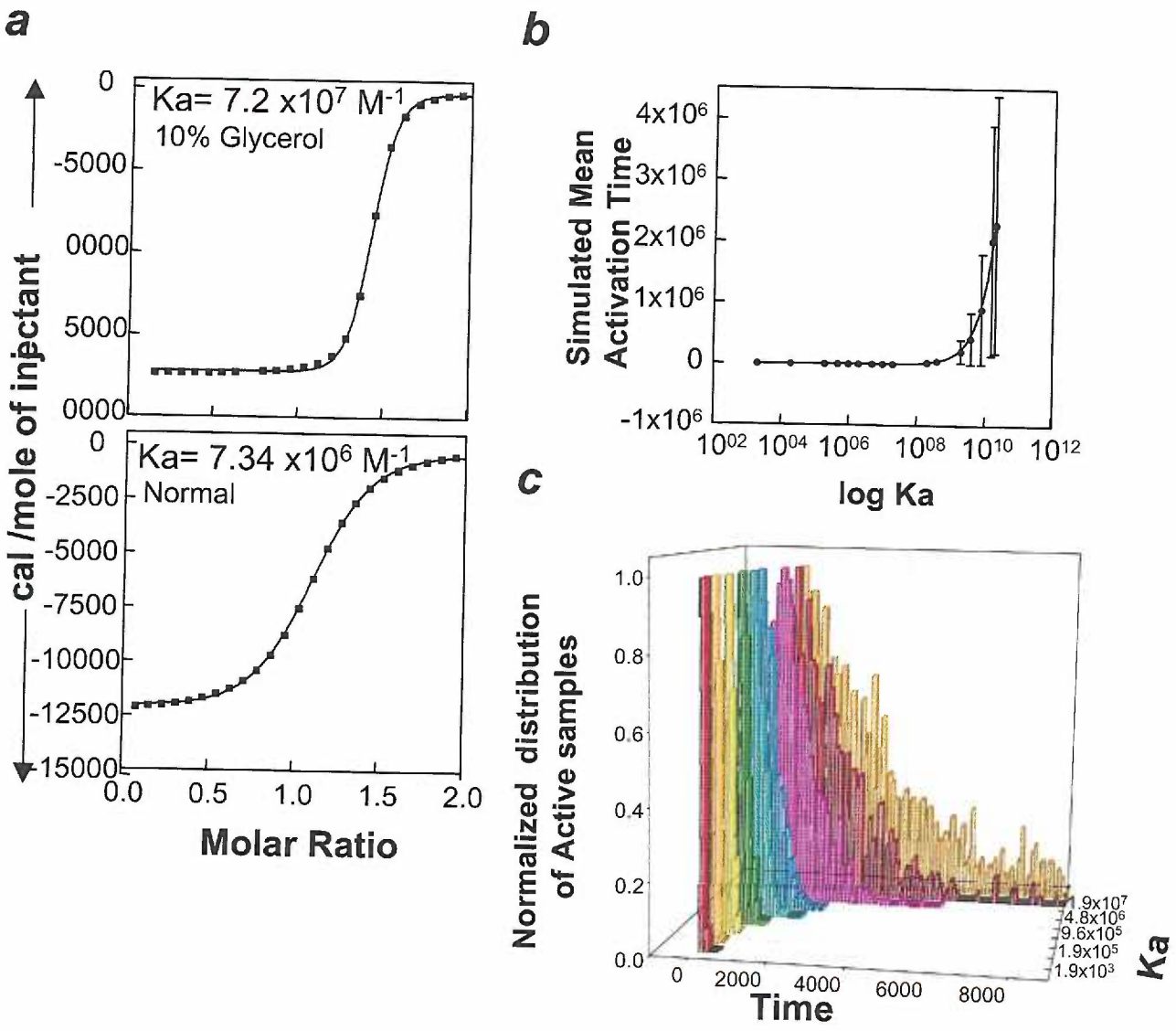
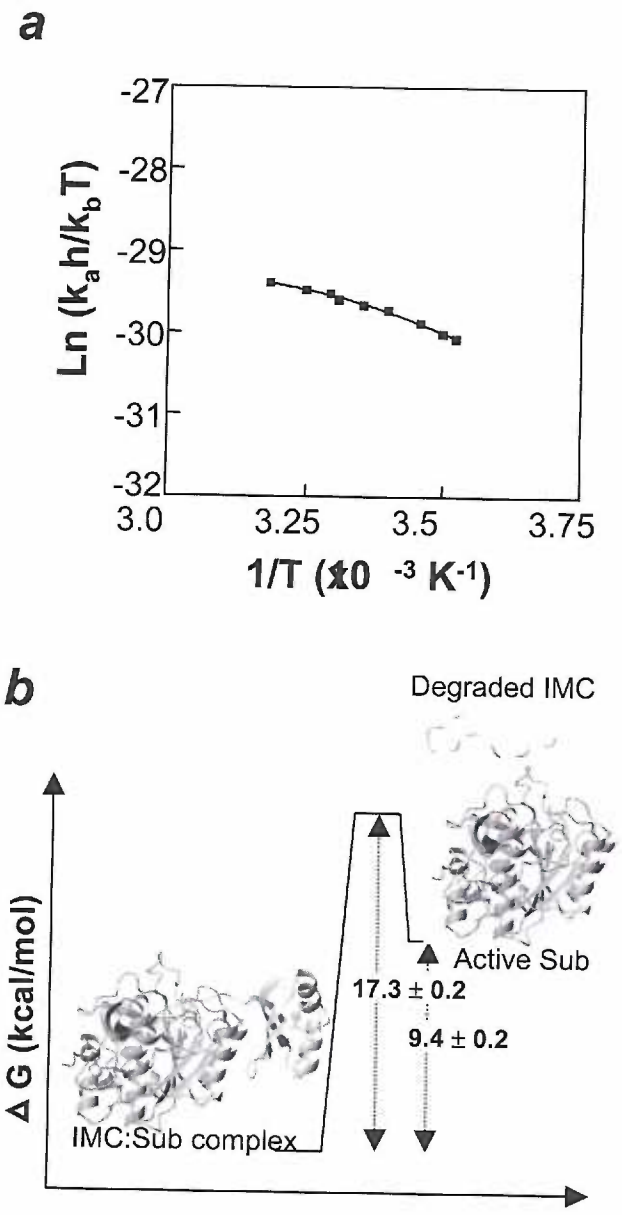

a
Ka= 7.2 x10^7 M^-1
10% Glycerol
Ka= 7.34 x10^6 M^-1
Normal
cal /mole of injectant
Molar Ratio
b
Simulated Mean Activation Time
log Ka
c
Normalized distribution of Active samples
Time
Ka
1.9x10^7
4.8x10^6
9.6x10^5
1.9x10^5
1.9x10^3

Fig. 2.10. Energy landscape for protease maturation. (a) Arrhenius plots for activation rate. **(b)** Free energy diagram of protease activation. Activation requires transition of the protease from its thermodynamically stable inhibition complex to a kinetically trapped active form. Variation in activation time occurs due to the non-spontaneous nature of the RDS, namely release and degradation of the inhibitory IMC-domain.

Fig. 2.10


a
-27
-28
-29
-30
-31
-32
Ln ($k_a h/k_b T$)
3.0
3.25
3.5
3.75
1/T (x10 $^{-3}$ K^{-1})
b
Degraded IMC
Active Sub
17.3 ± 0.2
9.4 ± 0.2
IMC:Sub complex
Δ G (kcal/mol)

CHAPTER 3

ENERGY LANDSCAPE FOR UNIMOLECULAR PRO-SUBTILISIN E MATURATION

Ezhilkani Subbian, Yukihiro Yabuta[1] and Ujwal Shinde

Department of Biochemistry and Molecular Biology,

Oregon Health and Science University,

3181 SW Sam Jackson Park Road, Portland, OR 97239, USA

[1]Current Address: RIKEN Center for Developmental Biology, Kobe, Japan

(Manuscript in preparation)

I) ABSTRACT:

Efficient and precise assembly of polypeptides into their functional native states is critical for normal cellular processes. Understanding how a specific structure is encoded in the polypeptide sequence and what drives the structural progression to the native state is the first step towards curing of folding diseases and for design of novel proteins. In this manuscript, we investigate the elementary steps in the propeptide mediated maturation of Subtilisin E, a bacterial serine protease, and attempt to define its folding energy landscape. Our studies demonstrate that the initial structural acquisition of Pro-SbtE is driven by a rapid hydrophobic collapse and occurs through the formation of a molten globule-like intermediate state. Furthermore, induction of structure within the propeptide stabilizes both the molten globule-like folding intermediate and the native state, and appears to expedite initial stages of folding, purely through thermodynamic stabilization of the folded state. However, this induced structure is detrimental to the activation of the precursor and dramatically delays propeptide release, since this stage requires some degree of unfolding. The efficiency of autoprocessing, at higher temperatures, is also lowered by the structured propeptide although the activation barrier to autoprocessing is not significantly affected. Hence, based on our studies, we establish that the intrinsically unstructured propeptide of SbtE is optimized for a synergistic maturation process and not for faster folding kinetics.

II) INTRODUCTION:

Protein folding is one of the simplest and most fundamental examples of biological self-assembly. Understanding the principles of protein folding will not only help us outline the factors that drive this process, but is also the first step to cure folding diseases, and to design novel proteins. The broad question of folding is driven towards deciphering how a specific structure is encoded in the polypeptide sequence, and how the polypeptide attains this native state in a limited time (Alm, Morozov et al. 2002; Dobson 2004). Extensive studies have led to the new view of folding, which postulates that the native state is attained by a stochastic search of all possible conformations, that is aided by the stability of sub-structures that are formed (Dobson 2003; Onuchic and Wolynes 2004). However, such studies have been based on small proteins (≤00 residues) that fold with fast kinetics, and in the absence of stable intermediate states (Plaxco, Simons et al. 2000; Dobson 2003). On the other hand, the average biological molecule is large (~300 amino acids) and such proteins generally fold through the formation of stable folding intermediate states (Kuwajima 1989). Stable intermediates can provide valuable information about the development of the overall fold from a linear polypeptide, and can also help examine whether their existence facilitates the process, or serves as kinetic traps that impede folding (Shastry and Udgaonkar 1995; Matagne, Jamin et al. 2000; Nishimura C, Prytulla S et al. 2000). In this manuscript we attempt to define the folding energy landscape of the propeptide-dependent maturation of subtilisin E (SbtE) and analyze how stabilization of intermediates affects this landscape.

SbtE is a calcium-dependent extracellular serine protease that is secreted as a precursor (Pro-SbtE) and is an established model for understanding propeptide-dependent folding (Eder, Rheinnecker et al. 1993; Shinde and Inouye 2000; Bryan 2002; Yabuta, Subbian et al. 2003). In absence of its 77-residue propeptide, SbtE (275-residue) folds to a stable, molten-globule intermediate that is separated from its kinetically-trapped native state by a high-energy barrier (Ikemura, Takagi et al. 1987; Eder, Rheinnecker et al. 1993). Spontaneous conversion to the native state is slow ($T_{½}$ 500 yrs), but is catalyzed ~10^6-fold by the propeptide that functions as a dedicated intramolecular chaperone and lowers this high-energy barrier (Eder, Rheinnecker et al. 1993). However, the propeptide remains tightly associated with the protease and subsequent degradation of the propeptide releases enzymatic activity (Li, Hu et al. 1995; Yabuta, Takagi et al. 2001). Degradation of the propeptide enables the uncoupling of folding and unfolding pathways, and stabilizes the native state by increasing the transition state for the unfolding reaction (Eder, Rheinnecker et al. 1993; Bryan 2002). Furthermore, SbtE is a member of a large family of serine proteases termed subtilases (Siezen and Leunissen 1997), most members of which fold using propeptide-dependent mechanisms (Shinde and Inouye 1993; Anderson, Molloy et al. 2002; Bryan 2002). Similar mechanisms are now well established in numerous prokaryotic and eukaryotic proteases and non-proteases (Lee, Koike et al. 1994; Winther, Sorensen et al. 1994; Shinde and Inouye 2000; Anderson, Molloy et al. 2002). Much of our comprehension of the thermodynamics and kinetics of propeptide-mediated folding has however emerged through detailed analysis of bacterial

extracellular proteases (Bryan 2002; Jaswal, Sohl et al. 2002). Understanding propeptide mediated folding mechanisms not only give interesting insights into the energy landscape of proteins, but also help to understand the role of intrinsic chaperones in the overall folding process. Moreover these studies help elucidate why nature selects between chaperone dependent and independent folding pathways.

Studies pertaining to the folding kinetics and thermodynamics of SbtE have been done in a bimolecular reaction by the addition of the propeptide *in trans* (Bryan 2002). Such studies are limited by the formation of the initial propeptide-subtilisin complex (Bryan, Alexander et al. 1992) and do not provide information as to why propeptide-independent folding is slow, or how the propeptide helps to lower the transition barrier to folding. Furthermore, the biological folding reaction is clearly unimolecular. Here, we present a detailed characterization of the folding mechanism and energy landscape of unimolecular, propeptide-mediated maturation of Pro-SbtE, and establish how properties of the propeptide affect this landscape. Our studies demonstrate that a rapid hydrophobic collapse precedes acquisition of tertiary structure during the folding of Pro-SbtE. Furthermore, induction of structure within the propeptide stabilizes both the hydrophobic molten-globule intermediate and the native state and appears to expedite initial stages of folding, purely through thermodynamic stabilization of the folded state. It is interesting to note that the induced structure does not affect the activation energies of the folding. However, the induced structure in the unimolecular folding reaction is detrimental to the activation of the precursor, since this stage requires some

degree of unfolding. Our results suggest that the SbtE propeptide appears to have evolved to be intrinsically unstructured and to bind with its cognate protease with a specific affinity that is critical for biological regulation.

III) RESULTS:

Maturation pathway of Pro-SbtE:

A complete understanding of the energy landscape of maturation requires (i) determination of the structures of all conformational states along this pathway, including the denatured state, possible intermediates and the native state, (ii) a detailed analysis of the thermodynamics and kinetics of these states and, (iii) the nature of the transition barriers that separate these states. Our previous studies have established that the unimolecular maturation pathway of Pro-SbtE involves the initial folding of the polypeptide to a stable pre-processed complex (Fig. 3.1, Stage I, U←→N), cleavage of the peptide bond between the propeptide and the protease (Stage II, N←→A), and the rate limiting release and degradation of the propeptide from the processed complex (Stage III, A→M) (Yabuta, Takagi et al. 2001). Over the years we have isolated specific mutations that enable us to establish the nature of each of the individual stages of this pathway (Fig. 3.1). For example, the active site mutant Pro-S_{221}A-SbtE folds to a structured state (Fig. 3.2) but loses its ability to cleave and degrade its propeptide, and represents the maturation intermediate prior to being processed (Stage I) (Shinde and Inouye 1995). Similarly Pro-S_{221}C-SbtE variant folds into a structured state (Fig. 3.2) that can auto process its cognate propeptide but is unable to subsequently degrade it (Fig. 3.1). This represents the autoprocessed maturation intermediate (Stage II) (Jain, Shinde et al. 1998; Yabuta, Subbian et al. 2002). Hence, the above active-site variants together with wild type SbtE provide an ideal system to establish the thermodynamics

and kinetics of the individual steps of the maturation pathway, and to analyze the implications of the synergy between these different stages.

The structures of mature SbtE (Stage III) and Pro-S_{221}C-SbtE complex (Stage II) have been solved using x-ray crystallography (Chu, Chao et al. 1995; Jain, Shinde et al. 1998). These structures demonstrate that the conformations of the protease domains are identical in the last two stages of maturation and are unaffected by the active-site substitution. The structure of the un-cleaved precursor (Stage I) has not been solved. However, the structure of a closely related homologue (Comellas-Bigler, Maskos et al. 2004) establishes that the protease domain in the uncleaved precursor is similar to the wild type protease, but has a distorted 6-residue N-terminus that is covalently attached to the C-terminus of the propeptide (Shinde and Inouye 1995). Hence, high-resolution structures of the stable maturation intermediates are available. While the thermodynamic properties of the intermediates can be studied at equilibrium, the properties of the high-energy and metastable transition states can be inferred from kinetic studies.

Hydrophobic collapse coincides with secondary structure formation but precedes tertiary structure acquisition:

While earlier studies established that the propeptide is indispensable to the folding of SbtE, what initiates/drives the Pro-SbtE folding process remains unknown. It is widely accepted that hydrophobicity provides the major driving force for most protein folding reaction (Northey, Di Nardo et al. 2002; Calloni, Taddei et al. 2003). Since the kinetics of the hydrophobic core formation,

relative to those for secondary and tertiary structure formation, can provide insights into folding mechanisms (Shastry and Udgaonkar 1995), we performed a detailed analysis of the folding kinetics for Pro-SbtE. The formation of a hydrophobic core during folding of a protein can be monitored under both equilibrium and kinetic conditions by using dyes such as 1-anilino 8-napthalene sulphonic acid (ANS). ANS specifically interacts with exposed hydrophobic surfaces on proteins, and upon binding, exhibits an increase in the fluorescence emission intensity. This approach has been successfully utilized to establish the presence of partially structured molten-globule intermediates in several protein-folding models (Engelhard and Evans 1995; Shastry and Udgaonkar 1995). Fig. 3.3a establishes that when denatured Pro-SbtE (0.65 μM) is rapidly diluted (30 fold) into a refolding buffer that contains 6.0 μM ANS, a rapid burst in the fluorescence emission at 458 nm is observed. The fluorescence emission spectra can be fitted to a single exponential rate equation and provides a rate constant (k_h = 0.26 s^{-1}) for the formation of the hydrophobic core. It is important to note that in the presence of ANS, only partially folded Pro-SbtE (structure at 600s) and not its denatured state (6 M GdnHCl), or mature SbtE, exhibits significant fluorescence intensity (Fig. 3.3b). Hence, the increase in fluorescence reflects the rapid formation of the hydrophobic core of Pro-SbtE (Fig. 3.3a).

We next monitored the kinetics of secondary and tertiary structure acquisition of Pro-SbtE. Folding was initiated as described in Materials and Methods. Acquisition of secondary structure was monitored as a function of time, using the changes in circular dichroism (CD) ellipticity at 222nm, while the

tertiary structure was monitored using changes in intrinsic tryptophan fluorescence (Fig. 3.3a). The kinetics obtained using CD-spectroscopy best fit a single exponential rate equation with a rate constant of $0.35s^{-1}$ at 21°C. However, the intrinsic fluorescence profiles best fit a double exponential rate equation with rate constants of $0.021s^{-1}$ and $0.0016s^{-1}$ for the fast and slow phases, respectively. The two phases may be consequences of proline isomerization or may reflect the modular folding of the protease domain followed by the relatively slower folding of the propeptide domain (Bryan 2002). Interestingly, the rate constants describing the tertiary structure acquisition are significantly slower than the rate constant for the secondary structure ($0.28s^{-1}$) or hydrophobic collapse ($0.26\ s^{-1}$) (Table 3.T.1). Hence, these results suggest that a rapid hydrophobic collapse that serves to minimize conformational entropy of Pro-SbtE, precedes the relatively slower acquisition of tertiary structure.

It has been argued that hydrophobic dyes such as ANS may bind and stabilize normally transient folding species and shift their equilibrium towards stable intermediates (Ali, Prakash et al. 1999). To examine the effects of ANS on the maturation pathway, we monitored the kinetics of folding and autoprocessing of Pro-SbtE using quantitative SDS-PAGE (Fig. 3.3c). The extent of autoprocessing was measured as a function of time (Fig. 3.3d) and establishes that the kinetics remain unaffected in the presence of ~15-fold molar excess of ANS (10 μM). Hence, our results suggest that the hydrophobic dye does not appear to increase stability of the folding intermediate.

Energetic and Structural Analysis of Stage 1- Pro-subtilisin Folding:

Kinetics of Stage I analyzed using the Pro-S_{221}A-SbtE variant: The folding reaction-

To obtain insights into the nature of the folding transition state, we next examined the dependence of folding kinetics on the temperature of the reaction. Such an analysis can also yield information about the entropy, enthalpy, heat capacity, flexibility of the polypeptide backbone, as well as the solvation of a protein in its transition state (Matagne, Jamin et al. 2000). The kinetics of Pro-S_{221}A-SbtE was monitored as described in Materials and Methods, and the temperature dependence of the folding kinetics was established over the range of 0-25 °C. The Eyring plots thus obtained establish that the rate of folding (k_f) is non-linear over the temperature range investigated (Fig. 3.4a). Non-linear Eyring plots have been observed in several other proteins (Chen, Baase et al. 1989; Pappenberger, Aygun et al. 2001). Non-linearity has been attributed to one of several reasons, the most common of which are (i) change in heat capacity, (ii) change in the rate-limiting step or (iii) the presence of a folding intermediate. To examine possible causes, the kinetics were fitted to Eqn. 3.2 (See Materials and Methods), to obtain relative changes in the free energy ($\Delta G^{\mp}$), entropy ($\Delta S^{\mp}$) and heat capacity ($\Delta C_p^{\mp}$) and hence the overall enthalpy ($\Delta H^{\mp}$) of the transition states. The thermodynamic parameters obtained by best-fits of the Eyring plots are depicted in Table 3.T.1 and indicate that the net change in heat capacity between the unfolded and transition state is negligible (~0.1kcal/mol K). While a temperature dependent change in the rate-limiting

step cannot be overruled, the observed non-linearity in the Eyring plot appears to be most likely due to the presence of a stable folding intermediate (see next section).

Previous studies from our laboratory and others have demonstrated that the isolated propeptide is an intrinsically unstructured domain (Buevich, Shinde et al. 2001). Why the propeptide is intrinsically unstructured and how this affects the folding landscape is unknown. Studies establish that the propeptide folds into a well defined α-β structure in the presence of co-solvents such as 10% glycerol and 25% trifluoroethanol (Shinde, Li et al. 1993). Such concentrations of glycerol do not substantially alter the viscosity of the solution or the activity of the protease. Further, the co-solvent induced secondary structure appears to be similar to the structure of the propeptide in complex with the protease domain (Subbian, Yabuta et al. 2005). Introduction of specific mutations can also significantly increase the native-like secondary structure within the isolated propeptide domain (Ruvinov, Wang et al. 1997). The propeptide variants with increased structural stability are invariably better chaperones and tighter inhibitors (Li, Hu et al. 1995; Wang, Ruan et al. 1998). However, the effects of mutations are cumulative on the entire maturation pathway, and the delineation of their effects on individual stages becomes difficult. Hence, instead of introducing mutations in the propeptide, we analyze the folding reaction and the transition state (Stage I), in the presence of 10% glycerol as a co-solvent.

The folding reaction in the presence of glycerol also follows single exponential kinetics with a non-linear dependence upon temperature (Fig. 3.4a). The rate of folding is ~ 2-fold faster in the presence of 10% glycerol, and

is consistent with data obtained using stabilizing propeptide mutants (Li, Hu et al. 1995; Wang, Ruan et al. 1998). The thermodynamic parameters obtained from the fit of the folding kinetics to Eqn 3.1 (See Materials and Methods) are depicted in Table 3.T.1. The activation energy barriers of the folding reaction in the absence (ΔG^*=~18.0 kcal/mol) and presence of glycerol (ΔG^*=~17.8 kcal /mol) suggests that the induced structure does not significantly lower the free energy of the folding transition state. It is however interesting to note that the overall entropy of the transition state is reduced in the presence of glycerol and is compensated by a gain in enthalpy probably due to packing of the side chains (Table 3.T.1). This is again consistent with less exposed hydrophobic surface area in the presence of 10% glycerol (data not shown). Hence, our results suggest that the induction of structure in the propeptide does not significantly lower the activation energy barrier to folding.

Thermodynamic equilibrium of Stage I analyzed using the Pro-S_{221}A-SbtE variant:

To examine thermodynamic stability of the uncleaved precursor, the equilibrium unfolding of Pro-S_{221}A-SbtE was monitored as a function of denaturant concentrations (Fig. 3.4b). Under non-denaturing conditions, Pro-S_{221}A-SbtE adopts a well-defined structure as observed using CD (Fig. 3.2). Under increasing concentrations of denaturants Pro-S_{221}A-SbtE unfolds through a partially structured state, which represents a productive folding intermediate (I), as established earlier. This structural transition is completely reversible, and suggests that the folded but un-cleaved precursor adopts a thermodynamically stable conformation. The folding/unfolding profiles (Fig. 4b) best fit a three state

transition (Eqn. 3.2; Materials and Methods) and indicate that the complex at Stage I is stabilized by a free energy difference of ~ΔG^0= -4.6 kcal/mol, relative to the unfolded state (Table 3.T.2). The I-state appears maximally populated at denaturant concentration between ~1.0 to 1.2 M GdnHCl (Fig. 4b), and has a net free energy of stabilization of ~2.9 kcal/mol, relative to the folded state. The I-state therefore is marginally stable (ΔG=~1.7 kcal/mol) compared to the unfolded state and exhibits the characteristics of a partially structured molten-globule intermediate (Fig. 3.4b). It is important to note that the non-linearity observed in the temperature dependent folding kinetics (Fig. 3.4a) may be a consequence of this intermediate.

We next examined how inducing structure within the propeptide (using 10% glycerol as a co-solvent) affects equilibrium folding/unfolding of Pro-S_{221}A-SbtE. Even under these conditions the complex exhibits a distinct three-state transition (Fig 4b; Table 3.T.2). However, stabilities of the intermediate and native states are enhanced by ~2.0 and ~2.6 kcal/mol, respectively. Further, the stable I state is maximally populated at ~2.25 M GdnHCl (Fig. 4b). Hence, our data suggests that a structured propeptide stabilizes both the folded state and the I-state.

Kinetics of Stage I analyzed using the Pro-S_{221}A-SbtE variant: The unfolding reaction-

To obtain insights into the transition state separating the I- from the N-state, folded Pro-S_{221}A-SbtE was rapidly diluted into denaturant concentrations wherein the intermediate species are maximally populated. The loss in ellipticity at 222nm was monitored using CD spectroscopy and the unfolding kinetics was

analyzed using GraphPad Prism (inset). The temperature dependence of the unfolding kinetics from the N-state to I state follow strong non-linearity, both with and without glycerol (Fig. 3.4c). Using Eqn 3.1 the activation free energies, entropies, and activation heat capacities for the unfolding reaction were estimated. The activation enthalpies were then calculated from the free energies and the entropies. It is interesting to note that the rates of unfolding of the precursor to its intermediate state is faster in the presence of glycerol at lower temperatures (Fig. 3.4c), and the trend is reversed at higher temperature (above 17°C). While reasons for this change in dependence are currently unknown, one possible cause may be the presence of additional intermediates that are not resolved using equilibrium unfolding. Alternatively, it is possible that this non-linearity may be a consequence of a change in the rate-determining step of the unfolding reaction.

Energetic and Structural Analysis of Stage II: Pro-subtilisin autoprocessing:

Pro-subtilisin processing involves the autoproteolytic cleavage of the peptide bond between the C-terminus of the propeptide and the N-terminus of the protease domain and results in a non-covalently associated stoichiometric complex (Shinde and Inouye 1995; Yabuta, Subbian et al. 2002). Auto-cleavage is accompanied by structural rearrangements within the complex, which coincides with the formation of a high-affinity calcium-binding site (Yabuta, Subbian et al. 2002). To establish the energetics associated with autoprocessing, we measured the temperature dependence of this reaction

using Pro-S_{221}C-SbtE variant, both in the presence and absence of glycerol. This variant is ideal for these studies since its rate of folding is similar to Pro-SbtE, while its rates of autoprocessing are approximately 100-fold slower (Li and Inouye 1996; Jain, Shinde et al. 1998). Hence, unlike Pro-SbtE insignificant levels of autoprocessing are observed during the folding of the Pro-S_{221}C-SbtE variant. Moreover, Pro-S_{221}C-SbtE is unable to degrade its propeptide and this prevents *trans* proteolysis. This allows one to fold the precursor at a fixed temperature (4°C) and then measure the kinetics of auto-cleavage at any desired temperature using a quantitative SDS-PAGE analysis as described in Materials and Methods. Fig. 3.5a depicts a representative kinetic curve for autoprocessing and the rate for this reaction was obtained by fitting the profile to a single exponential equation. The temperature dependence of the rate constant thus obtained provides an Eyring plot for the autoprocessing reaction. The Eyring plot fits equally well to Eqn 3.1 and to a linear function (Pappenberger, Aygun et al. 2001) and the estimated parameters for the fit are listed in Table 3.T.1. It is interesting to note the presence of a 'crossover' between the Eyring plots obtained in the presence and absence of glycerol. At this 'crossover' temperature the rates of autoprocessing remain unaffected by glycerol, while at lower temperatures the rates of autoprocessing increase. At higher temperatures however, it appears that the increased structural stability of the propeptide slows autoprocessing. This mirrors the unfolding behavior of the precursor to its intermediate state (Fig. 3.4c). Although reasons for why the rates of unfolding are higher in the presence of glycerol are currently unknown, these results suggest that the kinetics of autprocessing appear to directly

correlate with the kinetics of unfolding of the precursor. Hence, the stabilization of the uncleaved complex appears to be detrimental to autoprocessing. In addition, the presence of glycerol in the folding conditions does not lower the activation energy barrier significantly as seen from Table 3.T.1.

To better understand this behavior we studied the autoprocessing reaction under equilibrium conditions in the presence of varying concentrations of GdnHCl at room temperature. Under non-denaturing conditions, the yield of mature enzyme in the presence of glycerol is lower than that observed in its absence. This suggests that the higher stability (lower unfolding rates) in the presence of glycerol adversely affects autoprocessing. Interestingly, if the stability is decreased by the addition of low concentrations of denaturants (0.75 M GdnHCl) the yield of SbtE is increased. This mirrors the trend observed in the kinetics of unfolding (Fig. 3.4c) and autoprocessing (Fig. 3.5b). This suggests that partial unfolding maybe essential for efficient autoprocessing. However this is not reflected in the activation energy barriers and requires additional analysis.

Thermodynamic equilibrium of the processed complex:

The thermodynamic stability of the autoprocessed complex was established by monitoring the loss in ellipticity as a function of denaturant concentration. It is important to note that this folding/unfolding reaction is reversible due to the presence of the propeptide, which can facilitate folding *in trans* (Zhu, Ohta et al. 1989). Analysis of the unfolding curves using Eqn. 3.3 establishes the free-energy of unfolding of the autoprocessed complex (Table 3.T.2 and Fig. 3.5b). The equilibrium unfolding was monitored both in the

presence and absence of 10% glycerol. The presence of glycerol significantly increases the stability of the processed complex ($\Delta\Delta G$= 4.6 kcal/mol). Thus, a structured propeptide increases the thermodynamic stability of both the uncleaved precursor and the cleaved complex.

Energetic and Structural Analysis of Stage III: Pro-subtilisin Activation:

The release of the propeptide and its subsequent degradation by the active protease represents the rate-determining step to propeptide release. We earlier established the energetics of Pro-SbtE activation (Subbian, Yabuta et al. 2005) . The temperature dependence of the activation kinetics demonstrated that the kinetic barrier of this reaction is ~17.3 ± 0.3 kcal mol^{-1}. However, this measured Ea signifies the limiting energy barrier for release and degradation. Since the proteolysis of the propeptide is rapid, it should be noted that the energy barrier for release is most likely higher. In the presence of 10% glycerol, the Pro-SbtE inhibition complex is extremely stable, with the protease being inactive till 8 days at 23 °C and 4 weeks at 4 °C. Hence we were unable to determine the kinetics of activation in the presence of glycerol. However, based on the extremely slow activation observed under these conditions, it is reasonable to state that the energy barrier to activation significantly increases in the presence of a structured propeptide domain.

The thermodynamic stability of the mature protease relative to the autoprocessed complex was established by monitoring the extent of binding of the propeptide to the mature domain. The K_a for this interaction under folding

conditions was $7.34 \times 10^6\ M^{-1}$ and (Subbian, Yabuta et al. 2005) and the propeptide binding with SbtE stabilizes the inhibition complex by ~9.3 kcal/mol. Thus, the free-energy diagram for activation favors formation of a thermodynamically stable inhibition complex over the active protease. This established that the native state of the protease is not at its thermodynamically stable conformation but is trapped in a local energy minimum by a high barrier to unfolding.

IV) DISCUSSION:

Energy Landscape for Pro-Subtilisin Maturation:

A complete study of the folding pathway involves the analysis of the structural acquisition from the unfolded state to the native state of a protein. In this study, we define the folding energy landscape of unimolecular Pro-SbtE maturation, using active site variants. The complete maturation involves the three distinct stages of folding, autoprocessing, and activation (Fig. 3.1). Our results establish that folding is initiated by a rapid hydrophobic collapse and occurs through the formation of a molten globule-like intermediate. Both the initial folding and the subsequent autoprocessing are thermodynamically driven processes Autoprocessing is accompanied by the formation of the high-affinity calcium-binding site and significantly increases the thermodynamic stability of the complex (Yabuta, Subbian et al. 2002). This results in a high enthalpic barrier to autoprocessing. Earlier analysis of Pro-SbtE activation demonstrated that the release of the propeptide from the complex is energetically unfavorable and its proteolytic removal results in the formation of a kinetically trapped native state (Subbian, Yabuta et al. 2005). Taken together, the energetics of Stage I, II and II, establish that the native state SbtE is also thermodynamically unstable. A similar native state has been observed with the propeptide mediated folding of alpha-lytic protease and appears to offer distinct functional advantages (Jaswal, Sohl et al. 2002).

Effect of a structured propeptide on the folding energy landscape:

The propeptide of SbtE is intrinsically unstructured and folds only in the presence of its cognate protease domain (Shinde, Li et al. 1993). To understand how this structural instability affects maturation we established the energy landscape in the presence of a structured propeptide. Our results demonstrate that the propeptide does not significantly contribute to the first transition state to folding and only functions subsequent to the formation of the stable intermediate. Further, inducing structure in the propeptide does not significantly affect the kinetic barriers to folding and autoprocessing but enhances the thermodynamic stability of the structured states. However such stabilization of the uncleaved complex appears to be detrimental to auto-cleavage and dramatically prolongs activation. While this suggests that autoprocessing of the propeptide, similar to its release, may require some structural instability within the propeptide, the absence of a significant change in the activation energy barrier necessitates further analysis.

Hence, although structured native states and thermodynamic stability are an accepted norm for a majority of proteins, the kinetically-determined folding pathway of Pro-SbtE appears to be optimized for a structurally unstable propeptide domain. In this case of protein folding, kinetic-stability, its consequent functional advantages, and structural instability appear to have co-evolved.

V) MATERIALS AND METHODS:

Circular dichroism measurements:

CD measurements were performed on an automated AVIV 215 spectrophotometer maintained at 23°C and spectra were taken between 190 to 260 nm as described earlier (Shinde, Liu et al. 1997; Fu, Inouye et al. 2000; Yabuta, Subbian et al. 2002). Protein concentrations were maintained between 0.25 mg to 0.4 mg ml^{-1}. A 1 mm path-length cuvette was used to measure spectra except in case of the propeptide, where a 0.5 mm path-length cuvette was used. Each plot represents the average of 3 independent scans.

Kinetics of Folding and Unfolding:

Proteins were expressed and purified as described earlier (Shinde and Inouye 1995; Jain, Shinde et al. 1998; Yabuta, Takagi et al. 2001). Folding kinetics were measured by monitoring changes in CD ellipticity, extrinsic fluorescence using ANS and intrinsic tryptophan fluorescence. To monitor the kinetics by CD, Pro-SbtE (in 6M GdnHCl, pH 4.8) was rapidly diluted into folding buffer (50 mM MES-NaOH, pH 6.5, 0.5 M $(NH_4)_2SO_4$, 1mM $CaCl_2$), with or without 10% Glycerol, and changes in CD ellipticity were monitored at 225nm. To establish the effect of temperature, folding kinetics were monitored as described above, from 4-25°C. Changes in fluorescence were measured on a Hi-Tech SF61-DX2 stopped flow fluorimeter using $\lambda_{excitation}$ of 295nm and 310nm band-pass filter, using similar conditions as the CD measurements. To monitor

the initial hydrophobic collapse, 5μM of ANS was included in the folding buffer. The sample was excited at 390nm and emission was monitored at 468nm.

To estimate the kinetics of unfolding, fully folded Pro-S_{221}A-SbtE was rapidly diluted into folding buffer containing either 1.1M GdnHCl or, 2.25M GdnHCl and 10% glycerol, and the loss in structural content was monitored by CD. The concentrations of denaturant represent conditions at which maximum fractions of intermediate (I) are seen at equilibrium. Folding curves were analyzed using GraphPad Prism and the temperature dependence of the microscopic rate constant (k_i) was analyzed according to the following equation (Chen, Baase et al. 1989; Matagne, Jamin et al. 2000):

$$k_i = \exp[A_i + B_i(T°/T) + C_i(\ln(T°/T) + \ln T] \qquad \text{Equation (3.1)}$$

where:

$$A_i = [-\Delta C_{pi}^{\mp} + \Delta S_i^{\mp}(T°)]/R + \ln(k_b/h)$$

$$B_i = [\Delta C_{pi}^{\mp} - \Delta S_i^{\mp}]/R - \Delta G_i^{\mp}(T°)/RT°$$

$$C_i = -\Delta C_{pi}^{\mp}/R$$

Effect of ANS on folding:

Pro-SbtE (200nM) was folded in the presence of increasing amount of ANS (0-10μM). 20μl aliquots were taken at different time points after folding initiation, the reaction was stopped by TCA, and samples were analyzed through SDS-PAGE and gel-scanning densitometry.

Equilibrium unfolding:

Equilibrium folding-unfolding was monitored (CD at 222nm) to obtain free-energy differences for each transition as described (Eder, Rheinnecker et

al. 1993; Eder, Rheinnecker et al. 1993; Raschke and Marqusee 1997; Fujiwara, Arai et al. 1999; Takei, Chu et al. 2000).The transitions are represented by at least three-states (Fujiwara, Arai et al. 1999), the native (N), intermediate (I) and unfolded states (U):

$$N \underset{}{\overset{K_{NI}}{\rightleftharpoons}} I \overset{K_{IU}}{\rightleftharpoons} U \qquad \text{Equation (3.2)}$$

The observed ellipticity (A_{obs} *(c)*) at any concentration of the denaturant is given by the sum of the contributions from the three states as (Mizuguchi, Arai et al. 1998; Takei, Chu et al. 2000):

$$A_{obs}(c) = \left[\frac{A_N + A_I \exp[-(\Delta G_{NI}^{H_2O} - m_{NI}c)/RT] + A_U \exp[-(\Delta G_{NU}^{H_2O} - m_{NU}c)/RT]}{1 + \exp[-(\Delta G_{NI}^{H_2O} - m_{NI}c)/RT] + \exp[-(\Delta G_{NU}^{H_2O} - m_{NU}c)/RT]} \right]$$

Equation (3.3)

Where f_N *(c)*, f_I *(c)* and f_U *(c)* are the fractions of the three states at a GdnHCl concentration of *c* ($f_N + f_I + f_U = 1$), and A_N, A_I and A_U are the ellipticity values of the N, I, and U states, respectively. The f_N, f_I and f_U terms are related to the equilibrium constants, K_{NI} and K_{NU}, of the unfolding transitions from N to I and from N to U, respectively. $\Delta G_{NI}^{H_2O}$ and $\Delta G_{NU}^{H_2O}$ are the ΔG_{NI} and ΔG_{NU} at 0M GdnHCl, and $m_{NI}c$ & $m_{NU}c$ represent the dependence of the respective free energy changes on *c*. The data were fitted using Prism Graphpad.

Kinetics of autoprocessing:

The kinetics of autoprocessing were monitored using Pro-S_{221}C-SbtE. 20μl aliquots were taken at increasing times after folding initiation and the

samples were analyzed using SDS-PAGE and quantitative gel-scanning densitometry to estimate the kinetics (Fu, Inouye et al. 2000; Yabuta, Takagi et al. 2001). To monitor the effect of temperature on autoprocessing kinetics, the samples were transferred to respective temperatures after 5 mins (completion of folding), and the kinetics were monitored at appropriate temperatures as above. The autoprocessing curves were analyzed using GraphPad Prism. To establish the efficiency of autoprocessing, the protein was diluted into folding buffer containing appropriate amount of denaturant. Reaction was allowed to proceed for 30 mins, after which the samples were analyzed as above.

Kinetics of activation:

20 μl of denatured Pro-SbtE (100 μM in 6M GdnHCl, pH 4.8) was rapidly mixed in 19980 μl of the folding buffer at 23°C with 0.5 mM synthetic substrate (N-succ-AAPF-pNA). After 15 min the stirring was stopped and 200 μl aliquots of the reaction mixture were transferred to a 96-well microplate. Subtilisin activity was monitored by release of *p*-nitroanilide measured at 405 nm in a microplate reader (Yabuta, Takagi et al. 2001; Yabuta, Subbian et al. 2002; Yabuta, Subbian et al. 2003) maintained at the desired temperature. The time of activation was calculated from the X-axis intercept of the 'none-to-all' transition curve as described earlier(Fu, Inouye et al. 2000; Inouye, Fu et al. 2001). Ea was calculated from the Arrhenius plots obtained by measuring the activation rate at different temperatures. The activation was also monitored using SDS-PAGE. To analyze the effect of a structured propeptide on

activation, 10% glycerol was added to the folding buffer and, activation was monitored as discussed above.

VI) FIGURES

Fig. 3.1. Schematic representation of Pro-SbtE maturation pathway. Maturation of Pro-SbtE can be dissected into the three stages of folding, autoprocessing and, release and degradation. Pro-S_{221}A-SbtE represents an active site mutant that folds to a structured state and is blocked at the stage of folding (Uncleaved complex). Pro-S_{221}C-SbtE mutant is able to fold and undergo autoprocessing to result in a structured inhibited complex (Cleaved complex; 1SCJ.pdb) but is unable to release and degrade its propeptide. Wildtype Pro-SbtE undergoes complete maturation to yield active protease (1Sbt.pdb)

Fig. 3.1

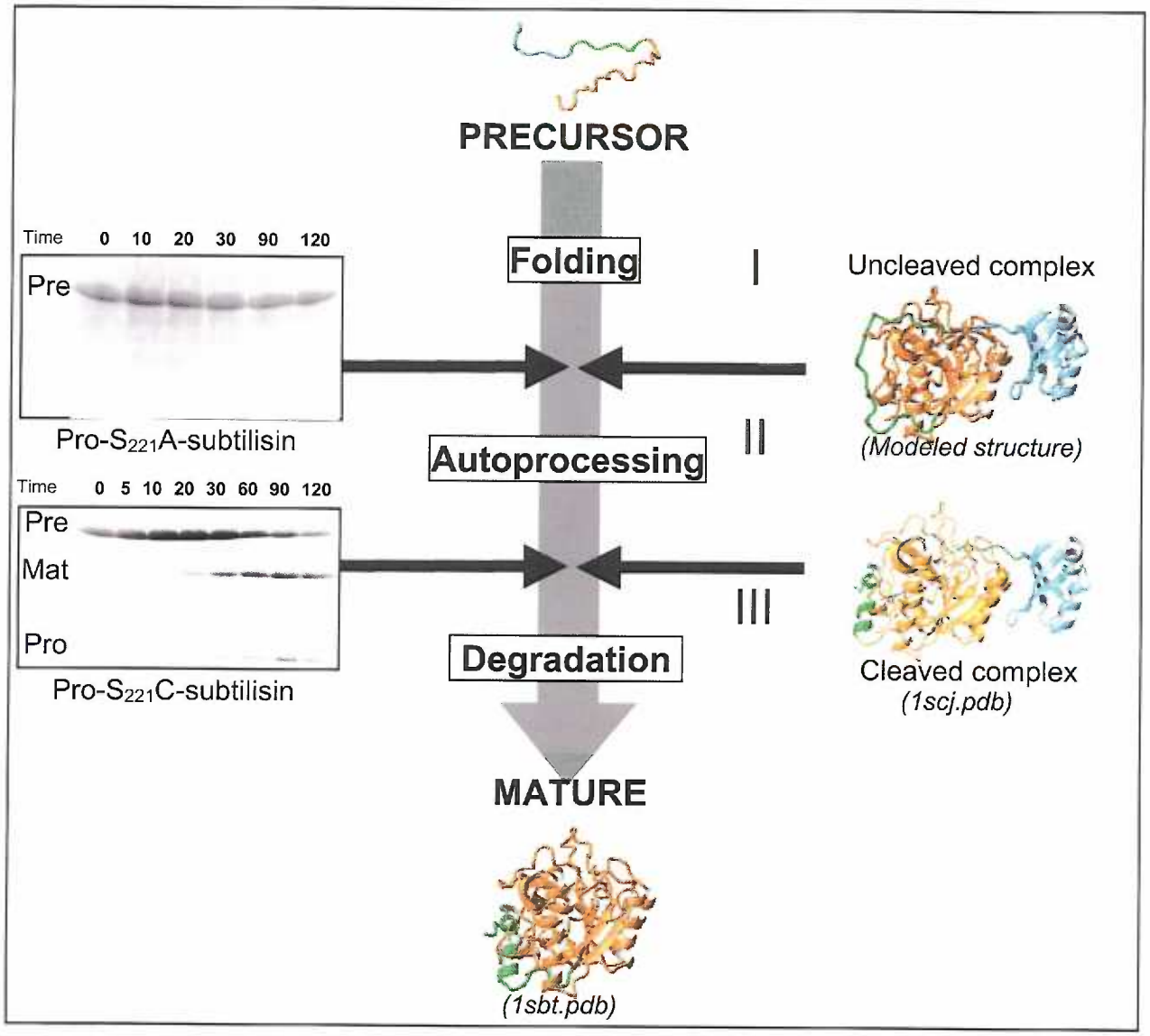

PRECURSOR
Folding
I
Uncleaved complex
(Modeled structure)
Time 0 10 20 30 90 120
Pre
Pro-S221A-subtilisin
II
Autoprocessing
Time 0 5 10 20 30 60 90 120
Pre
Mat
Pro
Pro-S221C-subtilisin
III
Degradation
Cleaved complex
(1scj.pdb)
MATURE
(1sbt.pdb)

Fig. 3.2. Secondary structure characterization using circular dichroism (CD) spectroscopy. (a) CD spectra of Pro-SbtE and its active-site mutants: Pro-S_{221}C-SbtE inhibition complex (solid black line), Pro-S_{221}A-SbtE uncleaved complex (dotted grey line), SbtE (dotted black line), isolated propeptide (solid grey line) and differential spectra of propeptide in complex (broken line).

Fig. 3.2

a

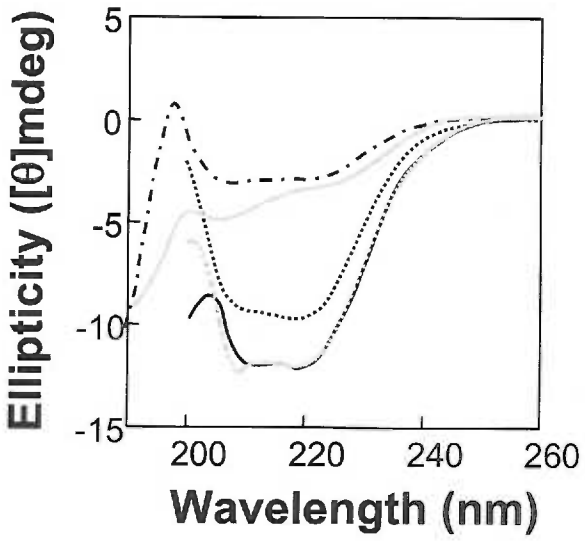


Fig. 3.3. Characterization of initial collapse and folding of the precursor: (a) Kinetics of Pro-SbtE folding, monitored by changes in ANS fluorescence (upper), CD ellipticity (middle) and intrinsic fluorescence (lower) **(b)** Fluorescence emission spectra of ANS_: dotted grey line – ANS control; solid black line –Pro-SbtE 600s after folding initiation; dotted black line – denatured Pro-SbtE; and solid grey line – mature subtilisin. **(c)** SDS-PAGE of aliquots taken at 0, 2 and 5min after initiation of folding, in presence of denoted amounts of ANS. **(d)** Extent of autoprocessing, as a function of increasing ANS concentrations, estimated from SDS-PAGE analysis by quantitative gel-scanning densitometry; open circles-0μM ANS; open diamond-0.1μM ANS; open triangles-1μM ANS; inverted triangles-10μM ANS.

Fig. 3.3

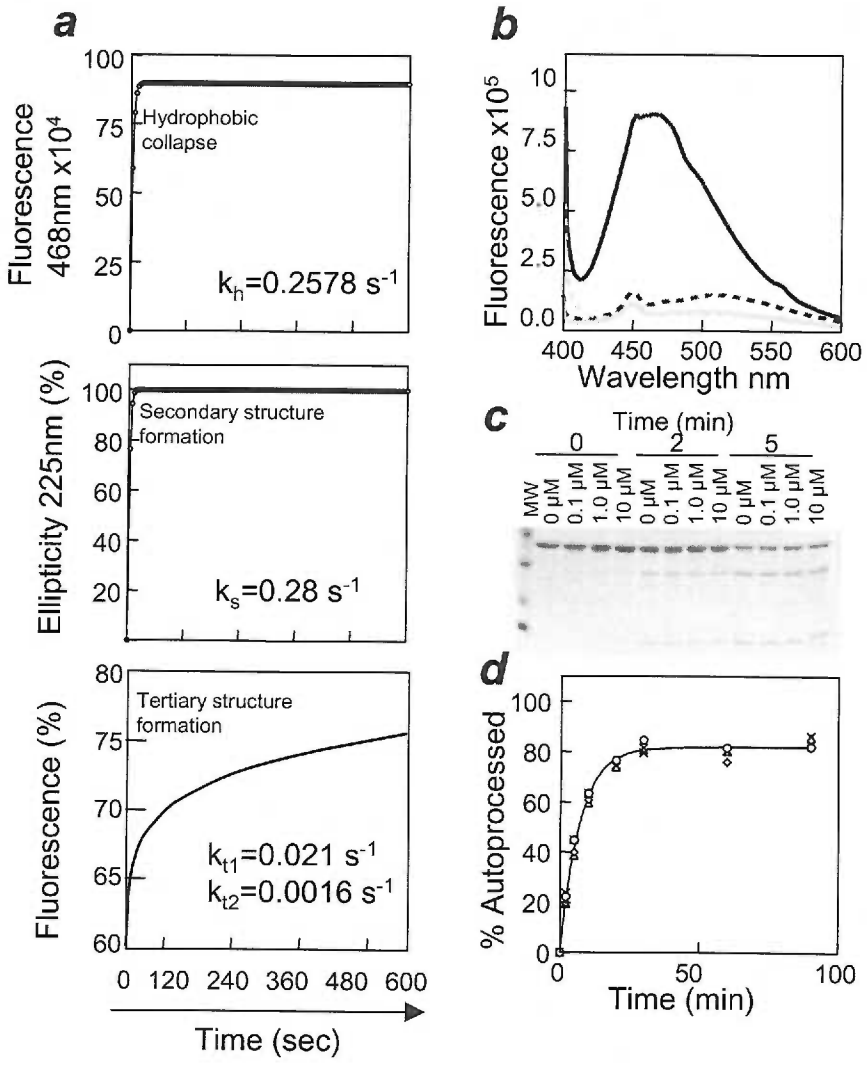

a
Fluorescence 468nm x10^4
Hydrophobic collapse
k_h=0.2578 s^-1
Ellipticity 225nm (%)
Secondary structure formation
k_s=0.28 s^-1
Fluorescence (%)
Tertiary structure formation
k_t1=0.021 s^-1
k_t2=0.0016 s^-1
Time (sec)
b
Fluorescence x10^5
Wavelength nm
c
Time (min)
MW
0 µM
0.1 µM
1.0 µM
10 µM
d
% Autoprocessed
Time (min)

Fig. 3.4. Energetics of stage I- Folding. (a) Temperature dependence of the kinetics of folding under normal conditions (filled squares), and in the presence of 10% glycerol (filled triangles) **(b)** Equilibrium unfolding of folded Pro-S_{221}A-SbtE under normal conditions (filled circles), and in the presence of 10% glycerol (open circles). Deconvoluted constituent fractions under both normal and glycerol conditions are depicted; native (N) - dotted line; intermediate (I) – solid black line and unfolded (U) – broken line. **(c)** Temperature dependence of the kinetics of unfolding from the U-state to the I-state under normal conditions (squares), and in the presence of 10% glycerol (triangles).

Fig. 3.4

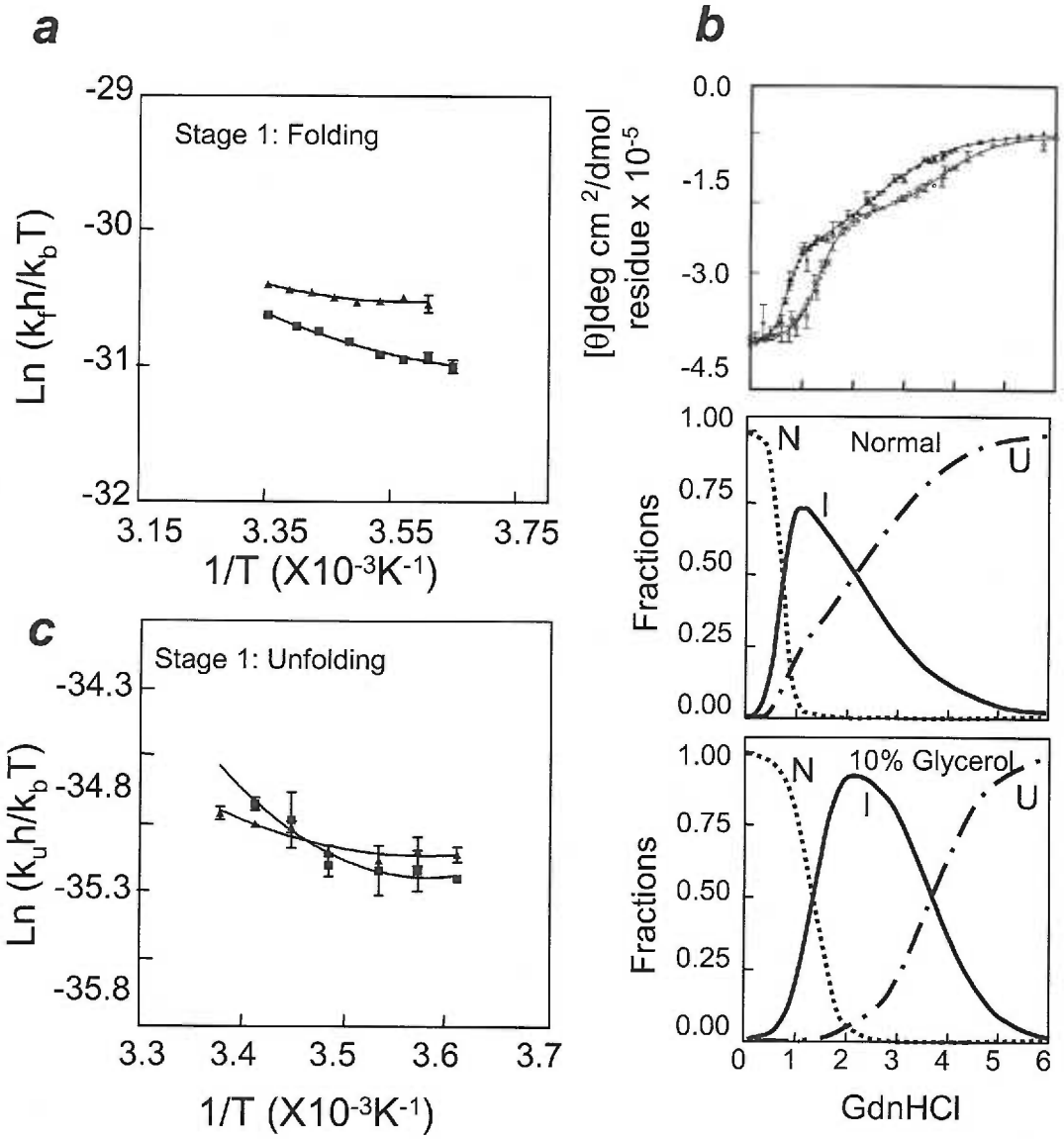

a
Stage 1: Folding
Ln ($k_f h/k_b T$)
-29
-30
-31
-32
3.15
3.35
3.55
3.75
1/T (X10^{-3}K^{-1})
b
[θ]deg cm ^{2}/dmol residue x 10^{-5}
0.0
-1.5
-3.0
-4.5
Fractions
1.00
0.75
0.50
0.25
0.00
N
I
U
Normal
10% Glycerol
0
1
2
3
4
5
6
GdnHCl
c
Stage 1: Unfolding
Ln ($k_u h/k_b T$)
-34.3
-34.8
-35.3
-35.8
3.3
3.4
3.5
3.6
3.7
1/T (X10^{-3}K^{-1})

Fig. 3.5. Energetics of stage II - Autoprocessing. (a) Kinetics of autoprocessing estimated from SDS-PAGE (Inset) through quantitative gel-scanning densitometry. Mature protein – open circles; precursor – closed circles; **(b)** Temperature dependence of autoprocessing kinetics under normal (squares), and glycerol (triangles) conditions. **(c)** Efficiency of autoprocessing estimated from the relative yield of mature under increasing amounts of denaturant, in the presence (open circles), and absence (filled circles) of glycerol. **(d)** Equilibrium unfolding of the Pro-S_{221}C-SbtE cleaved complex under normal conditions (upper panel), and in the presence of 10% glycerol (lower panel). The individual fractions obtained from deconvolution of folding curves are Native (N)- dashed grey line; Intermediate (I) - solid grey line and unfolded (U) - broken grey line

Fig. 3.5

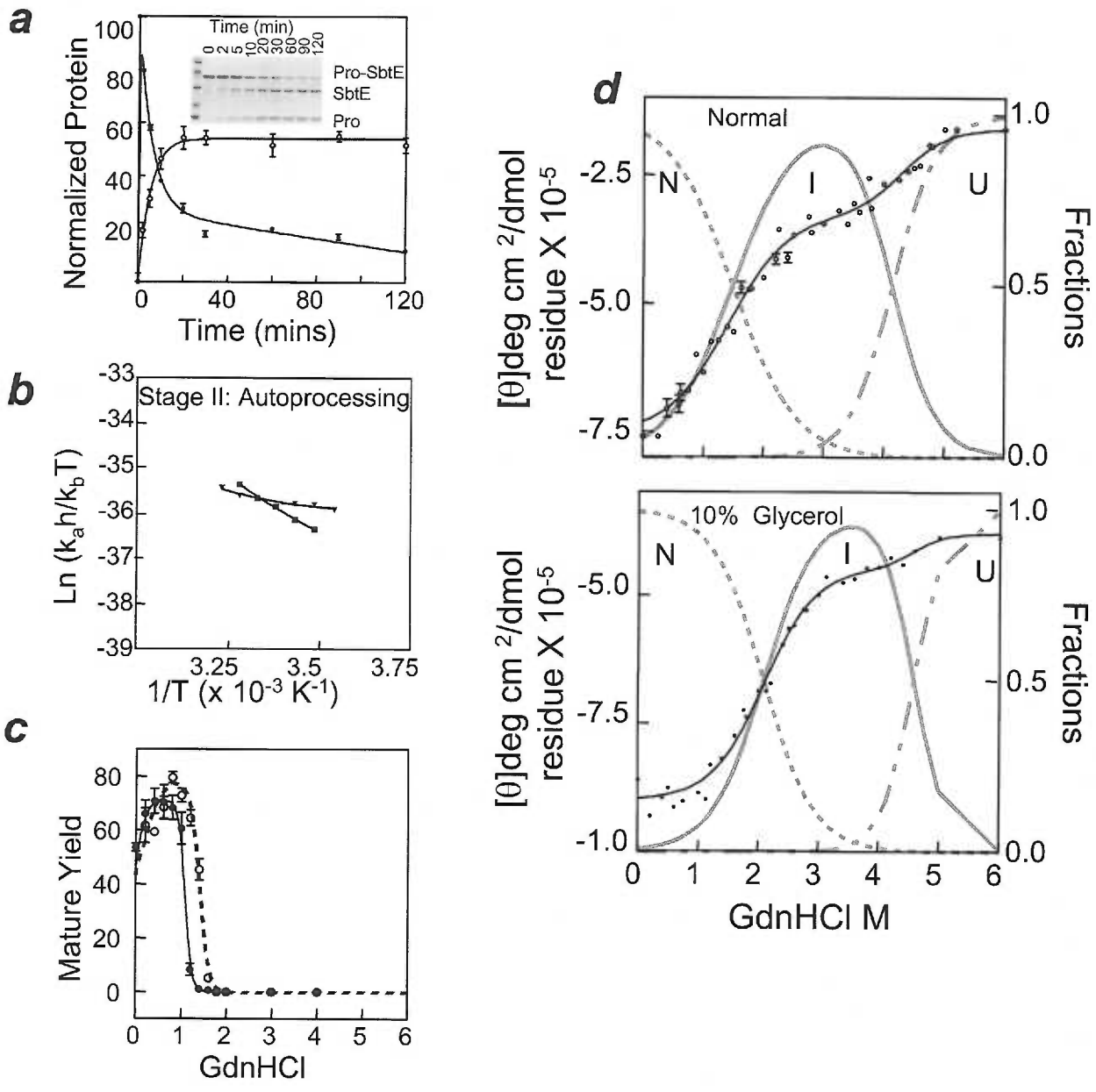
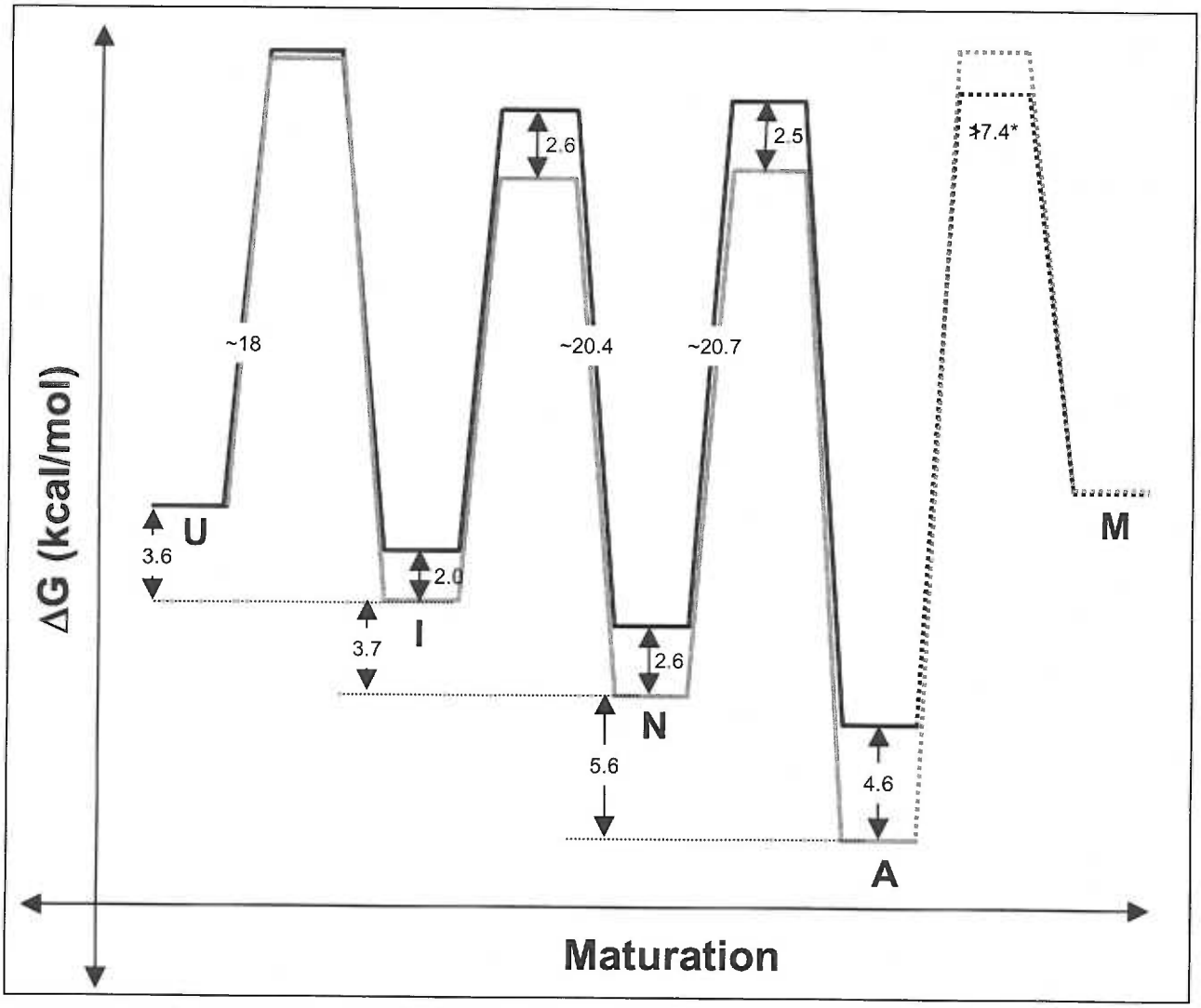

a
Normalized Protein
Time (min)
0 2 5 10 20 30 60 90 120
Pro-SbtE
SbtE
Pro
Time (mins)
b
Stage II: Autoprocessing
Ln ($k_a h/k_b T$)
1/T (x 10^{-3} K^{-1})
c
Mature Yield
GdnHCl
d
Normal
N
I
U
[θ]deg cm ^{2}/dmol residue X 10^{-5}
Fractions
10% Glycerol
GdnHCl M

Fig. 3.6. Folding energy landscape for Pro-SbtE maturation. Folding of Pro-SbtE (N) is driven by a hydrophobic collapse and proceeds through a distinct structured intermediate (I). Autoprocessing of the propeptide results in a thermodynamically stable inhibited complex, A. Release and degradation of the propeptide drives the transition to a kinetically stable native state M (solid black line). In the presence of a structured propeptide (solid grey line), the inherent stabilities of I - and N-states are increased and the folding reaction is favored by ~2.5kcal/mol. Similarly, the stability of the autoprocessed complex is also significantly enhanced (~4.6 kcal/mol). However, the release of the propeptide is limited by its structural stability, and this enhances the barrier to activation. Since the activation in the presence of a structured propeptide is an extremely slow process we were unable to measure its kinetics (broken grey line)

Fig. 3.6


ΔG (kcal/mol)
~18
~20.4
~20.7
+7.4*
2.6
2.5
2.0
2.6
U
I
N
A
M
3.6
3.7
5.6
4.6
Maturation

VII) TABLES

Table 3.T.1: Thermodynamic parameters for Stages I, II, and III of Pro-SbtE maturation

	Stage I Folding (U – N)		Stage I Unfolding (N – U)		Stage II Autoprocessing (N – A)		Stage III Activation (A – M)	
	Normal	10%gly	Normal	10%gly	Normal	10%gly	Normal	10%gly
ΔG kcal/mol	17.963	17.820	20.395	20.450	20.710	20.810	17.390	NA
ΔH kcal/mol	3.692	1.980	10.900	4.510	11.610	4.043	3.880	NA
ΔS cal/mol K	-48.200	-53.500	-31.000	-53.000	-30.700	-56.600	-45.600	NA
ΔC_p kcal/mol K	0.112	0.114	0.657	0.273	0.184	0.156	-0.123	NA

Table 3.T.2: Parameters for equilibrium unfolding of Pro-S_{221}A and Pro-S_{221}C SbtE

	$^{H_2O}\Delta G_{NI}$ kcal mol^{-1}	m_{NI} kcal mol^{-1}	$^{H_2O}\Delta G_{IU}$ kcal mol^{-1}	m_{IU} kcal mol^{-1}	$^{H_2O}\Delta G_{NU}$ kcal mol^{-1}	m_{IU} kcal mol^{-1}
Pro-S_{221}A SbtE *Normal*	2.939	4.180	1.647	0.721	4.586	4.901
Pro-S_{221}A SbtE *10%Gly*	3.604	2.664	3.665	0.058	7.269	3.662
Pro-S_{221}C SbtE *Normal*	1.666	1.179	6.77	1.571	8.436	2.750
Pro-S_{221}C SbtE *10%Gly*	2.869	1.373	10.081	2.197	12.950	3.570

CHAPTER 4

FOLDING PATHWAY MEDIATED BY AN INTRAMOLECULAR CHAPERONE: A FUNCTIONAL PEPTIDE CHAPERONE DESIGNED USING SEQUENCE DATABASES

Yukihiro Yabuta[1], Ezhilkani Subbian, Catherine Oiry[2] and Ujwal Shinde

Department of Biochemistry and Molecular Biology,

Oregon Health and Science University,

3181 SW Sam Jackson Park Road, Portland, OR 97239, USA

[1]Current Address: RIKEN Center for Developmental Biology, Kobe, Japan

[2]Current Address: LAPP, Faculté de pharmacie, 15 Av Charles Flahault, BP 14491, 34093 Montpellier, FRANCE

Journal Biological Chemistry

Vol. 278-No. 17, 15246–15251; April 25 2003

I) ABSTRACT:

Catalytic domains of several prokaryotic and eukaryotic protease families require dedicated N-terminal propeptide domains or "intramolecular chaperones" to facilitate correct folding. Amino acid sequence analysis of these families establishes three important characteristics: (i) propeptides are almost always less conserved than their cognate catalytic domains, (ii) they contain a large number of charged amino acids, and (iii) propeptides within different protease families display insignificant sequence similarity. The implications of these findings are, however, unclear. In this study, we have used subtilisin as our model to redesign a peptide chaperone using information databases. Our goal was to establish the minimum sequence requirements for a functional subtilisin propeptide, because such information could facilitate subsequent design of tailor-made chaperones. A decision-based computer algorithm that maintained conserved residues but varied all non-conserved residues from a multiple protein sequence alignment was developed and utilized to design a novel peptide sequence (ProD). Interestingly, despite a difference of 5 pH units between their isoelectric points and despite displaying only 16% sequence identity with the wildtype propeptide (ProWT), ProD chaperones folding and functions as a potent subtilisin inhibitor. The computed secondary structures and hydrophobic patterns within these two propeptides are similar. However, unlike ProWT, ProD adopts a well defined α-β conformation as an isolated peptide and forms a stoichiometric complex with mature subtilisin. The CD spectrum of this complex is similar to ProWT subtilisin. Our results establish that despite low sequence identity and dramatically different charge distribution,

both propeptides adopt similar structural scaffolds. Hence, conserved scaffolds and hydrophobic patterns, but not absolute charge, dictate propeptide function.

II) INTRODUCTION:

Subtilisin E is an alkaline serine protease isolated from *Bacillus subtilis* (Wong and Doi, 1986). *In vivo* this protein is produced as pre-prosubtilisin (Zhu, Ohta et al., 1989), wherein the pre-region (signal peptide) facilitates protein secretion and the pro-region (propeptide) functions as an intramolecular chaperone that guides correct folding of subtilisin (Shinde, Li et al., 1993; Bryan, Wang et al., 1995; Ruan, Hoskins et al., 1999; Yabuta, Takagi et al., 2001). Upon completion of folding, the propeptide is autoproteolytically removed (Ikemura, Takagi et al., 1987). This is necessary because the propeptide is a potent inhibitor of subtilisin activity (Yabuta, Takagi et al., 2001). Propeptide-mediated folding mechanisms exist in several unrelated protease families. However, there is no sequence conservation in propeptide domains within these families (Shinde and Inouye, 2000). This functional conservation across unrelated protease families suggests that propeptides have evolved in multiple parallel pathways and may share a common mechanism of action (Eder and Fersht, 1995). Subtilisin (Shinde, Li et al., 1993; Bryan, Wang et al., 1995; Ruan, Hoskins et al., 1999; Yabuta, Takagi et al., 2001) and α-lytic protease (Silen and Agard, 1989; Baker, Sohl et al., 1992; Jaswal, Sohl et al., 2002) constitute the best-studied examples of propeptide-mediated protein folding. Bacterial subtilisins are models for the subtilase family that pans prokaryotes and eukaryotes. Amino acid sequence analysis of this family establishes that although the subtilisin domain is conserved, the cognate propeptides can vary significantly (Shinde and Inouye, 2000). Because propeptides perform functions different from the catalytic domains, they may be subjected to different

mutational frequencies because of different functional constraints (Miyata and Yasunaga, 1980; Li, 1985; Li, Wu et al., 1985). However, given that they impart structural information to their catalytic domains (Shinde, Liu et al., 1997; Shinde, Fu et al., 1999; Muller, Cameron et al., 2000), propeptides within one family could adopt similar structural scaffolds despite digressions in their polypeptide sequences. This makes propeptides attractive models for protein redesign and for understanding the relation between sequence and structure. To establish minimum sequence requirements for a functional propeptide, in this study we have designed and developed a peptide chaperone using information databases on the subtilase family. This family is ideal because enormous structural and functional information on individual members is already available (Shinde and Inouye, 2000). The peptide sequence (ProD) was designed through a decision-based computer algorithm that maintained conserved residues but varied all non-conserved residues from a multiple protein sequence alignment. The DNA sequence coding for ProD was synthesized and subsequently expressed in *Escherichia coli*, and the resulting peptide was purified. Despite a difference of 5 pH units between their isoelectric points (pI) and despite displaying only 16% sequence identity with the wild-type propeptide (ProWT), ProD chaperones folding and functions as a potent subtilisin inhibitor. The computed two-dimensional structure and the distribution of hydrophobic residues within these two propeptides are similar. However, unlike ProWT, ProD adopts a well defined α-β conformation as an isolated peptide. The CD structure of ProD-subtilisin was found to be similar to ProWT-subtilisin. Hence, despite low sequence identity and opposite net charge, ProD

and ProWT adopt similar structural scaffolds. Our results suggest conserved structural scaffolds and hydrophobicity, but not absolute charge, dictates propeptide function.

III) RESULTS:

Criteria for Designing and Selecting ProD:

The algorithm described under Materials and Methods was used to obtain a matrix that was analyzed using the following criteria. If "X" is the most abundant amino acid residue and "Y" is the second most abundant residue at a particular position in the sequence, then residue "X" is always chosen over Y UNLESS (frequency(X) ≤ A*frequency(Y)) AND residue X is present in ProWT, where "A" represents a "cutoff factor." Fig. 4.1*D* depicts the relation between the cutoff factor "A" and the sequence identity of the computed sequences to ProWT. It is evident from Fig. 4.1*D* that as the cutoff factor increases the percent identity decreases according to two different slopes. In the present analysis the value of the cutoff factor was selected as 2 because the transition in slopes occurs when A = 2. If for example, frequency(X) = 70%, then frequency(Y) ≤ 30%; then according to our criterion X was selected because (frequency(X) ≤ 2* frequency(Y)). If, however, frequency(X) = 60% and frequency(Y) = 40% AND residue X was present in ProWT, then residue Y was selected over X because (frequency(X) ≤ 2 * frequency(Y)).However, if frequency(X) = 60% and frequency(Y) = 40% AND residue X was NOT present in ProWT, then residue X was selected in the computer-generated sequence. Although it is possible that the frequencies of X or Y at any position may be identical, during the analysis of 176 subtilisin homologues this scenario did not arise when A = 2. Because no two residues at any position displayed identical X

or Y frequencies, this algorithm provided us with a unique sequence for ProD (Fig. 4.1*A*).

ProD Displays Low Sequence Identity with ProWT:

Fig. 4.1*A* compares the primary sequence of ProD with ProWT. N1 and N2 represent motifs that were identified using closely related subtilases and appear to be fairly well conserved (Shinde and Inouye, 1993). The sequence identity in Motif N1 and the C-terminal part of Motif N2 is significant, and these motifs are located in β1 and α2-β4, as seen from the x-ray crystallographic structure of ProWT (Jain, Shinde et al., 1998). NMR structure analysis suggests that β1 and α2-β4 display conformational rigidity and may represent potential folding nucleation sites in ProWT (Buevich, Shinde et al., 2001). Hence, motifs that nucleate the folding process appear to be conserved in ProD. When the difference in the total number of individual amino acids between the two propeptides is examined (Fig. 4.1*B*), it is clear that ProWT (pI = 9.76) has an excess of basic residues, whereas ProD (pI = 4.52) is rich in acidic amino acids. However, the hydrophobic patterns within these two peptides are similar (Fig. 4.1*C*), and the residues that constitute the buried hydrophobic core in ProS and ProD are always nonpolar (Fig. 4.1*A*). Properties of ProD and ProWT are summarized in Table 4.T.I.

ProD Can Inhibit Subtilisin Activity and Chaperone Folding of Denatured Subtilisin:

It has been established that ProWT inhibits subtilisin through slow binding inhibition kinetics (Li, Hu et al., 1995). Because the primary sequence and the amino acid composition between ProWT and ProD are significantly different, we examined whether ProD can inhibit the proteolysis of the synthetic substrate and whether this inhibition follows slow binding kinetics (Fig. 4.2*A*). From Fig. 4.2*A*, it is evident that in 10 min the activity of subtilisin (65 nM) drops ~12- and 135-fold in the presence of 0.9 μM of ProWT and ProD, respectively. Hence under these conditions, ProD is an ~10-fold better inhibitor than ProWT and does not demonstrate slow binding inhibition. At lower concentrations, however, ProD demonstrates slow binding inhibition (Fig. 4.2*A*, *inset*), whereas the same concentration of ProWT is not inhibitory (data not shown). We next examined whether ProD can chaperone folding of subtilisin in a trans folding reaction. Denatured subtilisin was folded using ProD and ProS as a chaperone, and the subtilisin activity that was recovered was measured (see Materials and Methods). Fig. 4.2*B* demonstrates that ProWT and ProD can both chaperone folding of subtilisin *in trans*, and the activity recovered increases with the chaperone concentration. When a 40-fold excess of the chaperone was used, ProWT and ProD recovered ~40 and 10 units of activity, respectively. Hence, under these conditions ProD is ~25% as efficient as ProWT in folding subtilisin (Table 4.T.I). We next examined the kinetics of trans degradation of ProD and ProWT as complexes with S_{221}C-subtilisin by TPCKtreated trypsin. Fig. 4.2*C* depicts the kinetics measured using quantitative gel-scanning densitometry as described under Materials and Methods. The S_{221}C-subtilisin is proteolytically inactive and is unable to degrade the inhibitory propeptides. This enables one

to monitor the kinetics of trans degradation of ProD and ProWT, in the absence of autodegradation. The data were fitted to a single exponential rate equation, and rate constants for degradation of ProD and ProWT were estimated as 0.09 and 0.31 min^{-1}, respectively. This suggests that in a complex with $S_{221}C$-subtilisin, ProD is degraded ~3-fold slower than ProWT, which may be attributed to increased structure in the isolated ProD, higher affinity for subtilisin, and fewer cleavage sites for trypsin. To examine whether lower activity recovered by folding mediated by ProD occurs because of incomplete degradation of ProD, an SDS-PAGE analysis of the refolding reaction before and after digestion by TPCKtreated trypsin was performed. Fig. 4.2*D* establishes that both ProD and ProWT (at 20-fold molar excess) are completely degraded by 1 h of trypsin digestion. Moreover, the amount of active subtilisin obtained after refolding and trypsin treatment of the ProD folding reaction (Fig. 4.2*D*, *lane 7*) is lower when compared with that of the ProWT reaction (Fig. 4.2*D*, *lane 4*). It is also important to note that active subtilisin produced during folding can also degrade ProD and ProWT. However, allowing the folded subtilisin to degrade the excess of ProD and ProWT does not significantly improve the units of activity recovered (data not shown). Hence, the data presented in Fig. 4.2*D* suggest that ProD folds subtilisin less efficiently than ProWT.

ProD Is Structured as an Isolated Peptide and Forms a Complex with Subtilisin:

Because ProD can function both as a chaperone and an inhibitor of subtilisin, we next attempted to analyze the secondary structure of ProD in its isolated form and as a complex with S_{221}C-subtilisin using CD spectroscopy. Equimolar concentrations of ProD and ProWT were added to S_{221}C-subtilisin, and the samples were incubated overnight. The S221C substitution at the active site lowers the proteolytic activity by 10,000-fold and is insufficient to degrade the inhibitory propeptide (Jain, Shinde et al., 1998). Far-UV CD spectroscopy establishes that unlike ProWT, which is completely unstructured (Shinde, Li et al., 1993), ProD adopts a well defined α-βconformation (Fig. 4.3*A*). Deconvolution of this spectrum suggests that ProD contains 20% α-helices and 26% β-sheets. Hence, as shown in Eqn. 4.1, the equilibrium (*K*eq) between the folded (Profolded) and unfolded states (Prounfolded), which normally favors the unfolded state in Pro-WT (Ruan, Hoskins et al., 1999), appears to have shifted toward the folded state in the case of ProD,

$$\mathrm{Pro}_{\mathrm{unfolded}} \xleftrightarrow{K_{eq}} \mathrm{Pro}_{\mathrm{folded}} \qquad \text{Equation (4.1)}$$

Where, $K_{eq} = k_f/k_u$.

However, the CD spectra of the complexes of ProD and ProWT with S_{221}C-subtilisin appear to be similar. Because the x-ray structure of the ProWT- S_{221}C-subtilisin complex establishes that the conformation of S_{221}C-subtilisin does not change through its interactions with ProWT (Jain, Shinde et al., 1998), Fig. 4.3*A* suggests that the structure of ProD as a complex with S_{221}C-subtilisin is similar to that of ProWT. ProD contains one tryptophan, four tyrosine, and five phenylalanine residues. Because phenylalanine, tyrosine, and tryptophan contribute to the fluorescence of a protein and are useful probes for monitoring

the solvent accessibility of the side chains, we next examined the intrinsic fluorescence to detect conformational differences between the folded and denatured states of ProD. The protein was excited at 295, and the emission spectra were recorded (Fig. 4.*3B*). ProD shows an emission maximum at 352 and 349 nm for denatured and folded ProD, respectively. The shift in maxima to a lower wavelength (blueshift) suggests that the environment around the tryptophan side chain becomes non-polar upon folding and is consistent with a conformational change within the protein. The 10% decrease in fluorescence intensity of folded ProD may occur because of quenching of the neighboring aromatic residues (Fig. 4.4, *A–C*). We next measured the tertiary structure of ProD under denaturing and non-denaturing conditions using near-UV CD spectroscopy. The near-UV CD profiles of denatured and nondenatured ProD differ only marginally and suggest that ProD may not possess a well-defined tertiary. However, the lack of a tertiary structure in small peptides in not unusual (Greenfield and Fasman, 1969).

Molecular Model of ProD:

From Fig. 4.2, *A–D* and Fig. 4.3, *A–C*, it is evident that (i) ProD functions as an inhibitor and a chaperone of subtilisin similar to ProWT, and (ii) ProD and ProS adopt similar conformations as complexes with S_{221}C-subtilisin. Hence we attempted to build a homology model of ProD- S_{221}C-subtilisin (Fig. 4.4*A*) using the x-ray structure of the ProWT-subtilisin complex (Jain, Shinde et al., 1998) as described under Materials and Methods. The goal of this model is to obtain the spatial orientation of residues within ProD and ProWT, because such

models can guide mutagenesis experiments and hypotheses about structure-function relationships (Baker and Sali, 2001). From this model it is evident that the distribution of the electrostatic potential around the complexes formed by ProD and ProWT is remarkably different (Fig. 4.4*B*). Whereas ProD appears to be completely negative (acidic), ProWT displays a strongly positive electrostatic potential. This is consistent with the finding that ProD contains a large number of acidic amino acids and ProWT displays a bias for basic residues (Table 4.T.I; Fig. 4.1*A*). However, it is interesting to note that the hydrophobic core within ProD and ProWT appears to be maintained (Figs. 1*A* and 4*C*). Hydrophobic interactions can make significant contributions to folding and stability of proteins. The average hydrophobicity of ProD is greater than ProWT. Furthermore, the computed free energy for the transfer of the 10-residue hydrophobic core (Fig. 4.4*C*) from an aqueous to a hydrophobic environment stabilizes ProD by ~2.5 kcal/mol (Table 4.T.I).

IV) DISCUSSION:

We have used the subtilase family as a model to analyze the relation between sequence, charge, structure, and function within propeptides. An indigenously developed algorithm provided a novel peptide, ProD, that displays only 16% sequence identity and has a pI 5 pH units lower than ProWT (Table 4.T.I). However, the overall hydrophobic patterns, the computed secondary structures, and the potential folding nucleation motifs (N1 and N2) appear to be conserved in ProD (Fig. 4.1, *A–C*). These properties seem to be important for function because, despite low sequence identity, ProD and ProWT can function as peptide chaperones and potent protease inhibitors (Fig. 4.2, *A* and *B*). However, ProD is a less efficient chaperone and a more potent inhibitor than ProWT. This confirms that the chaperone and inhibitory functions are not obligatorily related (Fu, Inouye et al., 2000). The differences in the amino acid composition are also reflected in the secondary structures of the isolated ProWT and ProD measured using CD spectroscopy (Fig. 4.3*A*). Unlike ProWT, ProD adopts a well-defined α–β conformation as an isolated peptide and can form a stoichiometric complex with mature subtilisin. The CD spectra of these two complexes are similar and suggest that ProD and ProWT adopt similar structural scaffolds. Based on our results, we propose that the chaperoning activity of the propeptide of the subtilisin family appears to be encoded in the ability to adopt a specific conformation. It is important to note that although the sequence similarity between ProWT and ProD is low, the sequence similarity of the residues that form interface with subtilisin (as judged by three-dimensional structure 1SCJ) is at least 70%. As a result, one would expect the net binding

energy of ProWT and ProD to subtilisin to be very similar. However, in the case of ProWT its binding to subtilisin is coupled to its folding. Hence a part of the interaction energy will be spent on ProWT folding, thereby lowering the ProWT-subtilisin affinity (or inhibition constant). The ProD is already folded; it does not need to expend as much energy on its folding when compared with ProWT. This in turn may result in a higher affinity and hence a 10-fold tighter inhibition constant. Because propeptide domains appear to have emerged through convergent evolution (Eder and Fersht, 1995), this raises another important question. Are there common structural features that define propeptide function in different protein families? To examine this possibility we compared the structures of the propeptides from subtilisin (PDB identifier, 1SCJ), α-lytic protease (1P02), and carboxypeptidase B (1KWM). Fig. 4.5*A* suggests that all three proteins display similar structural scaffolds despite there being no sequence and structural similarities between their cognate protease domains (Baker, 1998). Interestingly, potent subtilisin inhibitors (chymotrypsin inhibitor II (2SNI), streptomyces subtilisin inhibitor (3SSI), and eglin C (1CSE)) that have no known chaperone function also display similar structural scaffolds without any sequence similarity (Fig. 4.5*B*). Does this structural similarity imply a common evolutionary domain? In other words, have propeptides emerged from a primordial protease inhibitor? The conservation of the structural scaffold in ProD, a chaperone designed to digress from the ProWT sequence while maintaining function, suggests that the ability to adopt a specific α-β conformation along with certain specific interactions appears to be crucial for propeptide function. These functions are to chaperone and to inhibit the

protease domain. We suggest that the structural scaffold may contribute to the inhibitory function and certain residue specific interactions may be important for its chaperone function. Such specific interactions may be provided by motifs N1 and N2 in the subtilase family and may vary between families. Although the structural similarities are striking in Fig. 4.5, *A* and *B*, it is also possible that they are purely co-incidental. The functional conservation in ProD, although having a difference of 5 pH units, suggests that charge *per se* may not be important for propeptide function. So why is there a large bias for charge in propeptide domains? Also, if the propeptides within a family adopt similar structures and perform similar functions, why do the primary sequences of propeptide diverge faster than their cognate catalytic domains? One possibility is that the charged N terminus increases the solubility of the precursor protein. Another possibility is that the environment in which these proteins are required to become active influences sequence digressions. For example, subtilisin is secreted out by *B. subtilis* for the purpose of scavenging food, whereas mammalian cells transport the eukaryotic subtilisin homologue furin into the trans Golgi network for processing precursor proteins (Anderson, Molloy et al., 2002). Although the catalytic domains of bacterial subtilisin and their eukaryotic counterparts are similar and robustly stable (Shinde and Inouye, 2000), the propeptides in the above example are removed in dramatically different environments (Yabuta, Takagi et al., 2001). We propose that the amino acid differences between their propeptides allow these proteins to respond to different environments to modulate activation precision of proteases. Another noteworthy point is that our

work raises questions regarding the lower limit for sequence identity that is acceptable for two proteins to perform similar functions.

IV) MATERIALS AND METHODS

Algorithm for Designing a Peptide Chaperone:

The PSI-BLAST (position-specific iterated basic local alignment search tool) program (Altschul, Madden et al., 1997) was used to obtain a multiple sequence alignment of proteins that are similar to ProWT (77 residues) using a threshold E-value of 0.8. This alignment was treated as a matrix and utilized to obtain the percentage of every individual amino acid at every position in the 77-residue polypeptide. The matrix contained 176 homologues of subtilisin. The computed percentage was employed in a decision-based algorithm that searched for the following criteria. If "X" is the most abundant amino acid residue and "Y" is the second most abundant residue at a particular position in the sequence, then residue "X" is always chosen over Y UNLESS (frequency(X) ≤ A* frequency(Y)) AND residue X is present in ProWT, where "A" represents a "cutoff factor." In the present analysis A = 2 and no two residues at any position display identical frequencies. The aim of the algorithm was to establish the absolute minimum sequence requirement within divergent propeptide sequences. The computed sequence ProD, which was designed to digress from ProWT, displays 16% sequence identity and opposite net charge (Fig. 4.1*A*).

Gene Synthesis of ProD:

The nucleotide sequence corresponding to ProD was obtained using the BACK-TRANSLATE module from the Genetics Computer Group Wisconsin

Package program (GCG) and was based on the *E. coli* codon usage table. The nucleotide sequence was synthesized using a combination of DNA synthesis and PCR. The product was cloned into pET11a under the control of the T7-promoter (Yabuta, Takagi et al., 2001), and the nucleotide sequence was verified on an ABI-310 DNA sequencer.

Expression and Purification of ProD and S_{221}C-subtilisin:

E. coli BL21(DE3) cells were transformed with plasmids pET-Pro- S_{221}C-subtilisin and pET-ProD to obtain S_{221}C-subtilisin and ProD, respectively. Cells were grown in M9 medium supplemented with 50µg/ml ampicillin (Fu, Inouye et al., 2000). At an absorbance of 0.6 *A*600 nm, the culture was rapidly cooled to 12 °C. Adding isopropyl β-D-thiogalactopyranoside to a final concentration of 1 mM induced protein expression. After 16 h at 12 °C, the cells were harvested as described earlier (Yabuta, Takagi et al., 2001). Under these conditions ProWT-S_{221}C-subtilisin and ProD remain in the soluble fraction. The samples were then centrifuged at 20,000 x *g* to remove the insoluble cell debris. The supernatant that contained ProWT-S_{221}C-subtilisin was then incubated at 4 °C to allow the degradation of the inhibitory ProWT through cellular proteases. The cell lysates were dialyzed overnight (using a 3-kDa cutoff) against 100 vol of 50 mM Tris-HCl, pH 8.6, and 1mM CaCl2. The proteins were loaded onto a HighQ column (BioRad) equilibrated with 50 mM Tris-HCl, pH 8.6, and then eluted using a 0–1M NaCl gradient. Fractions containing the desired proteins were pooled together and dialyzed against 50 mM Tris-HCl, pH 7.0, containing 0.5 M ammonium sulfate, 2 mM CaCl2 and concentrated by ultra-filtration using a 3-

kDa cutoff membrane. Concentrated ProD was further purified using a Superdex-75 gel-filtration column (Amersham Biosciences). Both ProD and S_{221}C-subtilisin were purified using the above approach. The only difference was that a Superdex-200 gel-filtration column was used to obtain mature S_{221}C-subtilisin.

Trans Folding of Denatured Subtilisin:

6 M guanidine hydrochloride denatured subtilisin (0.9 µM) was mixed with varying concentrations of ProD (0–36 µM) to a final volume of 100 µl. ProWT was used as a control. The mixture was dialyzed for 16 h against 50 mM MES, pH 6.5, containing 0.5 M ammonium sulfate and 1 mM CaCl2. A 20-µl aliquot of the renaturation mixture was treated with ~7.0 units/ml of TPCK-treated trypsin for 1 h to degrade the excess of propeptide. This is necessary because the propeptide can also inhibit subtilisin activity(Shinde, Li et al., 1993).Enzyme activity was measured using *N*-suc-AAPF-pNA as a synthetic substrate (Marie-Claire, Yabuta et al., 2001). TPCK-treated trypsin does not cleave the synthetic substrate and hence does not interfere with the activity assay.

Inhibition of Subtilisin E Activity:

Active subtilisin (65 nM) was rapidly added to the protease assay buffer that contained 0.5 µM synthetic substrate and 0.9 µM of ProD and ProWT. Protease activity was monitored as described earlier (Fu, Inouye et al., 2000; Marie-Claire, Yabuta et al., 2001; Yabuta, Takagi et al., 2001).

Complex Formation between ProD and S_{221}C-subtilisin—ProD and ProWT (6.4 μM) were added to an equimolar amount of mature S_{221}C-subtilisin in the folding buffer (50 mM MES, pH 6.5, 0.5 M ammonium sulfate, 2 mM CaCl2, and 1 mM β mercaptoethanol). The mixtures were incubated overnight at 4 °C to allow complex formation.

Trypsin-mediated Trans Degradation of ProD and ProWT Complexes with S_{221}C-subtilisin:

The trans complex was obtained as discussed above. The ProD- and ProWT- S_{221}C-subtilisin complexes (50 nM) were incubated with TPCK-treated trypsin (2.0 units/ml), and aliquots were removed at different time intervals. The reaction was terminated through trichloroacetic acid precipitation as described earlier (Yabuta, Takagi et al., 2001), and the aliquots were separated using SDS-PAGE. The amount of residual propeptide was estimated using quantitative gel-scanning densitometry.

Circular Dichroism and Fluorescence Measurements:

CD measurements were performed on an automated AVIV 215 spectrometer at 25 °C. Protein concentration was maintained between 0.05–0.2 mg/ml (Yabuta, Takagi et al., 2001). Spectra were taken between 190–260 nm using a 1-mm path length cuvette (Fig. 4.2*C*) and represent averages of three independent scans. For the fluorescence measurements, the denatured and folded ProD (7.0μM) were excited at 295 nm, and the emission spectra from

310 to 410 nm were recorded as described earlier (Marie-Claire, Yabuta et al., 2001).

Molecular Modeling:

Homology modeling was performed using LOOK, Version 2.0. The sequence alignment was performed using the BestFit module of the Genetics Computer Group software (GCG), and the alignment was imported into LOOK. A structural model for ProD-S_{221}C-subtilisin was built, and the energy minimization was carried out using the GROMOS 43B1 force field using the Swiss-Model software, Version 3.7 (Guex and Peitsch, 1997). Because the ProWT and ProD contain more than 36% charged residues, the electrostatic potentials were computed using coulomb interactions by taking into account the charged residues and assuming a pH of 7.0 (Guex and Peitsch, 1997). The electrostatic charges were mapped to the molecular surface (Fig. 4.4*B*).

VI) FIGURES

Fig. 4.1. Design criteria and computational characterization of ProD. (A), primary sequence alignment between the subtilisin propeptide (*ProWT*) and the redesigned propeptide (*ProD*). Identical residues (*black background*), highly conserved substitutions (*gray background*), and less conserved substitutions (*boldface*) were identified using the program BestFit. The secondary structure of ProWT (obtained from the x-ray structure) when it forms a complex with subtilisin is depicted *above* the sequence alignment; the secondary structure computed using PREDICT-PROTEIN is depicted *below* the sequence alignment. *C*, *E*, and *H* denote coils, β-sheets, and α -helices, respectively. The predicted structure in boldface represents predictions that do not match with the secondary structure obtained from the crystallized complex of ProWT-subtilisin. Motifs N1 and N2 represent the conserved domains within the subtilase family. *Asterisks* denote residues that constitute the hydrophobic core within the propeptide. **(B)**, the difference in the amino acid composition between ProWT and ProD. *Positive values* represent those residues that are in excess in ProWT; *negative values* indicate residues that are in excess in ProD. **(C)**, hydropathy plot comparison between ProD (*black line*) and ProWT (*gray line*). **(D),** change in sequence identity of the computed peptide with ProWT as a function of Cutoff Factor (A).

Fig. 4.1

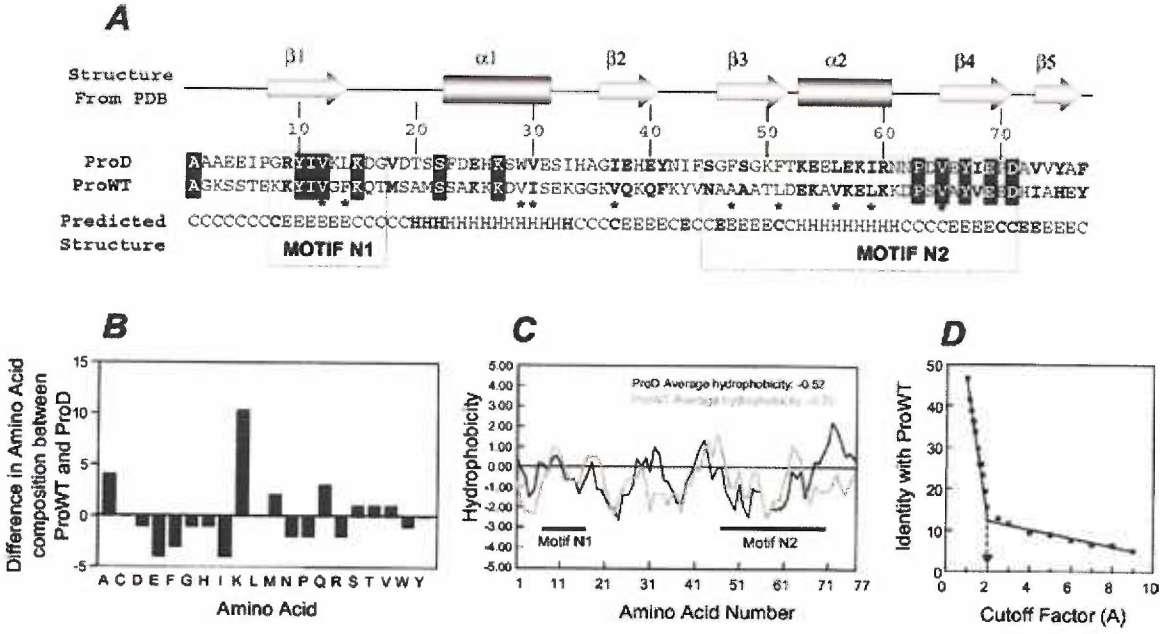

A
Structure From PDB
β1
α1
β2
β3
α2
β4
β5
10
20
30
40
50
60
70
ProD
ProWT
Predicted Structure
MOTIF N1
MOTIF N2
B
Difference in Amino Acid composition between ProWT and ProD
A C D E F G H I K L M N P Q R S T V W Y
Amino Acid
C
ProD Average hydrophobicity: -0.52
Hydrophobicity
Motif N1
Motif N2
Amino Acid Number
D
Identity with ProWT
Cutoff Factor (A)

Fig. 4.2. Biochemical characterization of ProD. **(A)**, inhibition of subtilisin activity as a function of time. Subtilisin (65 nM) rapidly degrades *N*-suc-AAPF-*p*NA to produce *p*-NA, which can be monitored spectroscopically (*black line without symbols*). The ProWT (*open circles* and *triangles* represent 0.8 and 1.2 μM of ProWT, respectively) can inhibit proteolysis of the substrate, and this inhibition follows slow binding kinetics. However, 0.8 (*filled circles*) and 1.2 μM (*filled triangles*) of ProD almost completely blocks subtilisin activity. *Inset* depicts slow binding inhibition of subtilisin by low concentrations of ProD (*inverted triangles*, 0.45 μM; *filled diamonds*, 0.30 μM). **(B)**, units of subtilisin activity recovered by refolding denatured subtilisin E using ProD (*filled circles*) and ProWT (*open circles*) as peptide chaperones. Refolding was carried out as described under Materials and Methods, and the amount of denatured subtilisin was maintained at 0.9 μM, whereas the concentrations of ProD and ProWT were varied from 0 to 36 μM. **(C)**, kinetics of trypsin mediated degradation of ProD (*filled circles*) and ProWT (*open circles*) from their cognate inhibition complexes with S221C-subtilisin. Amount of residual ProD and ProWT were estimated from SDS-PAGE using quantitative gel-scanning densitometry as described under Materials and Methods. **(D)**, SDS-PAGE analysis of a trans refolding reaction. Mature subtilisin (*band A*) was refolded using ProWT (*band C*) and ProD (*band D*) as described under Materials and Methods. The subtilisin-propeptides (1:20 ratio) (*lanes 2* and *5*) were dialyzed against refolding buffer. After 16 h, subtilisin folded by ProWT and ProD (*lanes 3* and *6*) were treated with trypsin (*band B*) as described under Materials and Methods. Trypsin degrades the excess of ProWT and ProD within 1 h (*lanes 4* and *7*) to give active subtilisin.

Fig. 4.2

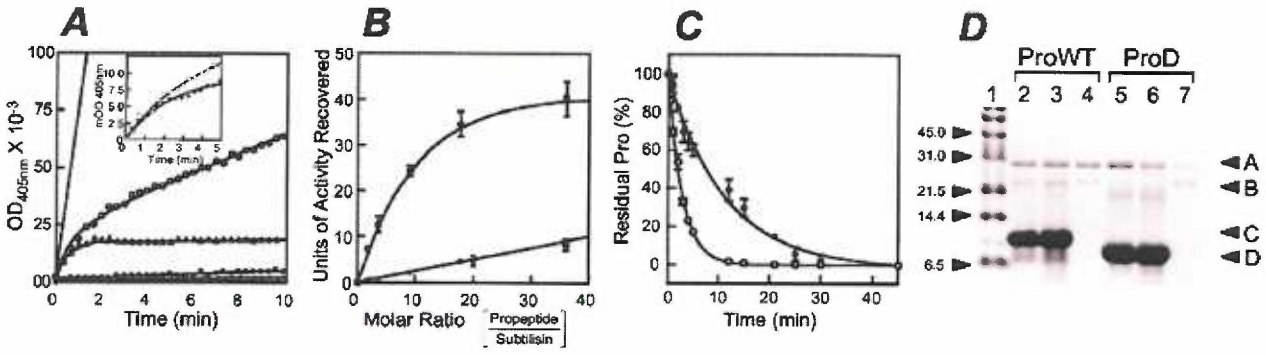

A
B
C
D
Time (min)
Molar Ratio
Propeptide / Subtilisin
Units of Activity Recovered
Residual Pro (%)
ProWT
ProD
1 2 3 4 5 6 7
45.0
31.0
21.5
14.4
6.5
A
B
C
D

Fig. 4.3. Biophysical characterization of ProD. (A), CD spectroscopy of ProD (*thick black*) and ProWT (*gray*) alone (*dotted*) and as stoichiometric complexes (*solid*) with mature S221C-subtilisin (*thin black line*). The difference spectra between ProD-S221C-subtilisin and mature subtilisin alone give the spectra of ProD (*black dashes*) in a complex with subtilisin. The gray dashed line represents the structure of ProWT when complexed with S221C-subtilisin obtained using the difference spectra. **(B)**, tryptophan fluorescence spectra of denatured (*dotted line*) and folded ProD (*dashed line*) excited at 295 nm. *Arrows* indicate the fluorescence maxima. **(C)**, near-UV CD spectra for denatured (*dotted line*) and folded ProD (*dashed line*). Denatured (*thick solid line*) and folded (*thin solid line*) subtilisin is used as a control for folded and unfolded protein.

Fig. 4.3

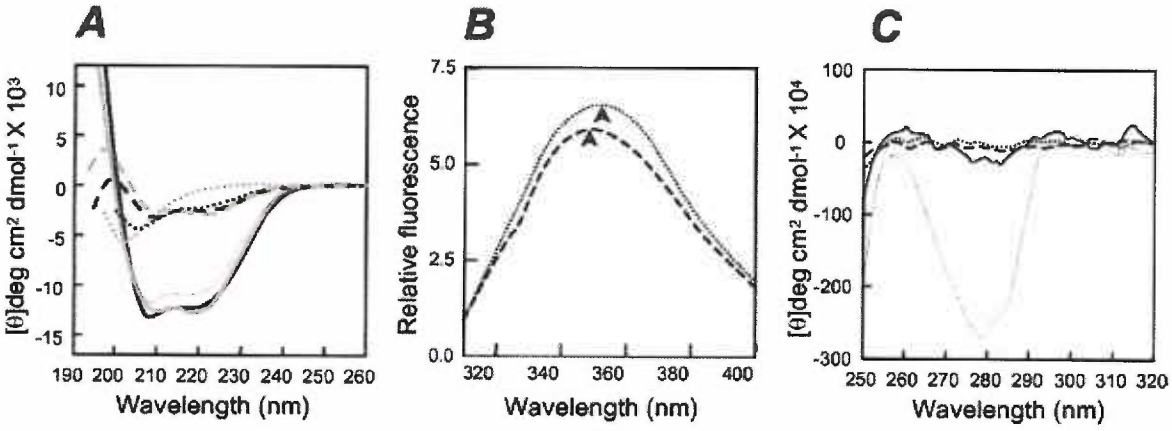

A
B
C
[θ]deg cm² dmol⁻¹ X 10³
10
5
0
-5
-10
-15
190 200 210 220 230 240 250 260
Wavelength (nm)
Relative fluorescence
7.5
5.0
2.5
0.0
320 340 360 380 400
Wavelength (nm)
[θ]deg cm² dmol⁻¹ X 10⁴
100
0
-100
-200
-300
250 260 270 280 290 300 310 320
Wavelength (nm)

Fig. 4.4. **(A)**, ribbon diagram depicting x-ray structure of the ProWT·subtilisin complex (*left*) and the homology model for ProDsubtilisin (*right*). **(B)**, the electrostatic potential mapped to the molecular surfaces of the two complexes. *Red* depicts negative charges. *Blue* indicates positive charges. **(C)**, the side chains that contribute to the hydrophobic core in ProWT and ProD. The extended C terminus interacts with the substrate binding loops to inhibit subtilisin activity

Fig. 4.4

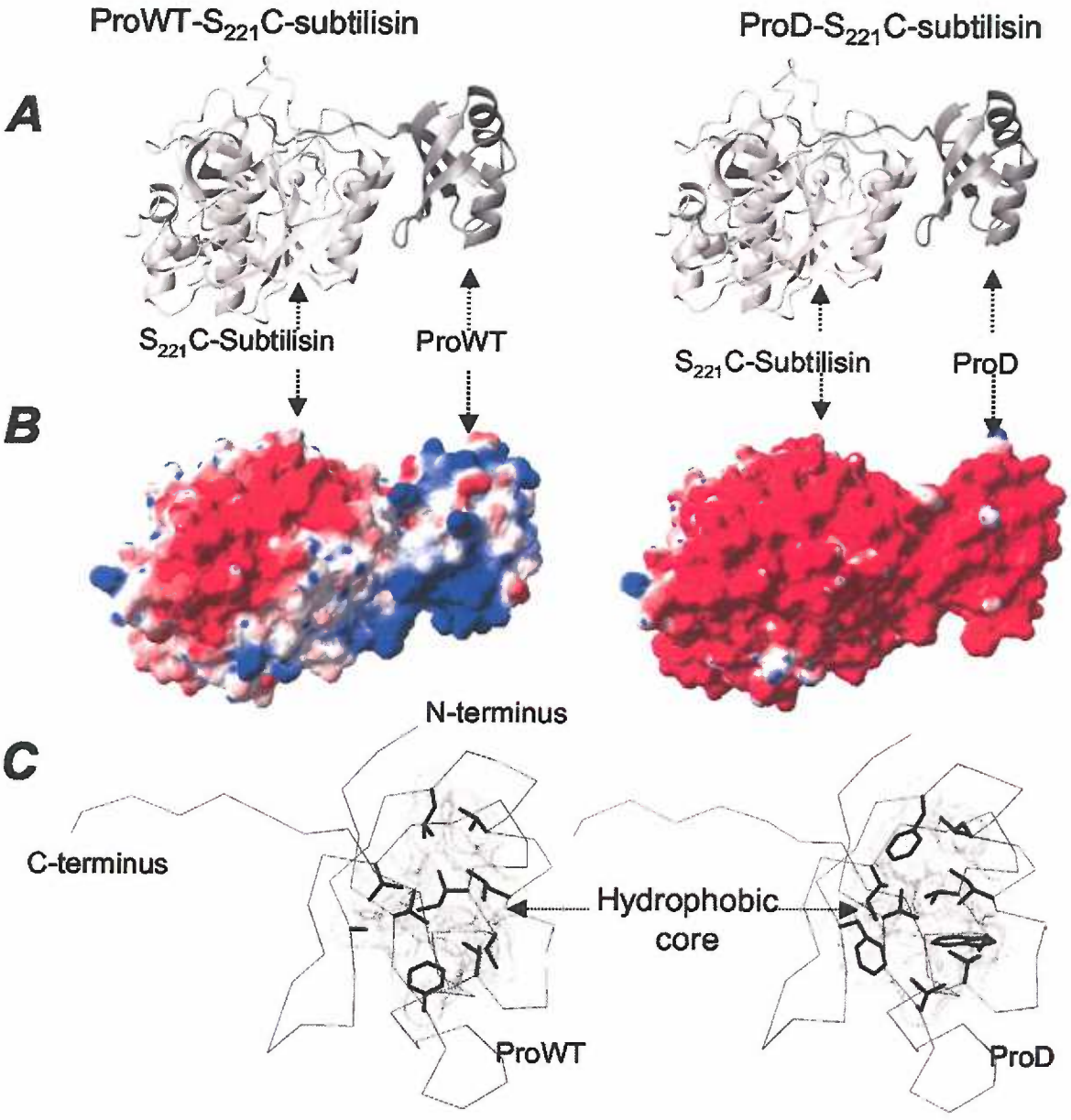
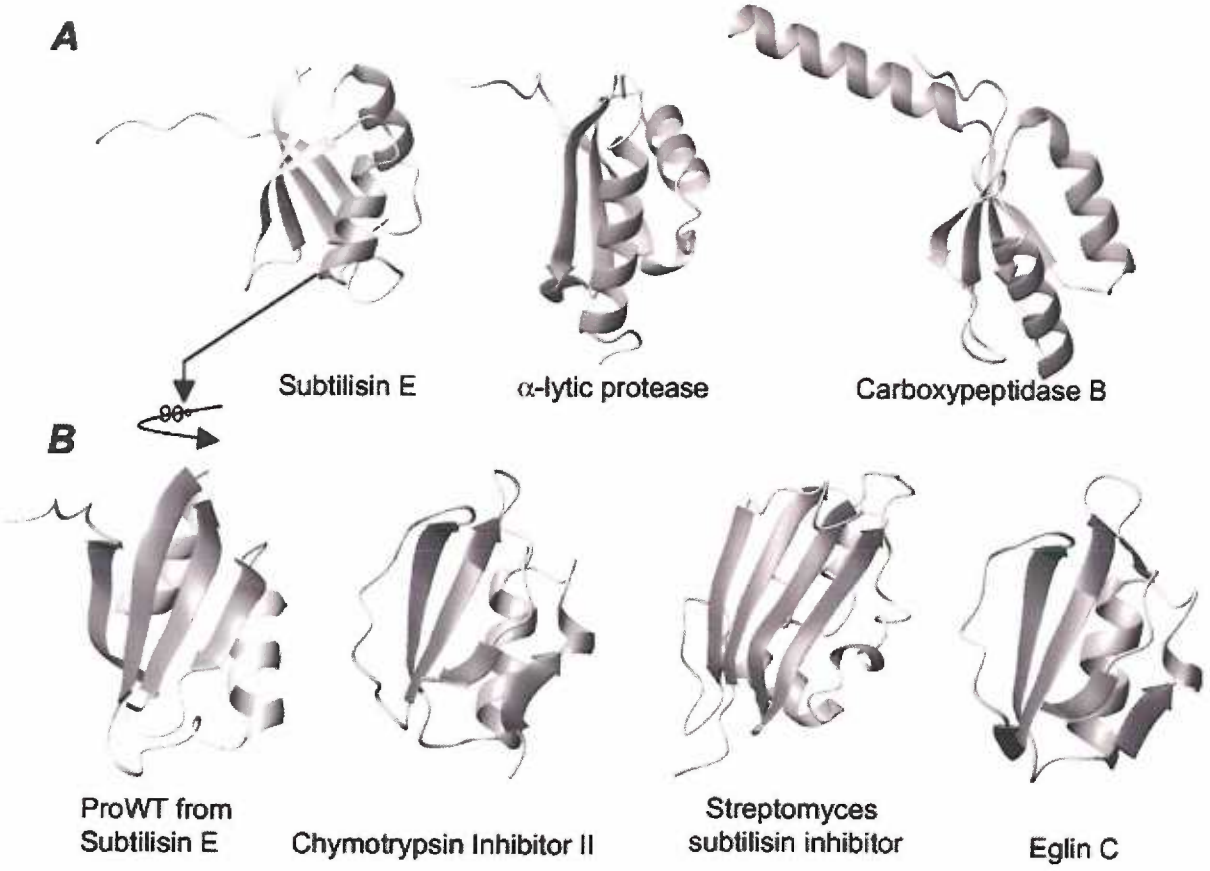

ProWT-S221C-subtilisin
ProD-S221C-subtilisin
A
S221C-Subtilisin
ProWT
S221C-Subtilisin
ProD
B
N-terminus
C
C-terminus
Hydrophobic core
ProWT
ProD

Fig. 4.5. The ribbon diagrams for (**A**) propeptides of subtilisin, α-lytic protease, and carboxypeptidase B and (**B**) ProWT, chymotrypsin inhibitor II, *streptomyces* subtilisin inhibitor, and eglin C.

Fig. 4.5


A
Subtilisin E
α-lytic protease
Carboxypeptidase B
90°
B
ProWT from
Subtilisin E
Chymotrypsin Inhibitor II
Streptomyces
subtilisin inhibitor
Eglin C

VII) TABLES

Table 4.T.1 Properties of the isolated peptide chaperones

	ProWT	ProD
Length	77	77
M.W.	8495	8764
Calculated pI	9.76	4.52
Identity to ProWT	100%	15.6%
Hydrophobicity		
Total average [a]	-0.52	-0.70
Core [b]	-22.8	-25.3
Structure	random	α–β
$K_i \times 10^{-9}$ M	8.36	0.74
Folding efficiency	100%	25%

[a] Kyte & Doolittle method

[b] Computed ΔG for burial of hydrophobic core (kcal/mol) using free energy transfer values defined by Bigelow

CHAPTER 5

POSITIVE SELECTION DICTATES CHOICE BETWEEN KINETIC AND THERMODYNAMIC PROTEIN FOLDING AND STABILITY IN SUBTILASES

Ezhilkani Subbian, Yukihiro Yabuta[1] and Ujwal Shinde

Department of Biochemistry and Molecular Biology,

Oregon Health and Science University,

3181 SW Sam Jackson Park Road, Portland, OR 97239, USA

[1]Current Address: RIKEN Center for Developmental Biology, Kobe, Japan

Biochemistry

Vol. 43(45), 14348-60; Nov 16 2004

I) ABSTRACT:

Subtilisin E (SbtE) is a member of the ubiquitous super-family of serine proteases called subtilases and serves as a model for understanding propeptide mediated protein-folding mechanisms. Unlike most proteins that adopt thermodynamically stable conformations, the native state of SbtE is trapped into a kinetically stable conformation. While kinetic stability offers distinct functional advantages to the native state, the constraints that dictate the selection between kinetic and thermodynamic folding and stability remain unknown. Using highly conserved subtilases, we demonstrate that adaptive evolution of sequence dictates selection of folding pathways. Intracellular and extracellular serine proteases (ISPs and ESPs, respectively) constitute two subfamilies within subtilases that have highly conserved sequences, structures and catalytic-activities. Our studies on the folding pathways of Subtilisin E (SbtE), an ESP, and its homologue Intracellular Serine protease 1 (ISP1), an ISP, show that although topology, contact order, and hydrophobicity that drive protein-folding reactions are conserved, ISP1 and SbtE fold through significantly different pathways and kinetics. While SbtE absolutely requires the propeptide to fold into a kinetically-trapped conformer, ISP1 folds to a thermodynamically stable state over a million times faster and independent of a propeptide. Furthermore, kinetics establishes that ISP1 and SbtE fold through different intermediate states. An evolutionary analysis of folding constraints in subtilases suggests that observed differences in folding pathways might be mediated through positive selection of specific residues that map mostly onto the protein surface. Together, our results demonstrate that closely related subtilases can

fold through distinct pathways and mechanisms, and suggest that fine sequence details can dictate the choice between kinetic versus thermodynamic folding and stability.

II) INTRODUCTION:

The question of how proteins find their unique native structures by using information encoded within their amino acid sequences lies at the heart of molecular biology. Understanding mechanisms of folding/unfolding are of immense scientific significance because of their potential, in providing new approaches to combat protein-misfolding disorders and in facilitating the design of novel proteins that catalyze desired reactions. In addition the protein folding problem is ostensibly the most fundamental example of biological self-assembly, and symbolizes the first step towards tackling one of the most arduous questions addressable by contemporary science - *How have complex biological systems evolved the ability to self-assemble into robust and predictable systems?* The current literature suggests that each sequence folds through a unique energy landscape that is dictated by the intrinsic properties of the polypeptide, and by the extrinsic influence of the folding environment {Dobson, 2004 }. The challenge now is to recognize how an energy landscape that is inherent to a given protein is defined by its amino acid sequence, and to what extent do external factors influence this process.

A popular approach to gain insights into general protein-folding principles and, to determine what characteristics of this process are under evolutionary control is to compare folding of homologous proteins (Shakhnovich, Abkevich et al. 1996; Nishimura C, Prytulla S et al. 2000; Plaxco, Simons et al. 2000; Gunasekaran, Eyles et al. 2001). Several elegant studies have demonstrated that homologous proteins fold through similar folding pathways and transition states (Shakhnovich, Abkevich et al. 1996; Raschke and Marqusee 1997;

Gunasekaran, Eyles et al. 2001; Mirny and Shakhnovich 2001; Larson, Ruczinski et al. 2002). There is also increasing evidence that the folding rates of proteins with diverse sequences strongly correlate with more global parameters such as contact order, topology, average hydrophobicity and protein length (Makarov and Plaxco 2003). In essence, fine sequence details of the protein do not affect folding pathways and mechanisms (Plaxco, Simons et al. 2000; Tseng and Liang 2004).

While the current view of folding has emerged through studies based on proteins that fold into their global energy minima (thermodynamically stable states), there is increasing evidence that several proteins can fold to kinetically stable local minima (Baldwin, Ziegler et al. 1993; Carr, Chaudhry et al. 1997; Shinde, Liu et al. 1997; Sohl, Jaswal et al. 1998; Im, Seo et al. 1999; Huntington, Read et al. 2000; Shinde and Inouye 2000; Bryan 2002; Cunningham and Agard 2003; Cunningham and Agard 2004). While thermodynamically stable protein conformations are generally the norm, kinetically stable conformations may be selected for specific functional advantages. For example, kinetic stability provides mechanisms for protease longevity in α-lytic protease (Sohl, Jaswal et al. 1998; Bryan 2002; Jaswal, Sohl et al. 2002), for biological regulation in serpins (Im, Seo et al. 1999; Huntington, Read et al. 2000) and, for membrane insertion of viral glycoproteins (Carr, Chaudhry et al. 1997). In case of the broadly specific α-lytic protease, the uncoupling of the folding and unfolding pathways upon propeptide degradation enables the native state to be locked in a metastable state that confers low conformational entropy. This enhances the longevity of the proteases by

reducing its autoproteolytic susceptibility (Sohl, Jaswal et al. 1998; Jaswal, Sohl et al. 2002). Studies on plasminogen activator inhibitor 1 have shown that the active inhibitory form converts to a latent inactive form over a period of several hours. The latent form can be converted back to the active form by denaturation and renaturation. Thus the active inhibitory form of the protein is not at its lowest energy state but slowly converts to the latent form that appears to have higher stability (Baker and Agard 1994; Berkenpas, Lawrence et al. 1995; Lawrence, Ginsburg et al. 1995). In some serpins, this conversion from the active to the latent form occurs upon proteolysis (Im, Seo et al. 1999; Huntington, Read et al. 2000; Im, Woo et al. 2002). These provide unique methods for inhibition of serine protease activity. In case of hemagglutinin, the trimeric envelope glycoprotein undergoes a dramatic loop to coiled-coil transition upon exposure to low pH. The initially synthesized native state has a disordered loop region that appears to be separated from a more stable state by a kinetic barrier. The conformational change upon pH change facilitates membrane insertion (Carr, Chaudhry et al. 1997; Shental-Bechor, Danieli et al. 2002; Swalley, Baker et al. 2004). Hence, in the cases discussed above, the kinetically trapped native conformations enhance the biological functionality of proteins. Since the folding landscape is encoded in the polypeptide sequence, the above examples raise an important question whether nature selects for specific folding pathways through fine sequence details. Agard and coworkers have compared folding of the kinetically trapped α-lytic protease with thermodynamically stable trypsin (Light and al-Obeidi 1991) to understand the functional basis of thermodynamic versus kinetic stability (Cunningham, Jaswal

et al. 1999; Cunningham, Mau et al. 2002; Cunningham and Agard 2003; Cunningham and Agard 2004). However, although α-lytic protease and trypsin adopt very similar folds, their amino acid chains display insignificant sequence similarity (Jaswal, Sohl et al. 2002). This prevents an analysis of how primary sequence evolution dictates selection between thermodynamic versus kinetic stability because, α-lytic protease and trypsin may have evolved via convergent evolution (Eder and Fersht 1995). The key to understanding how protein sequences have evolved to select between thermodynamic and kinetic stability is to identify protein sub-families closely related by sequence and structure, but different in their folding mechanisms, such that one sub-family folds into thermodynamically stable native states, while the other adopts kinetically trapped conformers. This would facilitate a detailed analysis of the relation between primary sequences, native structures, folding pathways and evolutionary constraints that dictate the choice between thermodynamic and kinetic stability.

Subtilases constitute a large ubiquitous super-family of calcium dependent serine proteases that span prokaryotes, eukaryotes and archea (Siezen and Leunissen 1997; Shinde and Inouye 2000; Bryan 2002). On the basis of their sequence homology, subtilases can be divided in at least six families, all of which have conserved catalytic domains (Siezen and Leunissen 1997). Although levels of overall sequence identity may vary between these families, they all adopt very similar three-dimensional structures. Most members of the subtilase super-family are produced with dedicated propeptide-domains located down-stream to signal peptides. These propeptides are demonstrated to be

essential for correct folding of their cognate protease domains in several prokaryotic and eukaryotic subtilases (Siezen and Leunissen 1997; Shinde and Inouye 2000). Propeptide deletion variants of these subtilases are correctly localized but are unable to fold into their native states, suggesting that propeptides do more than merely assist signal peptides to correctly localize their protease domains. Contemporary understanding of propeptide mediated folding mechanisms of subtilases have emerged from analyses of bacterial subtilisins (Shinde and Inouye 2000; Bryan 2002).

In the present manuscript we investigate folding in the highly conserved subtilase super-family (Siezen and Leunissen 1997; Shinde and Inouye 2000). Like other subtilases, subtilisin-E (SbtE) is a calcium-dependent ESP that is secreted as a precursor (pro-SbtE), and is an established model for understanding propeptide-dependent folding (Bryan, Alexander et al. 1992; Eder, Rheinnecker et al. 1993; Shinde and Inouye 1995; Shinde, Liu et al. 1997; Siezen and Leunissen 1997; Jain, Shinde et al. 1998; Fu, Inouye et al. 2000; Shinde and Inouye 2000; Bryan 2002; Yabuta, Subbian et al. 2002; Yabuta, Subbian et al. 2003). In absence of its propeptide (77-residue), SbtE (275-residue) folds into a stable, molten-globule intermediate that is separated from its kinetically-trapped native state by a high-energy barrier (Eder, Rheinnecker et al. 1993; Bryan 2002). Spontaneous conversion to the native state is slow ($t_{½}$~1500 yrs), but is catalyzed ~10^6-fold by the propeptide, which lowers this high-energy barrier (Eder, Rheinnecker et al. 1993; Bryan 2002). Hence, propeptides function as dedicated intramolecular chaperones essential for efficient folding of their cognate protease-domains to kinetically-trapped native

states (Baker and Agard 1994; Shinde and Inouye 2000; Yabuta, Takagi et al. 2001; Jaswal, Sohl et al. 2002). A database search of more than 500 homologues of SbtE identified several bacterial intracellular serine proteases that are highly conserved in sequence but lack classical propeptide domains. Through detailed analyses of amino acid and nucleotide sequences, protein structures, and folding kinetics, we demonstrate that Intracellular Serine protease 1 (ISP1), an ISP, and its homologue SbtE, an ESP from the same organism (*Bacillus subtilis*), can fold through significantly different pathways and kinetics. Despite displaying conserved topology, contact order, and hydrophobicity, SbtE absolutely requires assistance from its propeptide to fold into a kinetically-trapped conformer while, ISP1 folds to a thermodynamically stable state over a million times faster and independent of a propeptide. Analysis of the themodynamics and kinetics suggests that the folding transition state of ISP1 is different from that of SbtE. Interestingly, the differences between ESPs and ISPs appear localized mostly on the protein surfaces, with ISPs displaying a bias towards acidic amino acid residues (pI <5.5) and ESPs preferring basic amino acid residues (pI > 8.0). In addition, ISPs lose the specificity and affinity for the propeptides of ESPs but continue to bind tightly with known subtilisin inhibitors. Evolutionary analysis suggests that the primary sequence of ISPs display significant adaptive evolution of surface residues. Together our results suggest that surface residues dramatically influence folding, and their positive-selection can dictate folding pathways, mechanisms, and the choice between kinetic versus thermodynamically stable folds in subtilases.

III) RESULTS:

Sequence, Structural and Functional Analysis of Bacillus ESPs and ISPs:

A database search for SbtE-like bacterial proteases identified two subfamilies (ISPs and ESPs) within subtilases that exhibit significant sequence, structure, and functional conservation (Siezen and Leunissen 1997; Shinde and Inouye 2000). High-levels of sequence identity (~50%) within these subgroups (Fig. 5.1a), and availability of crystallographic structures for several ESPs, including SbtE (Jain, Shinde et al. 1998), allows development and validation of reliable homology models (Baker and Sali 2001) for ISPs (Fig. 5.2a). Subtilases with even low sequence identity have been established to adopt similar three-dimensional structures (Siezen and Leunissen 1997). For example, the eukaryotic subtilases Kex2 and furin (prohormone convertases) adopt structural scaffolds (Henrich, Cameron et al. 2003; Holyoak, Wilson et al. 2003) that are similar to bacterial subtilisins despite low sequence identity (26% with Kex and 37% with furin). Since ISP1 and SbtE have >50% sequence identity we modeled the structure of ISP1 using subtilase templates. Due to the high levels of sequence identity the modeled structure of ISP1 can be considered with confidence. When sequence conservation seen in the multiple sequence alignment (Fig. 5.1a) is mapped on the structure of ESPs and ISPs, conserved sites (black and grey highlight) map to the protein core, while sequence differences between the two families (Blue and Red typeface) map predominantly onto the protein surface (Fig. 5.2b and Fig. 5.2c). Analysis of the

amino acid distribution in the two families demonstrates that the percentage of hydrophobic residues within ESPs and ISPs are similar (see amino-acid distribution in Fig. 5.1a). Further, quantitation (Table 5.T.1) of the hydrophobic core establishes similar core strength (see Material and Methods for details, Fig. 5.2b) within ESPs and ISPs. Calcium-binding sites that are critical for ESP-stability (Bryan 2002; Yabuta, Subbian et al. 2002), are also well-conserved within ISPs (green bars in Fig. 5.1a). However, although ISPs and ESPs share significant sequence/structural similarities, ISPs lack propeptides that are critical for ESP-folding. The propeptides of ESPs are extremely charged [36% charge] and intrinsically unstructured polypeptides (Shinde and Inouye 1993; Shinde, Li et al. 1993). Interestingly, ISPs are biased for negatively charged residues as evident by their acidic isoelectric points (Fig. 5.1a; Table 1) and charges that are localized within propeptides of ESPs appear to be evenly distributed within the protease-domain of ISPs.

To obtain insights into structure and folding of ISPs, we cloned, expressed, and purified ISP1 (intracellular serine protease 1) from *Bacillus subtilis* (Material and Methods), to compare and contrast with SbtE, our ESP model from the same species (Shinde and Inouye 2000; Yabuta, Takagi et al. 2001). The secondary structure of SbtE, pro-SbtE, ISP1 and their active-site variants were compared using circular dichroism (CD) spectroscopy and the fractions of secondary structure content were estimated as described (Greenfield and Fasman 1969; Fu, Inouye et al. 2000; Yabuta, Takagi et al. 2001; Yabuta, Subbian et al. 2003). Fig. 5.3a demonstrates that the secondary structure spectra of ISP1 and SbtE are super-imposable and their deconvolution yields

similar overall secondary structure content (Table 5.T.1), and is consistent with our homology-model. Thermo-stability estimated using changes in CD ellipticity as a function of temperature indicates that the active site variants of ISP1 and SbtE display similar melting temperatures (Table 5.T.1). Both ISP1 and SbtE exhibit blue shift and increase in intrinsic tryptophan fluorescence with an emission maximum at 330nm, when compared with their unfolded states (Fig. 5.3b). Due to the presence of an additional tryptophan residue, native SbtE displays higher fluorescence relative to ISP1. The change in fluorescence observed upon denaturation and renaturation can be used to monitor folding-kinetics of ISP1 and SbtE.

We next compared enzymatic properties of ISP1 and SbtE. ISP1 is completely inhibited by Ethylenediaminetetraacetic acid and Phenylmethylsulfonyl fluoride similar to SbtE (data not shown) and exhibits a 15-fold higher specific activity towards a chromogenic substrate (Table 5.T.1). Since subtilases are known to be dependent upon Ca^{2+} ions for their stability and/or catalytic activity, the relative affinities of ISP1 and SbtE for Ca^{2+} were estimated as described in the Material and Methods. Both proteases display strong Ca^{2+}-dependence for their stability/activity, with ISP1 having only a 3-fold lower Ca^{2+}-dependent proteolytic stability (Fig. 5.4a).

To gain deeper insight into the similarities and differences between the enzymatic properties of ISP1 and SbtE, their binding affinities for known subtilase inhibitors were examined. *Streptomyces* Subtilisin Inhibitor (SSI), one of the strongest inhibitors of bacterial subtilases (Li, Hu et al. 1995) can interact with SbtE (K_I = 0.17nM) and ISP1 (K_I = 0.26nM) with similar affinity (Table 5.T.1,

Fig. 5.4b). The propeptide inhibits SbtE with a K_I of 18.5nM. While this affinity is approximately 10-fold weaker than that observed for SSI, it is similar to the affinity of other ESPs such as subtilisin BPN' and Carlsberg (Li, Hu et al. 1995), for the propeptide of SbtE. On the other hand, the propeptide inhibits ISP1 with an affinity (K_I = 1410 nM) that is ~100-fold weaker than that for SbtE. Interestingly, this affinity is similar to that of a non-specific protease substrate, such as Bovine Serum Albumin (BSA), for both ISP1 and SbtE (Table 5.T.1, Fig. 5.4b). The diminished affinity of the propeptide for ISP1 is a likely consequence of changes in specific surface residues that constitute the propeptide:protease interface. Hence together, the above experimental data suggests that ISP1 and SbtE display catalytic properties of classical serine proteases, but differ in their interaction with propeptides.

A comparison of the folding pathways of ISP1 and SbtE:

To examine reversibility of folding, denatured ISP1 was rapidly diluted into renaturation buffer (Yabuta, Takagi et al. 2001), and recovered activity was used as a direct indicator of folding. Fig. 5.4c establishes that ISP1 folding is reversible and its efficiency is ~70%, probably due to autoproteolysis. Under identical conditions the protease-domain of SbtE folds exceedingly slow, but folding is rapidly catalyzed by the propeptide added *in trans* (Fig. 5.4c). Such reversible folding-unfolding equilibrium provides sufficient evidence for a thermodynamically-stable native state (Anfinsen 1973; Baker and Agard 1994; Jaswal, Sohl et al. 2002). Hence, unlike ESPs whose folding is under kinetic

control (Shinde and Inouye 2000; Bryan 2002), ISPs adopt thermodynamically-stable folds independent of propeptides.

To investigate the thermodynamic versus kinetic folding of ISPs and ESPs, subsequent studies were performed using active-site variants that lack complications of proteolysis (Shinde and Inouye 1995). S_{246}A-ISP1 adopts a secondary structure super-imposable with ISP1, but lacks proteolytic activity (data not shown). Folding kinetics of S_{246}A-ISP1 and Pro-S_{221}C-SbtE (Shinde and Inouye 1995; Jain, Shinde et al. 1998), were monitored using stopped-flow fluorescence (Fig. 5.4d). Both best fit a double exponential rate equation, with rate constants of $20.23s^{-1}$and $0.012s^{-1}$ for S_{246}A-ISP1, and $0.021s^{-1}$ and $0.0016s^{-1}$ for Pro-S_{221}C-SbtE, respectively. Comparison of the two phases observed during folding suggests that the rapid phase of S_{246}A-ISP1 is ~10^3-fold faster than that of Pro-S_{221}C-SbtE. Observed slow phases may be consequences of proline-isomerisation (Eder, Rheinnecker et al. 1993). Unassisted folding of SbtE is extremely slow (Eder, Rheinnecker et al. 1993) ($t_{½}$~1500 yrs) and its kinetics cannot be accurately measured using fluorescence. Fersht and co-workers (Eder, Rheinnecker et al. 1993) have estimated that the propeptide enhances the folding rate of the protease-domain by ~10^6-fold. Hence, while they have conserved structures, ISP1 appears to fold ~10^9-times faster than propeptide-independent folding of SbtE.

Rigorous analysis of folding kinetics reveals a burst phase within the dead-time of the stopped-flow instrument. The occurrence of burst phase kinetics in several folding models suggests the presence of transient intermediates which are established counterparts of equilibrium intermediates (Fujiwara, Arai et al.

1999). Since analysis of intermediates can provide insights into folding pathways, we examined equilibrium folding-unfolding transitions of S_{246}A-ISP1 and Pro-S_{221}C-SbtE (Fig. 5.4e). Equilibrium folding-unfolding of S_{246}A-ISP1 is completely reversible with two distinct transitions, and is represented by at least three-states (Fujiwara, Arai et al. 1999), the native (N), intermediate (I) and unfolded states (U) [See Materials and Methods]. Parameters that best represent folding-unfolding of S_{246}A-ISP1 are listed in Table 5.T.1. Pro-S_{221}C-SbtE undergoes autoprocessing to give a thermodynamically-stable stoichiometric complex, whose structure has been solved (Jain, Shinde et al. 1998). This complex also shows a three-state equilibrium transition (Fig. 5.4e). Studies suggest that the I-state observed in S_{246}A-ISP1 is on-pathway (data not shown) as shown in case of Pro-SbtBPN (Eder, Rheinnecker et al. 1993). This I-state is more structured (70% of native ellipticity at 222 nm) than that observed in Pro-S_{221}C-SbtE (~33% of native ellipticity), and deconvolution (Fig. 5.4e) establishes that the I-state of S_{246}A-ISP1 is maximally populated in 0.8 M GdnHCl when compared with the I-state of Pro-S_{221}C-SbtE (2.5M GdnHCl). At high denaturant concentrations, deviations from linearity (roll-overs) are observed in the Chevron plots (Jackson and Fersht 1991) ($\ln[k_{obs}]$ versus denaturant) of both S_{246}A-ISP1 and Pro-S_{221}C-SbtE (Fig. 5.4f). This is typically ascribed to folding intermediates (Raschke and Marqusee 1997) and is consistent with our equilibrium studies. Extent of solvent exposure between the states involved in a transition is defined by m-values, and ratio of m_{IU}/m_{NU} obtained from Chevron plots (Raschke and Marqusee 1997; Takei, Chu et al. 2000) (Fig. 5.4f) suggest a more compact intermediate state for ISP1 (0.7716)

compared to Pro-SbtE (0.536), probably due to presence of the propeptide. Nonetheless, ISP1 folds to its thermodynamically-stable native state (ΔG_{NU}=5.3.kcal.mol^{-1}) through a compact transient intermediate, while Pro-SbtE folds and autoprocesses (Jain, Shinde et al. 1998) into a stable stoichiometric complex (ΔG_{NU}=9.9.kcal.mol^{-1}) via an expanded intermediate (Table 5.T.1). Degradation of the inhibitory propeptide from this complex results in a kinetically-trapped protease-domain (Eder, Rheinnecker et al. 1993; Yabuta, Takagi et al. 2001) whose unfolding does not show a stable intermediate (Table 5.T.1). Although, propeptide degradation lowers the thermodynamic-stability of SbtE (Table 5.T.1; $\Delta G_{release}$=10.4.kcal.mol^{-1}), it ensures proteolytic-stability (Yabuta, Takagi et al. 2001; Bryan 2002; Jaswal, Sohl et al. 2002) by increasing the unfolding energy-barrier (Fig. 5.5a). It is important to note that the observed differences in folding pathways of ISP1 and SbtE are not mere consequences of their cellular location, because, the intracellular or extra-cellular expression of SbtE does not yield active protease in the absence of its cognate propeptide (Zhu, Ohta et al. 1989). Hence, ISP1 and SbtE have similar sequences, structures and catalytic activities but fold through significantly different pathways. This suggests that the inability to spontaneously fold may not be a property of that particular scaffold, but may be a consequence of fine sequence details.

Evolutionary Analysis of Bacillus ESPs and ISPs:

To understand how two conserved proteins have evolved to fold differently, and to establish the constraints driving such evolution, we performed

a phylogeny-based statistical analysis on *Bacillus* ESPs and ISPs, to test for positive-selection along the ISP-subfamily (Fig. 5.6a). Ratio of the rate of non-synonymous (*dn*) to synonymous substitution (*ds*), ω, is an unequivocal indicator of evolution indicating neutral (ω=1), purifying (ω<1) or positive diversifying (ω>1) selection (Yang 1997; Swanson, Yang et al. 2001). Our dataset was analyzed using various codon-based Site and Branch-Site Models in PAML (Yang 1997) that estimate ω using maximum likelihood (see Materials and Methods). The major advantage with this analysis is that it allows for variable selective pressures along a gene, by assuming different classes of sites with varying ω ratios (Yang 1997; Swanson, Yang et al. 2001). Site-specific models allow ω to vary among sites, but not among branches. Branch-Site models are more powerful in detecting positive-selection as they allow variations in ω between both, various sites along the sequence, and different lineages along the tree (Bielawski and Yang 2003). Briefly, the site-models were M0 [one ω for all sites], M1 [allows ω=1 and ω=0], M2 [ω=1, ω=0 and ω>1], M3 [discrete model, with 3-site classes for ω estimated from data], M7 [beta-distribution for ω within (0,1)] and M8 [beta-distribution within (0,1), and independent ω estimated from data]. The branch-site models used were Model-A and Model-B [four different classes of ω; see Materials and Methods]. A Likelihood Ratio Test (LRT) is used to examine for positive selection, that is for presence of sites with ω > 1. This is achieved by comparing a null model that does not account for ω > 1 with a more general model that allows for ω > 1. For example M0, which allows one ω for all sites was compared with M3, which is a discrete model with 3-site classes for ω estimated from data. Positive-selection

is indicated when $\omega > 1$ AND LRT is significant (see Materials and Methods). When LRTs suggest positive-selection, the posterior probabilities for each site to have a specific ω-class can be calculated using Bayes-theorem. Our dataset was analyzed as described in the Materials and Methods and the parameters estimated are listed in Table 5.T.2.

Log-likelihoods for both branch-site models show significantly better fits ($p<0.05$) when compared with their null models [Model-A with M1; Model-B with M3] and show positive- selection along the ISP-clade (Fig. 5.6a, Table 5.T.2). Both models show that although most sites are neutral or conserved during evolution ($\omega \leq 1$), ~ 22% of sites exhibit strong positive-selection ($\omega > 1$) within ISPs. Hence, while subtilases are well-conserved, it appears that there has been significant adaptive evolution preceding ISPs, to allow propeptide-independent protein folding. We next examined posterior probabilities using an empirical Bayes-approach (Yang 1997; Swanson, Yang et al. 2001) to identify positively selected sites (Table 5.T.2). It is important to note that 75% sites with probability>0.95 are confined to the surface of ISP1 (Fig. 5.6b; Table 5.T.2). Due to their industrial importance ESPs have been extensively analyzed using mutagenesis and directed evolution, and sites that directly affect folding, activity and stability have been identified (Bryan 2000). Interestingly, several positively selected sites identified in our evolutionary analysis overlap with sites determined through such studies (Table 5.T.2 and Table 5.T.3) and are further consistent with earlier findings from our laboratory (Inouye, Fu et al. 2001). For example, residue 188 in ESPs (corresponding to site-213 in ISP1) has been established to interact with residue 17 (earlier referred to as position –60) in the

propeptide-domain during folding (Inouye, Fu et al. 2001). This site appears to have positively evolved within propeptide independent ISPs. Hence, while kinetics (Fig. 5.4) establishes that the structurally conserved proteases SbtE and ISP1 (Fig. 5.1,5.2 and 5.3), can fold through different pathways, our evolutionary analysis (Fig. 5.6; Table 5.T.2) suggests that this selection of pathways may be mediated through positive-selection of specific residues.

IV) DISCUSSION:

Subtilases constitute a ubiquitous super-family of serine proteases, with most members being synthesized with N-terminal propeptide extensions (Siezen and Leunissen 1997; Shinde and Inouye 2000). X-ray crystallography and homology modeling concur that catalytic domains of subtilases adopt very similar three-dimensional structures (Siezen and Leunissen 1997). The contemporary view is that structurally conserved subtilases should fold into their native states through similar transition states and folding pathways (Shakhnovich, Abkevich et al. 1996; Nishimura C, Prytulla S et al. 2000; Plaxco, Larson et al. 2000; Plaxco, Simons et al. 2000; Bucciantini, Giannoni et al. 2002; Calloni, Taddei et al. 2003; Friel, Capaldi et al. 2003). This is consistent with extensive work on prokaryotic subtilisins and eukaryotic subtilisin/kexin like pro-hormone convertases that emphasize the importance and necessity of propeptide-domains in facilitating correct protein folding (Shinde and Inouye 2000; Bryan 2002). Much of our mechanistic understanding of propeptide-mediated protein folding has emerged from analysis of the bacterial ESPs (Bryan, Alexander et al. 1992; Strausberg, Alexander et al. 1993; Gallagher, Gilliland et al. 1995; Ruvinov, Wang et al. 1997; Siezen and Leunissen 1997; Ruan, Hoskins et al. 1998; Ruan, Hoskins et al. 1999; Shinde, Fu et al. 1999; Fu, Inouye et al. 2000; Buevich, Shinde et al. 2001; Inouye, Fu et al. 2001; Yabuta, Takagi et al. 2001; Bryan 2002; Yabuta, Subbian et al. 2002; Yabuta, Subbian et al. 2003), which are extracellular proteases. Detailed *in vitro* and *in vivo* studies have demonstrated that propeptides can direct folding of ESPs into kinetically stable states and are not mere facilitators of correct localization

(Ikemura, Takagi et al. 1987; Zhu, Ohta et al. 1989; Shinde and Inouye 2000). Propeptides remain tightly associated with their cognate ESPs and have to be proteolytically degraded to release enzymatic activity (Li, Hu et al. 1995). Cleavage and degradation of the propeptide at the completion of maturation enables the decoupling of folding and unfolding pathways (Yabuta, Takagi et al. 2001) and stabilizes the native state by increasing the energy of the transition state for the unfolding reaction (Fig. 5.5a). Furthermore, proteolytic cleavage also facilitates formation of the high affinity calcium-binding site A, which contributes significantly to the increased energy of the unfolding transition state in ESPs(Siezen and Leunissen 1997).

A comprehensive database search for SbtE-like bacterial proteases helped identify two distinct subfamilies (ISPs and ESPs) within subtilases that exhibit significant sequence, structure, and functional conservation (Siezen and Leunissen 1997; Shinde and Inouye 2000). While ESPs and ISPs display high sequence identity (Fig. 5.1a) the latter represents an intracellular subfamily that lacks the classical subtilisin propeptide signature (Shinde and Inouye 1993), that is essential for correct folding of ESPs. The lack of the propeptide signature suggests that ISPs might fold independent of propeptide-domains. The present work represents the first description of the propeptide independent folding pathway of a Ca^{2+}-dependent subtilase and a detailed analysis of evolutionary constraints that affect the choice between thermodynamic and kinetic folding and stability.

Fig. 5.1a establishes that ESPs and ISPs have approximately 50% sequence identity. A striking feature of this alignment is that specific residues that are

conserved within one family are changed and, display significantly different amino acid properties within the other. For example, the loop connecting β5 with β6 has sites $D_{182}S$, $E_{189}T$, $E_{198}S$ that change from polar (ESPs) to acidic (ISPs). Similarly, $R_{136}E$, $Y_{137}W$, $E_{158}A$, $K_{164}D$ are substitutions that are localized within helices α4 and α5, which contribute interactions to the propeptide:subtilisin interface in ESPs. Specifically, mutations at site 112 in SbtE (corresponding to 136 in ISPs;$R_{136}E$ substitution) have been established to significantly lower the binding affinity of SbtE for the propeptide (Fu, Inouye et al. 2000). Thus while the two subfamilies have conserved structures, there are specific sites that appear to be selected for during their evolution. Interestingly such sites map mostly to the solvent accessible surface of the protein (Figures 2c and 6b). Furthermore, the level of sequence identity/conservation within the hydrophobic core is significant, and quantitation suggests that the average strength of the hydrophobic core, which is the major driving force in protein folding reactions (Agashe, Shastry et al. 1995; Ptitsyn 1995; Dill and Chan 1997), is similar in ESPs and ISPs (Table 5.T.1 and Fig. 5.2b).

In the current manuscript, ISP1 was extensively studied and compared with our ESP folding model, SbtE. The CD spectroscopic analysis demonstrates that ISP1 and SbtE adopt very similar secondary structures (Fig. 5.3a) and exhibit similar thermo-stability (Table 5.T.1). The inability to spontaneously fold has been attributed to high stability of ESPs bestowed by the presence of a high-affinity calcium-binding site (Siezen and Leunissen 1997). A calcium deletion variant of ESPs can fold independent of its propeptide (Bryan, Alexander et al. 1992; Bryan 2002), albeit very slowly (Ruvinov, Wang et al. 1997; Ruan,

Hoskins et al. 1998; Ruan, Hoskins et al. 1999) and its folding is catalyzed over 10^6-fold by the addition of the propeptide (Ruan, Hoskins et al. 1999). The sequence alignment of ESPs and ISPs also indicates significant conservation among residues that constitute the Ca^{2+} binding site. Consistent with this conservation (Fig. 5.1a), ISP1 displays a strong calcium dependent proteolytic stability. However, the affinity of ISP1 for calcium is approximately 3-fold less than SbtE, which may be due to an insertion prior to β2 in ISP1 (Fig. 5.1a). It is important to note that the presence of this Ca^{2+} binding site does not preclude ISP1 from spontaneous folding/unfolding (Fig. 5.4c & e) independent of any propeptide-domain.

The propeptide of subtilases are potent inhibitors of protease activity and exhibit significant cross reactivity. Therefore, the propeptide of SbtE can bind tightly with other ESPs such as subtilisin Carlsberg, which exhibits ~60% sequence identity. However, the SbtE propeptide is not a good inhibitor of ISP1. On the other hand, SSI, a potent subtilisin inhibitor binds to both SbtE and ISP1 with similar affinity (Fig. 5.4b; Table 5.T.1). The weak affinity of ISP1 for the propeptide, which is similar to the affinity of a non specific substrate BSA (Fig. 5.4b), can be attributed to distinct differences in the mode of interaction (Jain, Shinde et al. 1998; Shinde and Inouye 2000). The propeptide binds with the protease domain in a 'side-on' orientation through an interface mediated by helices α4 and α5 (Jain, Shinde et al. 1998). On the other hand, SSI binds to the active-site of subtilisin in a 'top-on' orientation (Takeuchi, Satow et al. 1991) through interactions that are completely different from that of the propeptide. Interestingly, the sequence alignment and the modeled structure establish that

these helices (α–4 and α-5) in ISPs exhibit differences in key residues (Fig. 5.2a & c) that are essential for high-affinity binding to the propeptide (Yabuta, Subbian et al. 2003). However, these helices share >90% identity within ESPs (Fig. 5.1a) and as a consequence, the propeptide can strongly inhibit other ESPs.

Comparative studies demonstrate that folding of ISP1 is propeptide-independent and thermodynamically reversible with ΔG_{UF} of -5.3 kcal/mol, while folding/unfolding of the protease domain of SbtE is thermodynamically reversible only in the presence of its cognate propeptide with ΔG_{UF} of ~-10.0kcal/mol. Interestingly, equilibrium folding/unfolding reactions of ISP1 and Pro-SbtE indicate the presence of distinct protein folding intermediates that are maximally populated in 0.8M and 2.5M GdnHCl, respectively (Fig. 5.4e). The dependence of the apparent folding rates on denaturant (Fig. 5.4f) shows a roll-over in the unfolding limbs for both proteins and supports the presence of stable intermediates in the folding pathway. Furthermore, the extent of solvent exposure obtained from the chevron plots (Fig. 5.4f) suggests that the folding intermediate of ISP1 is more compact, in contrast with the diffused state in SbtE. The presence of partially structured intermediates is in contrast with the hypothesis that broad specificity proteases such as subtilisin would select against folding via stable intermediates, because they would be excellent substrates for autolysis (Bryan 2002). While such partially structured states are commonly found in several protein models and are important intermediates in the overall folding pathway, their precise role in folding of ISP1 and Pro-SbtE remain unknown.

Our studies establish that, SbtE, an ESP that is secreted in harsh, extracellular, protease-rich environments folds to its native state with the help of a propeptide domain (Shinde, Liu et al. 1997; Shinde, Fu et al. 1999; Fu, Inouye et al. 2000; Shinde and Inouye 2000; Inouye, Fu et al. 2001; Marie-Claire, Yabuta et al. 2001; Yabuta, Takagi et al. 2001; Yabuta, Subbian et al. 2002; Yabuta, Subbian et al. 2003). The presence of the propeptide domain enables it to reach a kinetically trapped native state that has a high unfolding energy barrier and thus offers higher proteolytic stability. On the other hand, the intracellular homologue, ISP1, with similar sequence, topology, hydrophobic core, and catalytic activity can fold to its native state even without the help of a propeptide domain. Further, it reaches a thermodynamically stable native state through a more compact folding intermediate. While the contemporary view is that homologous proteins fold through similar pathways and folding transition states (Shakhnovich, Abkevich et al. 1996; Nishimura C, Prytulla S et al. 2000; Plaxco, Larson et al. 2000; Plaxco, Simons et al. 2000; Bucciantini, Giannoni et al. 2002; Calloni, Taddei et al. 2003; Friel, Capaldi et al. 2003), our results demonstrate that ISP1 (propeptide-independent) and SbtE (propeptide-dependent) reach similar native states through different folding pathways. Evolutionary analysis of the constraints underlying the evolution of such closely related sub-families shows that the ISPs have indeed positively diverged from the ESPs.

It is important to note that the observed differences in folding pathways of ISP1 and SbtE are not mere consequences of their cellular location, because, the intracellular or extra-cellular expression of SbtE does not yield active

protease in the absence of its cognate propeptide (Ikemura, Takagi et al. 1987; Shinde and Inouye 2000). While, the kinetic-barriers apparently enhance stability of ESPs in harsh extracellular-environments (Bryan 2002; Jaswal, Sohl et al. 2002), similar barriers leading to high-stability in ISPs may impede intracellular protein-turnover. This is consistent with our finding that intracellular expression of ESPs (Aqualysin 1) is detrimental to cell-growth due to extensive proteolysis (unpublished data). This suggests that biological requirements, and not just specific conformations dictate selection of folding pathways. A crucial question is how two homologous subtilase-subfamilies have evolved to select different folding pathways? Several studies demonstrate the importance of topology, contact order, hydrophobicity, and stability in driving protein folding pathways and kinetics (Shakhnovich, Abkevich et al. 1996; Nishimura C, Prytulla S et al. 2000; Plaxco, Simons et al. 2000; Gunasekaran, Eyles et al. 2001). Interestingly, the above properties appear conserved within ISPs and ESPs, and differences within their sequences map specifically onto the protein surface. Furthermore, evolutionary analysis shows that adaptive evolution, which precedes ISPs, is predominantly through positive-selection of specific surface residues. These differences may allow ISPs to fold in a propeptide-independent, thermodynamically-driven pathway.

Another important question is how changes in surface residues compensate for the loss of propeptides within ISPs? Both ESPs and ISPs are highly charged proteins but with extremely different isoelectric points (Table 5.T.1). Moreover, this charge is localized within propeptides of ESPs, while it is distributed over the protease-domain of ISPs. As a consequence protease domains of ISPs are

more polar when compared with ESPs. This difference may help to enhance conformational entropy of water around the folding polypeptide and assist the hydrophobic core in driving spontaneous folding of ISPs. When folding of ESPs is carried out in the absence of the propeptide, the protease domain adopts a partially structured molten-globule intermediate with solvent exposed hydrophobic surfaces (Shinde and Inouye 1995). The partially formed hydrophobic core within the protease domain of ESPs may be insufficient to drive folding without the solvation assistance bestowed by the surface charge. This leads to stabilization of the molten-globule intermediate that is prone to aggregation. This concept is consistent with our findings that charge per se, and not its polarity is critical for folding of subtilases (Marie-Claire, Yabuta et al. 2001; Yabuta, Subbian et al. 2003). This result to our knowledge is the first demonstration that surface residues and their charge may influence the selection of folding pathways and kinetics, and fine sequence details may dictate selection between kinetic and thermodynamically driven protein folding.

V) MATERIALS AND METHODS:

Sequence analysis and homology modeling:

Protein sequences were aligned using ClustalW and analyzed in GeneDoc. Homology modelling of ISP1 was performed using Swiss Model, a program that predicts structures reliably (RMSD<2Å) for sequences with 50-60% identity (Schwede, Kopp et al. 2003). Structure was validated using ANOLEA, ProSAII, VADAR, and WHATCHECK programs and displayed using Pymol. Sites within the conserved hydrophobic-core have solvent accessibility <15% AND hydrophobicity core index (HCI)>60%. Where

$$HCI = \left[\sum_{i=1}^{n} AA \times Hyd\right] \times \frac{N_{Hyd}}{N}, \qquad \text{Equation (5.1)}$$

AA=hydrophobic residues; *Hyd*=hydrophobicity of *AA*; N_{Hyd} = Number of *AA* at site *i* ; N=total number of residues at *i*.

Protein Expression and Purification:

ISP1 was amplified from the genomic DNA of *Bacillus subtilis (*American Type Culture Collection, Manassas, Virginia, ATCC No 23857D) using PCR and cloned into pET11a. Site-directed mutagenesis was done as described (Yabuta, Subbian et al. 2003). Nucleotide sequences were confirmed by sequencing both strands. Proteins were expressed in *Escherichia coli* BL21(DE3)] and purified from the soluble fraction using techniques as described (Yabuta, Takagi et al. 2001; Yabuta, Subbian et al. 2002; Yabuta, Subbian et al. 2003).

Structural Characterization:

CD-spectra were measured on an AVIV-215 spectrophotometer using a 1 mm path length quartz cuvette (Yabuta, Takagi et al. 2001). Proteins samples (7-15μM) were taken in renaturation-buffer [50mM Tris-HCl, pH 8.5, 0.5M $(NH_4)_2SO_4$, 1mM $CaCl_2$]. Thermostability was monitored through loss in ellipticity at 222nm. For intrinsic tryptophan fluorescence (PTI-fluorometer), samples were excited at 295nm and emission spectra between 300-400nm were obtained. Folded and unfolded samples were in renaturation buffer and in 6M Guanidine Hydrochloride (GdnHCl), pH 4.8, respectively. Fluorescence change upon folding/unfolding was used to monitor kinetics. Stopped-flow kinetics was measured on a Hi-Tech SF61-DX2 fluorometer using $\lambda_{excitation}$ of 295nm and 310nm band-pass filter. For the Chevron analysis, folded and unfolded S_{246}A-ISP1 and Pro-SbtE (1-4 μM) were diluted (1:10 ratio) into renaturation buffer containing different concentrations of GdnHCl (0-5M). Data were fitted to single and double exponential equations (GraphPad-Prism) to determine folding/unfolding rate constants (Yabuta, Takagi et al. 2001). Chevron plots best fit a three-state equation (Raschke and Marqusee 1997; Fujiwara, Arai et al. 1999; Takei, Chu et al. 2000). All experiments were performed at 23°C and are averages of 3-6 independent repeats.

Activity, Inhibition, and Calcium dependent stability:

Activity of ISP1 and SbtE were measured at 405nm using the chromogenic substrate, N-succ-AAPF-*pNA (Yabuta, Takagi et al. 2001; Yabuta, Subbian et al. 2003)*. Calcium dependence of ISP1 and SbtE was measured in

Chelex-treated buffer [50mM Tris-HCl, pH 8.5, 0.5M $(NH_4)_2SO_4$], to which various concentrations of calcium were added (Bryan, Alexander et al. 1992). Aliquots were removed as a function of time and protease activity was measured to obtain decay-rates (k_{decay}). Slopes obtained from plots of k_{decay} versus [Ca^{2+}/Protein] provide the relative calcium-dependent stability of ISP1 and SbtE. Inhibition constants (K_I) were measured by adding proteases [ISP1 (7.6nM) or SbtE (58.3nM)] to renaturation buffer containing Inhibitors [propeptide (0.05-35μM), SSI (0.05-12.5nM) or BSA (1.18-20μM)] and 0.5mM Nsuc-AAPF-*pNA (Yabuta, Subbian et al. 2003)*. K_I was obtained from EC_{50} plots of normalized rates of *p-nitroanilide* released versus inhibitor-concentration (GraphPad-Prism).

Kinetics and Equilibrium Folding/Unfolding:

Pro-SbtE and SbtE folding were monitored as described (Fu, Inouye et al. 2000; Yabuta, Takagi et al. 2001). Purified ISP1 was completely denatured (loss of structure and activity), after precipitation [55% $(NH_4)_2SO_4$] and re-suspension in 6M GdnHCl, pH 4.8. Denatured ISP1 (22μM) was rapidly diluted (50-fold) into renaturation buffer and recovered activity was assayed at various times after folding initiation. Figure4c represents maximum activity recovered after complete folding. Equilibrium folding-unfolding was monitored (CD at 222nm) to obtain free-energy differences and m-values for each transition as described (Eder, Rheinnecker et al. 1993; Eder, Rheinnecker et al. 1993; Raschke and Marqusee 1997; Fujiwara, Arai et al. 1999; Takei, Chu et al.

2000).The transitions are represented by at least three-states (Fujiwara, Arai et al. 1999), the native (N), intermediate (I) and unfolded states (U) as follows:

$$N \underset{}{\overset{K_{NI}}{\rightleftharpoons}} I \overset{K_{IU}}{\rightleftharpoons} U \qquad \text{Equation (5.2)}$$

The observed ellipticity (A_{obs} *(c)*) at any concentration of the denaturant is given by the sum of the contributions from the three states as (Mizuguchi, Arai et al. 1998; Takei, Chu et al. 2000):

$$A_{obs}(c) = \left[\frac{A_N + A_I \exp[-(\Delta G_{NI}^{H_2O} - m_{NI}c)/RT] + A_U \exp[-(\Delta G_{NU}^{H_2O} - m_{NU}c)/RT]}{1 + \exp[-(\Delta G_{NI}^{H_2O} - m_{NI}c)/RT] + \exp[-(\Delta G_{NU}^{H_2O} - m_{NU}c)/RT]} \right]$$

Equation (5.3)

Where f_N *(c)*, f_I *(c)*and f_U *(c)* are the fractions of the three states at a GdnHCl concentration of *c* ($f_N + f_I + f_U = 1$), and A_N, A_I and A_U are the ellipticity values of the pure N, I and U states, respectively. The f_N, f_I and f_U terms are related to the equilibrium constants, K_{NI} and K_{NU}, of the unfolding transitions from N to I and from N to U, respectively. $\Delta G_{NI}^{H_2O}$ and $\Delta G_{NU}^{H_2O}$ are the ΔG_{NI} and ΔG_{NU} at 0M GdnHCl and $m_{NI}c$ & $m_{NU}c$ represent the dependence of the respective free energy changes on *c*. The data were fitted using Prism Graphpad.

Statistical Analysis of Phylogeny:

Protein sequences of ESPs and ISPs (SwissProt, Hobacgen and Merops databases) were aligned using ClustalW and used to obtain nucleotide alignments. The DNAML-program (PHYLIP-package) was used to estimate tree-topology (Fig. 5.6a). Positive-selection (ω>1) was tested using several codon based likelihood models (codeml-program) in PAML(Yang 1997).

Likelihood-ratio tests (LRT) were used to establish best-fits by comparing general-models with their null-models (Swanson, Yang et al. 2001; Yang 2002; Bielawski and Yang 2003). The null distribution of the LRT–statistic ($2\Delta l$ where Δl equals difference between log-likelihood scores of the two models) can be approximated using the χ^2 distribution, with the degree of freedom being the difference in number of free parameters between the two models. Briefly, the site-models were M0 [one ω for all sites], M1 [allows ω=1 and ω=0], M2 [ω=1, ω=0 and ω>1], M3 [discrete model with 3-site classes for ω estimated from data], M7 [beta-distribution for ω within (0,1)] and M8 [beta-distribution within (0,1), and independent ω estimated from data]. The branch-site models used were Model-A and Model-B (Swanson, Yang et al. 2001; Yang 2002; Bielawski and Yang 2003). Branch site models test for adaptive evolution along a specific lineage, which is termed as the foreground branch. The basic branch-site model allows for four site classes. The first two classes (ω_0 and ω_1) are uniform over the entire phylogeny. The other two classes allow for some sites with ω_0 and ω_1 to change to positive selection (ω_2). In Model A, ω_0 and ω_1 are fixed, while in Model B they are estimated as free parameters (Yang 2002; Bielawski and Yang 2003).

The foreground branch for our analysis is highlighted in blue (Fig. 5.6a). When LRTs suggest positive-selection, the posterior probabilities for each site to have a specific ω-class can be calculated using Bayes-theorem (Swanson, Yang et al. 2001; Yang 2002; Bielawski and Yang 2003). Our evolutionary analysis of ISPs and ESPs (Branch-Site Models A and B) from *Bacillus* suggests that adaptive evolution followed gene duplication leading to ISPs. The

Bayes approach identified ~22% sites that have $\omega > 1$ and hence have positively evolved in the ISP clade. Interestingly most of these sites have been identified as being critical for folding/activity/stability of SbtE through independent mutagenesis and directed evolution studies (Bryan 2000). Table 5.T.3 lists sites that show a strong positive selection in ISPs together with the corresponding sites in ESPs and the role of that site in folding/stability/activity of ESPs as identified through independent studies. Consistent with studies on other protein families, trees generated using different programs (DNAPARS and DNADIST) do not affect the analysis (data not shown)

VI) FIGURES

Fig. 5.1. Sequence conservation between ISPs (blue) and ESPs (Red). **(a)** Multiple Sequence Alignment (MSA) between representative ISPs and ESPs. SUBT_BACSU (SbtE) and ISP1_BACSU (ISP1) are our models whose secondary structures are depicted above and below the MSA, respectively. Active-sites (yellow spheres), calcium-binding domains (green bars) and sequence conservation (black-highlight>75% identity; grey-highlight>70% similarity) are depicted. Specific regions are conserved (>65% identity) only within ESPs (red-type) or ISPs (blue-type). Amino acid composition and isoelectric points are shown N-terminus to each sequence (*ACID:*-D,E; *BASIC:*-K,R,H; *HYDRO:*-A,V,L,F,G,I,Y,W,P,M,C; *POLAR:*-S,N,Q,T)

Fig. 5.1

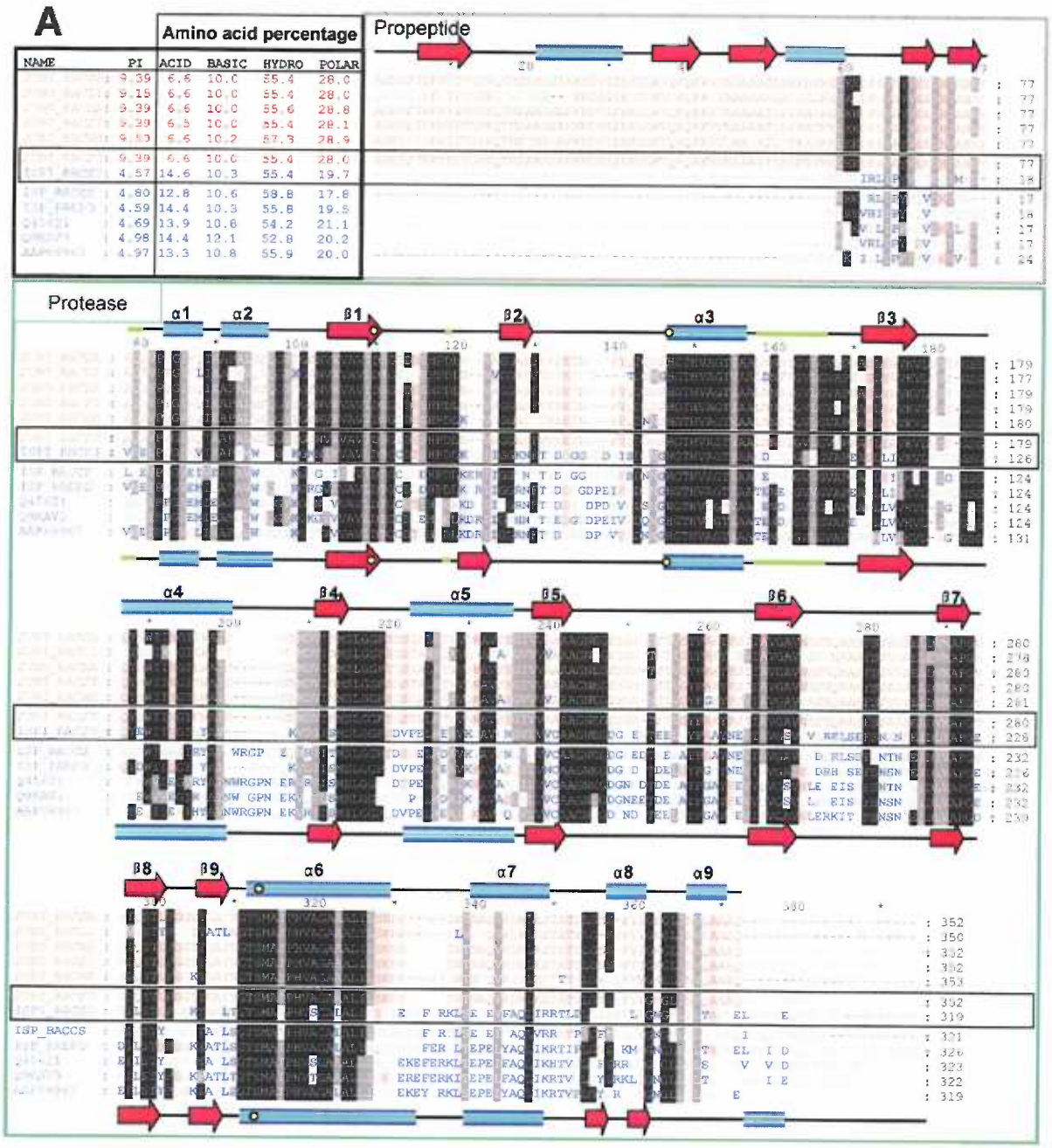

A
Amino acid percentage
NAME PI ACID BASIC HYDRO POLAR
9.39 6.6 10.0 55.4 28.0
9.15 6.6 10.0 55.4 28.0
9.39 6.6 10.0 55.6 28.8
9.39 6.5 10.0 55.4 28.1
9.39 6.6 10.0 55.4 28.0
4.57 14.6 10.3 55.4 19.7
4.80 12.8 10.6 58.8 17.8
4.59 14.4 10.3 55.8 19.5
4.69 13.9 10.6 54.2 21.1
4.98 14.4 12.1 52.8 20.2
4.97 13.3 10.8 55.9 20.0
Propeptide
Protease
α1 α2 β1 β2 α3 β3
α4 β4 α5 β5 β6 β7
β8 β9 α6 α7 α8 α9
ISP_BACCS

Fig. 5.2. Structural conservation between ISPs (blue) and ESPs (Red). **(a)** Ribbon models of SbtE (PDB:1SCJ) and ISP1 (homology model, validated as described in Materials and Methods). a4 and a5 are critical for propeptide interaction with SbtE **(b)** Conserved Hydrophobic-core (yellow space-filled) within ISPs and ESPs. **(c)** Sequence conservation from MSA mapped onto structures. Residues conserved within both ESPs and ISPs are colored grey, while red and blue depicts residues that are conserved only within ESPs or ISPs, respectively. Residues that are conserved within ESPs, but without selection-constraints in ISPs are shown in Cyan

Fig. 5.2

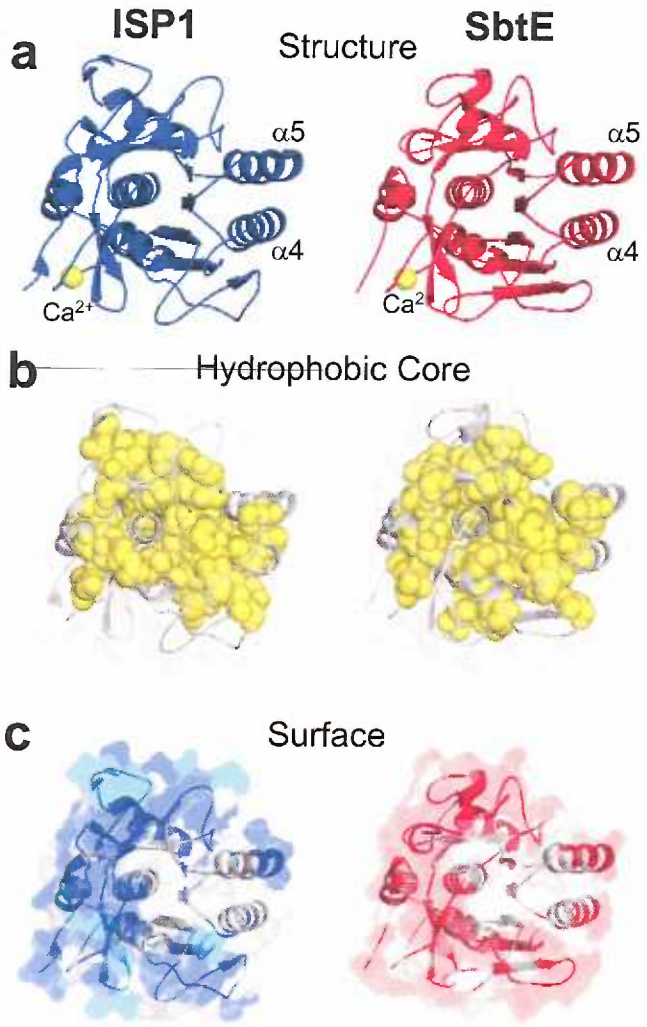

a
ISP1
Structure
SbtE
α5
α4
Ca2+
α5
α4
Ca2
b
Hydrophobic Core
c
Surface

Fig. 5.3. Secondary and tertiary structure comparison between ISP1 (blue) and SbtE (Red). (a) Secondary structure comparison using CD. Table 5.T.1 contains structural contributions. **(b)** Intrinsic tryptophan fluorescence. Native state (solid lines) SbtE displays higher fluorescence relative to ISP1 due an additional tryptophan. Unfolded proteins (dashed-lines) lose fluorescence intensity.

Fig. 5.3

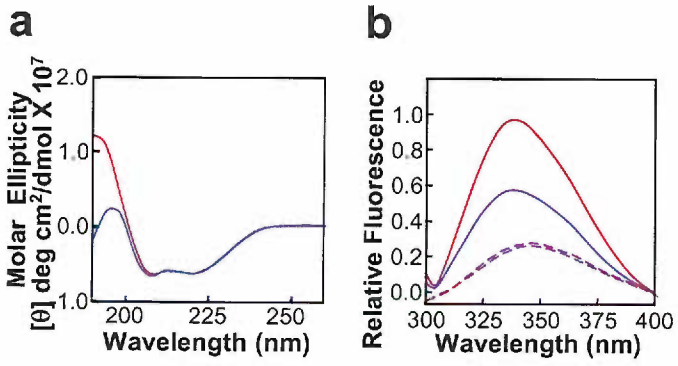

a
Molar Ellipticity
[θ] deg cm²/dmol X 10⁷
2.0
1.0
0.0
1.0
200
225
250
Wavelength (nm)
b
Relative Fluorescence
1.0
0.8
0.6
0.4
0.2
0.0
300
325
350
375
400
Wavelength (nm)

Fig. 5.4. Characterization of ISP1 (black) and SbtE (grey). (a) Activity and stability (measured as k_{decay}) as a function of Ca^{2+}. ISP1 activity decays 3-fold faster as measured from relative slopes. **(b)** EC_{50} plots for ISP1 and SbtE inhibition by SSI (open-circles), propeptide (solid-circles) and BSA (open diamonds). **(c)** Activity recovered upon refolding. Denatured ISP1 (U) recovers activity on refolding (F:black-stripes) and is unaffected by the propeptide (FP:solid-black). Unfolded SbtE (U) when refolded is inactive (F) unless folded in presence of its propeptide (FP:solid-grey). **(d)** Refolding kinetics of S_{246}A-ISP1 (Upper panel) and Pro-S_{221}C-SbtE (Lower panel). Residuals for best-fit curves are below each graph. **(e)** Equilibrium folding-unfolding of S_{246}A-ISP1 and Pro-S_{221}C-SbtE monitored using CD at 222nm. Resolved fractions of native (solid-line), unfolded (dashed-line) and intermediate (dotted-line) states are shown. *(f)* Chevron Plots fitted to a three-state model(Raschke and Marqusee 1997; Takei, Chu et al. 2000) yield m_{IU}=3.37; m_{NU}=4.50 for S_{246}A-ISP1 (black) and m_{IU}=1.78; m_{NU}= 3.33 for Pro-S_{221}C-SbtE (grey).

Fig. 5.4

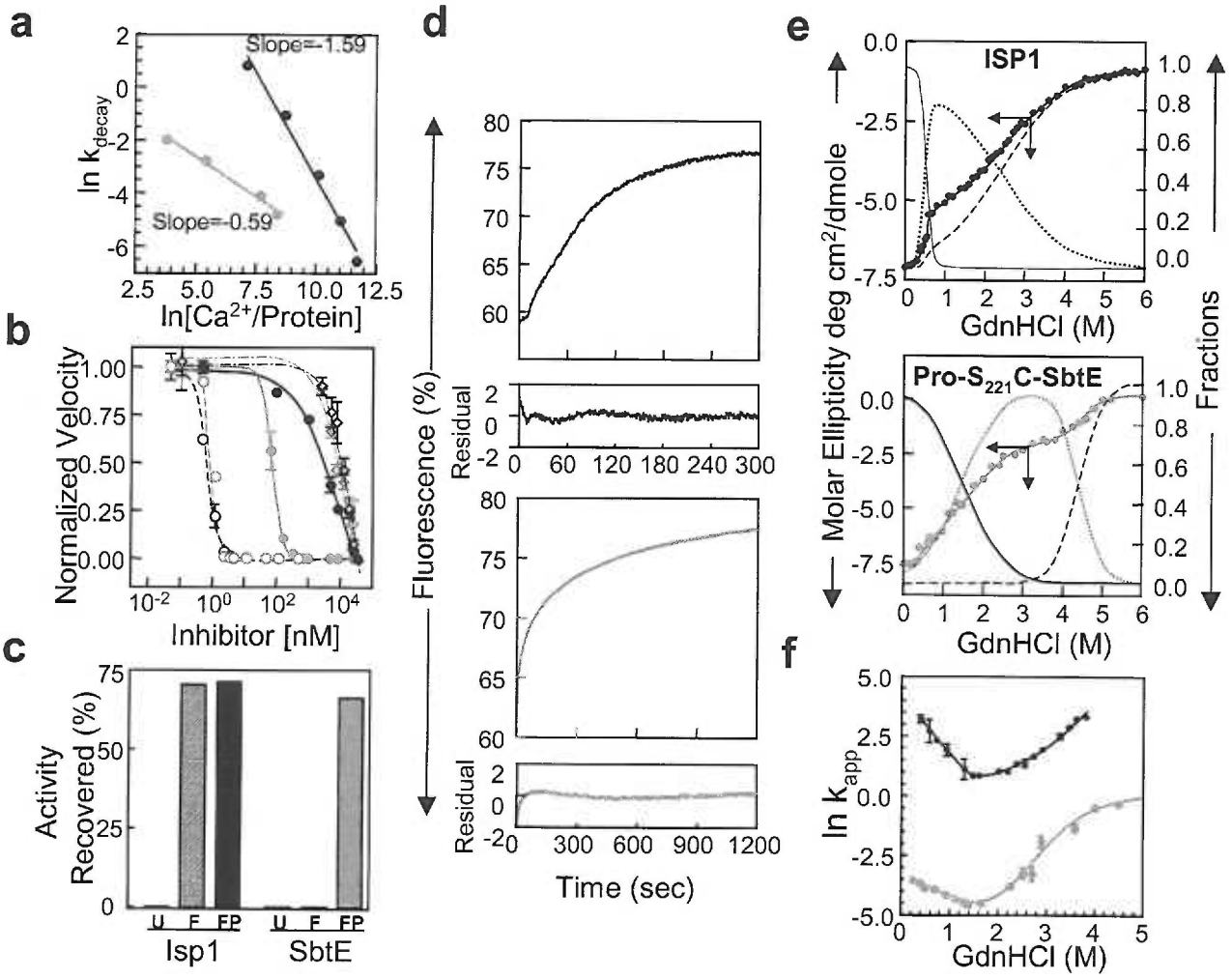

a
Slope=-1.59
Slope=-0.59
ln k_decay
ln[Ca2+/Protein]
b
Normalized Velocity
Inhibitor [nM]
c
Activity Recovered (%)
U F FP
U F FP
Isp1
SbtE
d
Fluorescence (%)
Residual
Time (sec)
e
ISP1
Molar Ellipticity deg cm2/dmole
GdnHCl (M)
Pro-S221C-SbtE
Fractions
f
ln kapp
GdnHCl (M)

Fig. 5.5. Representative folding energy landscape of ISPs and ESPs (a) Representative free energy landscape of kinetic and thermodynamically driven protein folding. Denatured ISP1 (U) folds through an intermediate (I) to its thermodynamically stable native state (N). Pro-SbtE folds into a thermodynamically stable complex (N + P). Propeptide degradation from this complex kinetically traps the native state (N) by increasing the unfolding energy-barrier and by uncoupling folding (Solid grey) and unfolding (broken grey) pathways.

Fig. 5.5

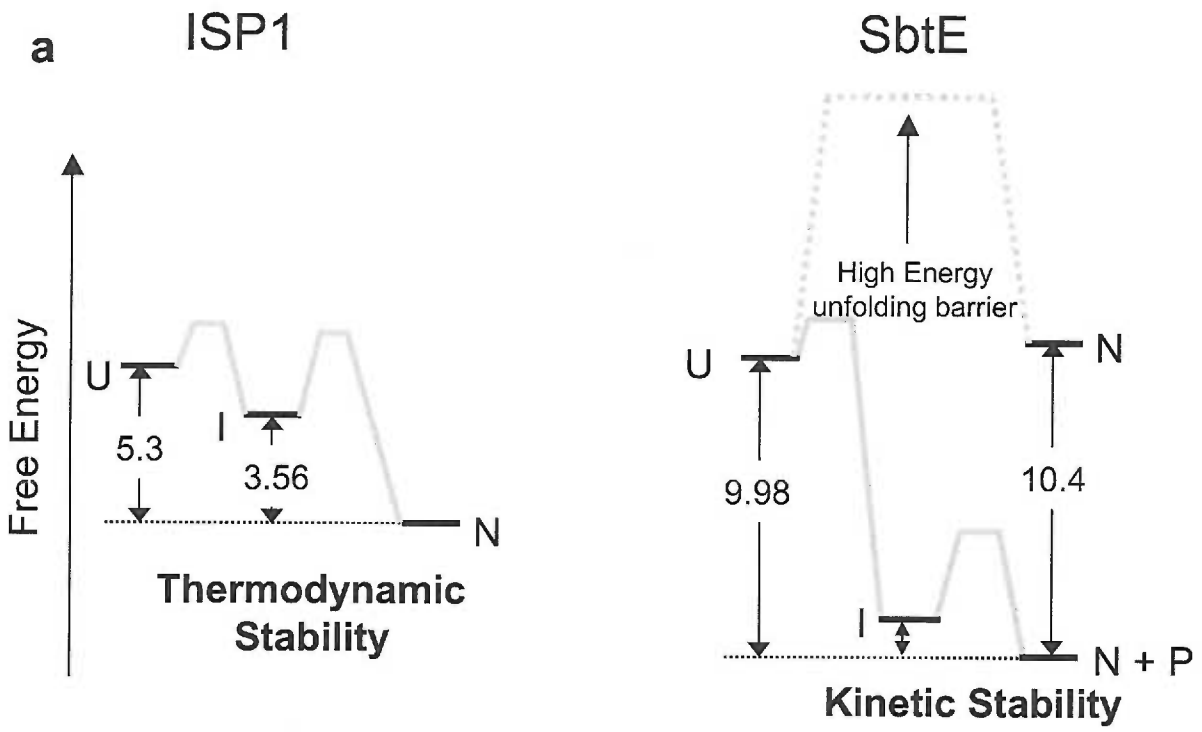
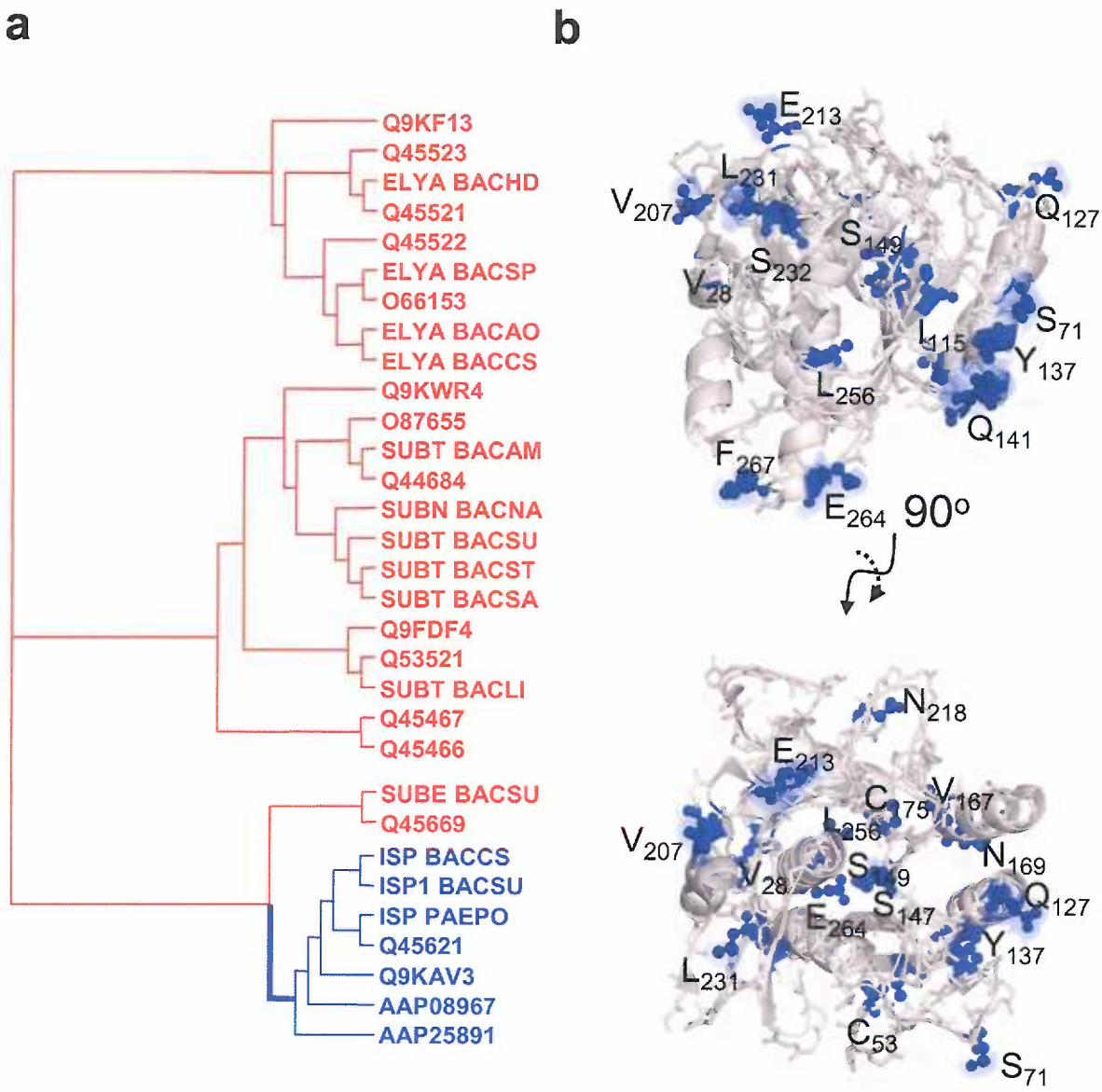

a
ISP1
Free Energy
U
I
N
5.3
3.56
Thermodynamic Stability
SbtE
High Energy unfolding barrier
U
N
9.98
10.4
I
N + P
Kinetic Stability

Fig. 5.6. Adaptive Evolution of ISPs **(a)** Phylogeny of ISPs (blue) and ESPs (Red) from *Bacillus*. Branch lengths are not to scale. Lineage leading to ISPs (thick blue-line) was tested for positive-selection. **(b)** Positively selected residues (posterior probability ≥ 0.95) are mapped onto ISP1 structure.

Fig. 5.6


a
Q9KF13
Q45523
ELYA BACHD
Q45521
Q45522
ELYA BACSP
O66153
ELYA BACAO
ELYA BACCS
Q9KWR4
O87655
SUBT BACAM
Q44684
SUBN BACNA
SUBT BACSU
SUBT BACST
SUBT BACSA
Q9FDF4
Q53521
SUBT BACLI
Q45467
Q45466
SUBE BACSU
Q45669
ISP BACCS
ISP1 BACSU
ISP PAEPO
Q45621
Q9KAV3
AAP08967
AAP25891
b
E213
L231
V207
Q127
S149
S232
V28
S71
L115
Y137
L256
Q141
F267
E264
90°
N218
E213
C175
V167
L256
V207
N169
S149
V28
S147
Q127
E264
Y137
L231
C53
S71

VII) TABLES

Table 5.T.1 Comparison between ISP1, SbtE and Pro-SbtE from *Bacillus subtilis* as models for ISPs and ESPs

Property	ISP1	SbtE	Pro: SbtE
Sequence			
Identity with ISP1 (%)	100	50	50
Total Length (residues)	319	275	352
Propeptide (residues)	17	0	77
Isoelectric point (pI)	4.7	6.3	8.5
Charged residues (%)	23.10	9.40	14.77
Contact order (CO)	27.36	28.41	NA
Hydrophobic core (HCI)	5.36	6.04	NA
Activity			
Optimum pH	8.5	8.5	NA[a]
K_I nM, for SSI	0.17	0.26	NA[a]
K_I nM, for ProS[b]	1410	18.50	NA[a]
K_I nM, for BSA	4700	2140	NA[a]
Specific Activity$_{N\text{-}suc\ AAPFpNA}$ (Units/ng)	2698	210	NA[a]
Structure			
Helix (%)	21.0	25.5	22.4
Sheets (%)	26.4	22.8	21.7
Beta turn (%)	17.3	17.4	17.4
RandomCoil (%)	35.3	34.3	38.8
Tm (°C)	58.7	58.3	54.5
Equilibrium Unfolding			
ΔG_{NI} (kcal/mol)	3.56	NA	1.62
m_{NI} (kcal/mol M)	7.14	NA	1.11
ΔG_{NU} (kcal/mol)	5.3	2.433[c]	9.98
m_{NU} (kcal/mol M)	7.85	0.73[c]	3.02

[a]Cleaved complex (pro:SbtE) cannot bind inhibitors

[b]$\Delta G_{release}$ for pro from SbtE = $-RTlnK_I$ = 10.4 kcal/mol

[c]Values are illustrative (SbtE unfolding is irreversible)

Table 5.T.2. Phylogeny based statistical analysis of Bacillus ESPs and ISPs using maximum likelihood models

Models	Likelihood	ω(dn/ds)	Positive-selection and Sites Selected[a]
SITE MODELS			
M0	-16311.68	0.038	Not Allowed
M1	-17475.09	P0=0.112, p1=0.888, ω0=0.000, ω1=1.000	Not Allowed
M2	-17469.63	P0=0.113, p1=0.843, p2=0.046 ω0=0.000, ω1=1.000,ω2=2.172	Yes, 169N, **231L**[b]
M3	-15873.44	P0=0.543, p1=0.457, p2=0.000, ω0=0.013, ω1=0.080, ω2=6.845	No
M7	-15815.78	P=0.737, q=14.409	Not Allowed
M8 (ncat=5)	-15817.58	P0= 1.000 p=0.709 q=13.607 (p1= 0.000) ω1=1.383	No
BRANCH- SITE MODELS			
A	-17422.28	P0=0.093, p1=0.664, p2=0.030, p3=0.214, ω0=0.000, ω1=1.000, ω2=999.00	Yes, 14 sites **28V**, 53C, **127Q**[b], **137Y**[b], 147S[b], **149S**[b], 167V,175C, **207V**[b], **213E**[b], **218N, 232S, 264E**[b], **267F**
B	-15856.14	p0=0.347, p1= 0.418, p2= 0.107, p3= 0.129,ω0=0.082 ω1=0.0128, ω2= 24.447	Yes, 12 sites 53C, **71S**[b], 115I[b], **137Y**[b], **141Q, 149S**[b], 175C,**218N, 232S,** 256L, **264E**[b], **267F**

[a]Only those sites with posterior probability>0.95 for positive-selection are shown. Residues in bold-type face are those on protein surface and constitute 75% of positively selected residues. There is > 55% overlap between residues identified from branch-site Model-A and B.

[b]Mutations at these sites have been shown to affect folding/stability/activity of subtilisins through independent directed evolutionary studies (Kobayashi and Inouye 1992; Bryan 2000; Inouye, Fu et al. 2001).

Table 5.T.3. Postively Selected Sites in ISPs, identified through phylogeny analysis, and their functional roles from independent mutagenesis studies in ESPs(Bryan 2000)

ISP1	SbtE	Implicated role in Subtilisins
28V	10Q	Not Determined
53C	35I	Not Determined
71S	51V	Stability
115I	92A	Activity, Stability
127Q	103Q	Activity, Affects Flexibility
137Y	113W	Folding, Stability
141Q	117N	Not Determined
147S	123N	Activity, Affects Flexibility
149S	125S	Activity
167V	143V	Not Determined
175C	151A	Not Determined
207V	182S	Activity, Stability
213E	188S	Folding,stability, Thermostability
218N	193G	Not Determined
231L	206Q	Stability
232S	207S	Not Determined
256L	231A	Not Determined
264E	239P	Activity, Stability
267F	242T	Not Determined

CHAPTER 6

CONCLUSIONS

Protein folding represents the crucial step in the transformation of genetic information into biological function. The complex, but specific, assembly of linear chains of amino acids into their precise orientations enables polypeptides to perform myriad of cellular functions. In this dissertation, using SbtE, we have addressed fundamental questions that underlie polypeptide folding. These studies offer interesting insights into the role of IMCs in protease folding and activation, and highlight how nature optimizes not only sequence and structure but also polypeptide folding for protein function.

I) How do IMCs facilitate efficient protease maturation?

Maturation of Pro-SbtE involves folding, autoprocessing, and subsequent autocatalytic activation to release active protease. Activation is the rate-limiting step to Pro-SbtE maturation and requires the release of the IMC from the inhibited complex, and the subsequent degradation of the IMC by the free protease. We demonstrate that protease activation is an inherently random process (Subbian, Yabuta et al., 2005). Such randomness in biological networks allows for effective modulation of the underlying processes, as the

intrinsic noise provides room for control. This is substantiated by the ability to regulate the *in-vitro* activation of Pro-SbtE by altering the structural dynamics of the IMC, through modulation of solvent environment. Intrinsically unstructured or unstable IMC domains have been established in other systems as well (Marie-Claire, Yabuta et al., 2001; Wiederanders, Kaulmann et al., 2003). For example, the IMC of thermostable aqualysin is structured at 4°C, but at 80°C, where the enzyme has optimum activity, the IMC is intrinsically unstructured (Marie-Claire, Yabuta et al., 2001). Furthermore, IMC-dependent proteases get activated in specific cellular environments such as the activation of pro-furin in TGN and that of pro-cathepsins in the lysosome (Thomas, 2002). Hence, we propose that apart from their role in folding, IMCs can function as regulators of enzyme activation. Structural instability within the IMCs may be selected to enable them to regulate protease activation through their own folding/unfolding (Subbian, Yabuta et al., 2005).

The intricate relation between structure and protease regulation is further substantiated by the folding energy landscape of Pro-SbtE, and through the effect of a structured IMC on this landscape (See Chapter 3). While the initial structural acquisition in Pro-SbtE is thermodynamically favored, the protease subsequently transitions to a thermodynamically unstable but kinetically-trapped native state, through the release of its IMC. This initial structural acquisition is driven by a hydrophobic collapse and occurs through a molten globule-like intermediate state. In the presence of a structured IMC, the molten-globule intermediate is stabilized and results in faster and more energetically-favored folding. Whether stable intermediates assist proteins in their folding or are

kinetic traps that inhibit the folding process has been a point of debate in protein folding. The faster folding kinetics of Pro-SbtE in the presence of a more stable intermediate suggests that the intermediate assists in polypeptide folding. However, under these conditions, the activation of the protease is limited by a higher kinetic barrier and is slow and delayed. Consequently, an intrinsically unstructured IMC, and hence a less-stable intermediate, may have been selected at least in the case of Pro-SbtE. This structural instability within the IMC may be the optimal for its transition from an unfolded polypeptide, to a structured inhibitory conformation, and eventually to an unstructured substrate. Thus, the IMC is optimized not for its individual chaperone, inhibitory or regulatory functions but to ensure the synergy of the overall maturation process.

II) How do differences in charge, structure, and primary sequence within IMC relate to their chaperone and inhibitory functions?

The selection of sequence and structural features within IMCs for their optimal function was further highlighted through the design of the *de novo* chaperone, ProD (Yabuta, Subbian et al., 2003). Using a protein design approach, we studied the minimum requirements in the propeptide for its chaperone function. The presence of the conserved motifs, N1 and N2, and the hydrophobic core within ProD that has only 16% sequence identity to WT-IMC, suggests that the motifs and conserved core are critical for its chaperoning function. Another essential requirement for efficient chaperoning may be the presence of charged residues. As stated earlier, the IMC of SbtE is an

extremely charge polypeptide. Hence, it is interesting to note that although ProD has a pI that is 5 pH units apart from that of WT IMC, it has 35% charged residues in its sequence. This also suggests that not the nature of the charge, but the charge per se, maybe critical for IMC function. Based on this observation, it is tempting to speculate that the charged residues within IMCs play a direct role in lowering kinetic barriers to folding of proteases.

Studies on ProD also highlight the delicate interplay between IMC structure and inhibition. Unlike the IMC of SbtE, ProD is structured even in its isolated form. Interestingly, ProD is also a ten-fold better inhibitor of SbtE. As seen from the crystal structure, the side-on orientation of the IMC is critical for its role in inhibition. The presence of a structured IMC, such as ProD, enhances this interaction resulting in a tighter complex. This suggests that the $\alpha\beta$ scaffold maybe critical for the inhibitory function of IMCs. While our design approach highlights the minimum requirements necessary for an IMC, it should be noted that ProD is only a weak chaperone. Thus, optimizing the chaperoning efficiency of ProD through directed evolution studies can further enhance our understanding of IMC domains.

III) Why have specific proteins evolved dedicated IMCs to mediated folding pathways?

While the above studies highlight that both the sequence and structure of Pro-SbtE is optimized for a kinetically determined folding pathway, the studies on ISP1 highlight that such folding pathways may be specifically selected (Subbian, Yabuta et al., 2004). Analysis of the folding pathway of ISP1

establishes that, unlike its extracellular homologue (SbtE), ISP1 folds to a thermodynamically stable native state independent of an IMC. Hence, despite having structurally similar native states and ~50% primary sequence identity, the folding of isolated SbtE results in a trapped molten globule-like intermediate state unlike ISP1, which adopts a functional state. This establishes that similar native states can be reached through different folding pathways and with distinct folding requirements. Of importance is the functional advantage that the distinct folding pathways confer on ISP1 and SbtE. While the kinetic stability of the SbtE native state ensures its longevity in the harsh extracellular environments, the thermodynamic stability of ISP1 maybe optimal for intracellular turn-over. Hence, it appears that the folding pathways of proteins can be selected for optimization of function. Interestingly, an evolutionary analysis demonstrates that the difference in pathways may have evolved through the positive selection of specific residues that mostly map to the protein surface.

The occurrence of distinct folding pathways in closely related homologues is contrary to accepted ideas about folding of polypeptide sequences (Shakhnovich, Abkevich et al., 1996; Plaxco, Simons et al., 2000; Larson, Ruczinski et al., 2002). The new view of folding establishes that the native state is attained through a stochastic search of the many conformations accessible to the polypeptide, and that the stability of the sub-structures that are formed directs the protein to its native state (Dobson, 2004). Hence, the structure of the native state and the average separation in the sequence between residues that are in contact with each other in the native structure

(intrinsic contact order), are critical factors that determine the folding pathway (Plaxco, Simons et al., 2000). This highlights that topology of the native state dictates protein folding and that proteins with similar topology should fold through similar pathways. However, based on the folding pathways of ISP1 and SbtE, it appears that not just topology, but also the fine sequence details determine folding pathways. A noteworthy point is that ISP1 has a high distribution of charged residues within its mature domain, which is similar to the charged property of IMCs. It is tempting to speculate that the charge within ISP1 substitutes for the absence of a covalently linked chaperone and may enhance folding by increasing the solvent dynamics around the folding polypeptide. Since IMCs function in intracellular environments the dynamics in the protease that may result from the charged residues may not be a significant deterrent to its proteolytic stability. In lieu of recent studies, which suggest that the transition state to folding of proteins is similar to their native state, it would be interesting to establish the similarities and differences in the transition states of SbtE and ISP1. This would help to elucidate if SbtE and ISP1 fold through partially overlapping, or distinctly different, folding pathways.

Thus through the application of biochemical, biophysical and computational analyses to a robust folding model we establish that,

1) IMCs may function as regulators of protease activation, and guide proteases to kinetically trapped native states through a synergistic maturation process

2) The propensity for charged residues within IMCs and the structural instability have been selected for their function as chaperones, inhibitors, and regulators, of enzyme activation.

3) Nature has optimized not just the protein sequence and structure, but also the folding pathway for function.

REFERENCES:

Agashe, V. R., M. C. Shastry, et al. (1995). "Initial hydrophobic collapse in the folding of barstar." Nature **377**(6551): 754-7.

Alexander, P. A., B. Ruan, et al. (2001). "Cation-dependent stability of subtilisin." Biochemistry **40**(35): 10634-9.

Alexander, P. A., B. Ruan, et al. (2001). "Stabilizing mutations and calcium-dependent stability of subtilisin." Biochemistry **40**(35): 10640-4.

Ali, V., K. Prakash, et al. (1999). "8-anilino-1-naphthalene sulfonic acid (ANS) induces folding of acid unfolded cytochrome c to molten globule state as a result of electrostatic interactions." Biochemistry **38**(41): 13635-42.

Alm, E., A. V. Morozov, et al. (2002). "Simple physical models connect theory and experiment in protein folding kinetics." J Mol Biol **322**(2): 463-76.

Almog, O., T. Gallagher, et al. (1998). "Crystal structure of calcium-independent subtilisin BPN' with restored thermal stability folded without the prodomain." Proteins **31**(1): 21-32.

Almog, O., A. Gonzalez, et al. (2003). "The 0.93A crystal structure of sphericase: a calcium-loaded serine protease from Bacillus sphaericus." J Mol Biol **332**(5): 1071-82.

Altschul, S. F., T. L. Madden, et al. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res **25**(17): 3389-402.

Ammerer, G., C. P. Hunter, et al. (1986). "PEP4 gene of Saccharomyces cerevisiae encodes proteinase A, a vacuolar enzyme required for processing of vacuolar precursors." Mol Cell Biol **6**(7): 2490-9.

Anderson, E. D., S. S. Molloy, et al. (2002). "The ordered and compartment-specfific autoproteolytic removal of the furin intramolecular chaperone is required for enzyme activation." J Biol Chem **277**(15): 12879-90.

Anfinsen, C. B. (1973). "Principles that govern the folding of protein chains." Science **181**(96): 223-30.

Ardelt, W. and M. Laskowski, Jr. (1985). "Turkey ovomucoid third domain inhibits eight different serine proteinases of varied specificity on the same ...Leu18-Glu19 ... reactive site." Biochemistry **24**(20): 5313-20.

Baker, D. (1998). "Metastable states and folding free energy barriers." Nat Struct Biol **5**(12): 1021-4.

Baker, D. and D. A. Agard (1994). "Kinetics versus thermodynamics in protein folding." Biochemistry **33**(24): 7505-9.

Baker, D. and A. Sali (2001). "Protein structure prediction and structural genomics." Science **294**(5540): 93-6.

Baker, D., A. K. Shiau, et al. (1993). "The role of pro regions in protein folding." Curr Opin Cell Biol **5**(6): 966-70.

Baker, D., J. L. Sohl, et al. (1992). "A protein-folding reaction under kinetic control." Nature **356**(6366): 263-5.

Baldwin, T. O., M. M. Ziegler, et al. (1993). "Contribution of folding steps involving the individual subunits of bacterial luciferase to the assembly of the active heterodimeric enzyme." J Biol Chem **268**(15): 10766-72.

Bassi, D. E., R. Lopez De Cicco, et al. (2001). "Furin inhibition results in absent or decreased invasiveness and tumorigenicity of human cancer cells." Proc Natl Acad Sci U S A **98**(18): 10326-31.

Berger, D. and T. Altmann (2000). "A subtilisin-like serine protease involved in the regulation of stomatal density and distribution in Arabidopsis thaliana." Genes Dev **14**(9): 1119-31.

Berkenpas, M. B., D. A. Lawrence, et al. (1995). "Molecular evolution of plasminogen activator inhibitor-1 functional stability." Embo J **14**(13): 2969-77.

Bielawski, J. P. and Z. Yang (2003). "Maximum likelihood methods for detecting adaptive evolution after gene duplication." J Struct Funct Genomics **3**(1-4): 201-12.

Bott, R., M. Ultsch, et al. (1988). "The three-dimensional structure of Bacillus amyloliquefaciens subtilisin at 1.8 A and an analysis of the structural consequences of peroxide inactivation." J Biol Chem **263**(16): 7895-906.

Bowers, P. M., L. E. Schaufler, et al. (1999). "A folding transition and novel zinc finger accessory domain in the transcription factor ADR1." Nat Struct Biol **6**(5): 478-85.

Bryan, P., P. Alexander, et al. (1992). "Energetics of folding subtilisin BPN'." Biochemistry **31**(21): 4937-45.

Bryan, P., L. Wang, et al. (1995). "Catalysis of a protein folding reaction: mechanistic implications of the 2.0 A structure of the subtilisin-prodomain complex." Biochemistry **34**(32): 10310-8.

Bryan, P. N. (2000). "Protein engineering of subtilisin." Biochim Biophys Acta **1543**(2): 203-222.

Bryan, P. N. (2002). "Prodomains and protein folding catalysis." Chem Rev **102**(12): 4805-16.

Bucciantini, M., E. Giannoni, et al. (2002). "Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases." Nature **416**(6880): 507-11.

Buevich, A. V., U. P. Shinde, et al. (2001). "Backbone dynamics of the natively unfolded pro-peptide of subtilisin by heteronuclear NMR relaxation studies." J Biomol NMR **20**(3): 233-49.

Calloni, G., N. Taddei, et al. (2003). "Comparison of the folding processes of distantly related proteins. Importance of hydrophobic content in folding." J Mol Biol **330**(3): 577-91.

Carmona, E., E. Dufour, et al. (1996). "Potency and selectivity of the cathepsin L propeptide as an inhibitor of cysteine proteases." Biochemistry **35**(25): 8149-57.

Carr, C. M., C. Chaudhry, et al. (1997). "Influenza hemagglutinin is spring-loaded by a metastable native conformation." Proc Natl Acad Sci U S A **94**(26): 14306-13.

Chen, B. L., W. A. Baase, et al. (1989). "Low-temperature unfolding of a mutant of phage T4 lysozyme. 2. Kinetic investigations." Biochemistry **28**(2): 691-9.

Cheng, S. W., H. M. Hu, et al. (1995). "Production and characterization of keratinase of a feather-degrading Bacillus licheniformis PWD-1." Biosci Biotechnol Biochem **59**(12): 2239-43.

Chu, N. M., Y. Chao, et al. (1995). "The 2 A crystal structure of subtilisin E with PMSF inhibitor." Protein Eng **8**(3): 211-5.

Comellas-Bigler, M., K. Maskos, et al. (2004). "1.2 A crystal structure of the serine carboxyl proteinase pro-kumamolisin; structure of an intact pro-subtilase." Structure (Camb) **12**(7): 1313-23.

Cui, Y., R. Hackenmiller, et al. (2001). "The activity and signaling range of mature BMP-4 is regulated by sequential cleavage at two sites within the prodomain of the precursor." Genes Dev **15**(21): 2797-802.

Cunningham, E. L. and D. A. Agard (2003). "Interdependent folding of the N- and C-terminal domains defines the cooperative folding of alpha-lytic protease." Biochemistry **42**(45): 13212-9.

Cunningham, E. L. and D. A. Agard (2004). "Disabling the folding catalyst is the last critical step in alpha-lytic protease folding." Protein Sci **13**(2): 325-31.

Cunningham, E. L., S. S. Jaswal, et al. (1999). "Kinetic stability as a mechanism for protease longevity." Proc Natl Acad Sci U S A **96**(20): 11008-14.

Cunningham, E. L., T. Mau, et al. (2002). "The pro region N-terminal domain provides specific interactions required for catalysis of alpha-lytic protease folding." Biochemistry **41**(28): 8860-7.

De Bie, I., D. Savaria, et al. (1995). "Processing specificity and biosynthesis of the Drosophila melanogaster convertases dfurin1, dfurin1-CRR, dfurin1-X, and dfurin2." J Biol Chem **270**(3): 1020-8.

Dhar, P., T. C. Meng, et al. (2004). "Cellware--a multi-algorithmic software for computational systems biology." Bioinformatics **20**(8): 1319-21.

Dill, K. A. and H. S. Chan (1997). "From Levinthal to pathways to funnels." Nat Struct Biol **4**(1): 10-9.

Dobson, C. M. (2003). "Protein folding and misfolding." Nature **426**(6968): 884-90.

Dobson, C. M. (2004). "Principles of protein folding, misfolding and aggregation." Semin Cell Dev Biol **15**(1): 3-16.

Dyson, H. J. and P. E. Wright (2002). "Coupling of folding and binding for unstructured proteins." Curr Opin Struct Biol **12**(1): 54-60.

Dyson, H. J. and P. E. Wright (2002). "Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance." Adv Protein Chem **62**: 311-40.

Eder, J. and A. R. Fersht (1995). "Pro-sequence-assisted protein folding." Mol Microbiol **16**(4): 609-14.

Eder, J., M. Rheinnecker, et al. (1993). "Folding of subtilisin BPN': characterization of a folding intermediate." Biochemistry **32**(1): 18-26.

Eder, J., M. Rheinnecker, et al. (1993). "Folding of subtilisin BPN': role of the pro-sequence." J Mol Biol **233**(2): 293-304.

Elowitz, M. B., A. J. Levine, et al. (2002). "Stochastic gene expression in a single cell." Science **297**(5584): 1183-6.

Engelhard, M. and P. A. Evans (1995). "Kinetics of interaction of partially folded proteins with a hydrophobic dye: evidence that molten globule character is maximal in early folding intermediates." Protein Sci **4**(8): 1553-62.

Fei, H., M. J. Luo, et al. (2001). "The Inhibitory Activities of Recombinant Eglin C Mutants on Kexin and Furin, Using Site-directed Mutagenesis and Molecular Modeling." Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai) **33**(6): 591-599.

Ferrari, E., S. M. Howard, et al. (1986). "Effect of stage 0 sporulation mutations on subtilisin expression." J Bacteriol **166**(1): 173-9.

Fersht, A. (1985). Enzyme Structure and Mechanism, W.H. Freeman.

Friel, C. T., A. P. Capaldi, et al. (2003). "Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: similarities and differences in the folding of homologous proteins." J Mol Biol **326**(1): 293-305.

Fu, X., M. Inouye, et al. (2000). "Folding pathway mediated by an intramolecular chaperone. The inhibitory and chaperone functions of the subtilisin propeptide are not obligatorily linked." J Biol Chem **275**(22): 16871-8.

Fuhrmann, C. N., B. A. Kelch, et al. (2004). "The 0.83A Resolution Crystal Structure of alpha-Lytic Protease Reveals the Detailed Structure of the Active Site and Identifies a Source of Conformational Strain." J Mol Biol **338**(5): 999-1013.

Fujiwara, K., M. Arai, et al. (1999). "Folding-unfolding equilibrium and kinetics of equine beta-lactoglobulin: equivalence between the equilibrium molten globule state and a burst-phase folding intermediate." Biochemistry **38**(14): 4455-63.

Gallagher, T., G. Gilliland, et al. (1995). "The prosegment-subtilisin BPN' complex: crystal structure of a specific 'foldase'." Structure **3**(9): 907-14.

GCG Genetics Computer Group. Madison, WI.

Gray, T. E., J. Eder, et al. (1993). "Refolding of barnase mutants and pro-barnase in the presence and absence of GroEL." Embo J **12**(11): 4145-50.

Greenfield, N. and G. D. Fasman (1969). "Computed circular dichroism spectra for the evaluation of protein conformation." Biochemistry **8**(10): 4108-16.

Guex, N. and M. C. Peitsch (1997). "SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling." Electrophoresis **18**(15): 2714-23.

Gunasekaran, K., S. J. Eyles, et al. (2001). "Keeping it in the family: folding studies of related proteins." Curr Opin Struct Biol **11**(1): 83-93.

Gupta, R., Q. K. Beg, et al. (2002). "Bacterial alkaline proteases: molecular approaches and industrial applications." Appl Microbiol Biotechnol **59**(1): 15-32.

Henrich, S., A. Cameron, et al. (2003). "The crystal structure of the proprotein processing proteinase furin explains its stringent specificity." Nat Struct Biol **10**(7): 520-6.

Henrich, S., I. Lindberg, et al. (2005). "Proprotein convertase models based on the crystal structures of furin and kexin: explanation of their specificity." J Mol Biol **345**(2): 211-27.

Holyoak, T., M. A. Wilson, et al. (2003). "2.4 A resolution crystal structure of the prototypical hormone-processing protease Kex2 in complex with an Ala-Lys-Arg boronic acid inhibitor." Biochemistry **42**(22): 6709-18.

Hong, Y. C., H. H. Kong, et al. (2000). "Isolation and characterization of a cDNA encoding a subtilisin-like serine proteinase (ahSUB) from Acanthamoeba healyi." Mol Biochem Parasitol **111**(2): 441-6.

Hou, W. S., D. Bromme, et al. (1999). "Characterization of novel cathepsin K mutations in the pro and mature polypeptide regions causing pycnodysostosis." J Clin Invest **103**(5): 731-8.

Hu, Z., K. Haghjoo, et al. (1996). "Further evidence for the structure of the subtilisin propeptide and for its interactions with mature subtilisin." J Biol Chem **271**(7): 3375-84.

Huber, A. H. and W. I. Weis (2001). "The structure of the beta-catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by beta-catenin." Cell **105**(3): 391-402.

Huntington, J. A., R. J. Read, et al. (2000). "Structure of a serpin-protease complex shows inhibition by deformation." Nature **407**(6806): 923-6.

Ichida, J. M., L. Krizova, et al. (2001). "Bacterial inoculum enhances keratin degradation and biofilm formation in poultry compost." J Microbiol Methods **47**(2): 199-208.

Ikemura, H. and M. Inouye (1988). "In vitro processing of pro-subtilisin produced in Escherichia coli." J Biol Chem **263**(26): 12959-63.

Ikemura, H., H. Takagi, et al. (1987). "Requirement of pro-sequence for the production of active subtilisin E in Escherichia coli." J Biol Chem **262**(16): 7859-64.

Im, H., E. J. Seo, et al. (1999). "Metastability in the inhibitory mechanism of human alpha1-antitrypsin." J Biol Chem **274**(16): 11072-7.

Im, H., M. S. Woo, et al. (2002). "Interactions causing the kinetic trap in serpin protein folding." J Biol Chem **277**(48): 46347-54.

Inouye, M., X. Fu, et al. (2001). "Substrate-induced activation of a trapped IMC-mediated protein folding intermediate." Nat Struct Biol **8**(4): 321-5.

Jackson, S. E. and A. R. Fersht (1991). "Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition." Biochemistry **30**(43): 10428-35.

Jain, S. C., U. Shinde, et al. (1998). "The crystal structure of an autoprocessed Ser221Cys-subtilisin E-propeptide complex at 2.0 A resolution." J Mol Biol **284**(1): 137-44.

Janzik, I., P. Macheroux, et al. (2000). "LeSBT1, a subtilase from tomato plants. Overexpression in insect cells, purification, and characterization." J Biol Chem **275**(7): 5193-9.

Jaswal, S. S., J. L. Sohl, et al. (2002). "Energetic landscape of alpha-lytic protease optimizes longevity through kinetic stability." Nature **415**(6869): 343-6.

Jean, L., F. Hackett, et al. (2003). "Functional characterization of the propeptide of Plasmodium falciparum subtilisin-like protease-1." J Biol Chem **278**(31): 28572-9.

Jerala, R., E. Zerovnik, et al. (1998). "pH-induced conformational transitions of the propeptide of human cathepsin L. A role for a molten globule state in zymogen activation." J Biol Chem **273**(19): 11498-504.

Julius, D., A. Brake, et al. (1984). "Isolation of the putative structural gene for the lysine-arginine-cleaving endopeptidase required for processing of yeast prepro-alpha-factor." Cell **37**(3): 1075-89.

Kaneda, M. and N. Tominaga (1975). "Isolation and characterization of a proteinase from the sarcocarp of melon fruit." J Biochem (Tokyo) **78**(6): 1287-96.

Khan, A. R. and M. N. James (1998). "Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes." Protein Sci **7**(4): 815-36.

Khatib, A. M., G. Siegfried, et al. (2002). "Proprotein convertases in tumor progression and malignancy: novel targets in cancer therapy." Am J Pathol **160**(6): 1921-35.

Kim, A. S., L. T. Kakalis, et al. (2000). "Autoinhibition and activation mechanisms of the Wiskott-Aldrich syndrome protein." Nature **404**(6774): 151-8.

Kobayashi, T. and M. Inouye (1992). "Functional analysis of the intramolecular chaperone. Mutational hot spots in the subtilisin pro-peptide and a second-site suppressor mutation within the subtilisin molecule." J Mol Biol **226**(4): 931-3.

Kojima, S., T. Minagawa, et al. (1998). "Tertiary structure formation in the propeptide of subtilisin BPN' by successive amino acid replacements and its close relation to function." J Mol Biol **277**(5): 1007-13.

Kristinsson, H. G. and B. A. Rasco (2000). "Fish protein hydrolysates: production, biochemical, and functional properties." Crit Rev Food Sci Nutr **40**(1): 43-81.

Kriwacki, R. W., L. Hengst, et al. (1996). "Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity." Proc Natl Acad Sci U S A **93**(21): 11504-9.

Kumar, C. G. and H. Takagi (1999). "Microbial alkaline proteases: from a bioindustrial viewpoint." Biotechnol Adv **17**(7): 561-94.

Kussie, P. H., S. Gorina, et al. (1996). "Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain." Science **274**(5289): 948-53.

Kuwajima, K. (1989). "The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure." Proteins **6**(2): 87-103.

Larson, S. M., I. Ruczinski, et al. (2002). "Residues participating in the protein folding nucleus do not exhibit preferential evolutionary conservation." J Mol Biol **316**(2): 225-33.

Lawrence, D. A., D. Ginsburg, et al. (1995). "Serpin-protease complexes are trapped as stable acyl-enzyme intermediates." J Biol Chem **270**(43): 25309-12.

Lee, Y. C., H. Koike, et al. (1994). "Requirement of a COOH-terminal pro-sequence for the extracellular secretion of aqualysin I (a thermophilic subtilisin-type protease) in Thermus thermophilus." FEMS Microbiol Lett **120**(1-2): 69-74.

Lee, Y. C., Y. Miyata, et al. (1991). "Involvement of NH2-terminal pro-sequence in the production of active aqualysin I (a thermophilic serine protease) in Escherichia coli." Agric Biol Chem **55**(12): 3027-32.

Lesage, G., A. Prat, et al. (2000). "The Kex2p proregion is essential for the biosynthesis of an active enzyme and requires a C-terminal basic residue for its function." Mol Biol Cell **11**(6): 1947-57.

Li, W. H., Lu,C.C., Wu,C.I (1985). Molecular Evolutionary Genetics, Plenum Press, New York.

Li, W. H., C. I. Wu, et al. (1985). "A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes." Mol Biol Evol **2**(2): 150-74.

Li, Y., Z. Hu, et al. (1995). "Functional analysis of the propeptide of subtilisin E as an intramolecular chaperone for protein folding. Refolding and inhibitory abilities of propeptide mutants." J Biol Chem **270**(42): 25127-32.

Li, Y. and M. Inouye (1996). "The mechanism of autoprocessing of the propeptide of prosubtilisin E: intramolecular or intermolecular event?" J Mol Biol **262**(5): 591-4.

Light, A. and A. M. al-Obeidi (1991). "Further evidence for independent folding of domains in serine proteases." J Biol Chem **266**(12): 7694-8.

Lobenko, A. A., A. A. Burov, et al. (1991). "[Use of immobilized elastoterase for treating acute purulent soft tissue diseases]." Klin Khir(1): 19-21.

Loperena, L., M. D. Ferrari, et al. (1994). "Study of Bacillus sp. culture conditions to promote production of unhairing proteases." Rev Argent Microbiol **26**(3): 105-15.

Louie, K. and P. A. Conrad (1999). "Characterization of a cDNA encoding a subtilisin-like serine protease (NC-p65) of Neospora caninum." Mol Biochem Parasitol **103**(2): 211-23.

Makarov, D. E. and K. W. Plaxco (2003). "The topomer search model: A simple, quantitative theory of two-state protein folding kinetics." Protein Sci **12**(1): 17-26.

Marie-Claire, C., B. P. Roques, et al. (1998). "Intramolecular processing of prothermolysin." J Biol Chem **273**(10): 5697-701.

Marie-Claire, C., Y. Yabuta, et al. (2001). "Folding pathway mediated by an intramolecular chaperone: the structural and functional characterization of the aqualysin I propeptide." J Mol Biol **305**(1): 151-65.

Masui, A., M. Yasuda, et al. (2004). "Enzymatic hydrolysis of gelatin layers on used lith film using thermostable alkaline protease for recovery of silver and PET film." Biotechnol Prog **20**(4): 1267-9.

Matagne, A., M. Jamin, et al. (2000). "Thermal unfolding of an intermediate is associated with non-Arrhenius kinetics in the folding of hen lysozyme." J Mol Biol **297**(1): 193-210.

Maubach, G., K. Schilling, et al. (1997). "The inhibition of cathepsin S by its propeptide--specificity and mechanism of action." Eur J Biochem **250**(3): 745-50.

McIver, K. S., E. Kessler, et al. (2004). "Identification of residues in the Pseudomonas aeruginosa elastase propeptide required for chaperone and secretion activities." Microbiology **150**(Pt 12): 3969-77.

Mechler, B., M. Muller, et al. (1982). "In vivo biosynthesis of the vacuolar proteinases A and B in the yeast Saccharomyces cerevisiae." J Biol Chem **257**(19): 11203-6.

Mei, H. C., Y. C. Liaw, et al. (1998). "Engineering subtilisin YaB: restriction of substrate specificity by the substitution of Gly124 and Gly151 with Ala." Protein Eng **11**(2): 109-17.

Meichtry, J., N. Amrhein, et al. (1999). "Characterization of the subtilase gene family in tomato (Lycopersicon esculentum Mill.)." Plant Mol Biol **39**(4): 749-60.

Meng, T. C., S. Somani, et al. (2004). "Modeling and simulation of biological systems with stochasticity." In Silico Biol **4**(2): 0024.

Mentrup, B., M. Londershausen, et al. (1999). "Isolation and characterization of insect PC2-like prohormone convertase cDNA." Insect Mol Biol **8**(3): 305-10.

Miller, S. A., E. M. Binder, et al. (2001). "A conserved subtilisin-like protein TgSUB1 in microneme organelles of Toxoplasma gondii." J Biol Chem **276**(48): 45341-8.

Miller, S. A., V. Thathy, et al. (2003). "TgSUB2 is a Toxoplasma gondii rhoptry organelle processing proteinase." Mol Microbiol **49**(4): 883-94.

Mirny, L. and E. Shakhnovich (2001). "Evolutionary conservation of the folding nucleus." J Mol Biol **308**(2): 123-9.

Miyata, T. and T. Yasunaga (1980). "Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application." J Mol Evol **16**(1): 23-36.

Mizuguchi, M., M. Arai, et al. (1998). "Equilibrium and kinetics of the folding of equine lysozyme studied by circular dichroism spectroscopy." J Mol Biol **283**(1): 265-77.

Mogridge, J., P. Legault, et al. (1998). "Independent ligand-induced folding of the RNA-binding domain and two functionally distinct antitermination regions in the phage lambda N protein." Mol Cell **1**(2): 265-75.

Morgunova, E., A. Tuuttila, et al. (1999). "Structure of human pro-matrix metalloproteinase-2: activation mechanism revealed." Science **284**(5420): 1667-70.

Muller, L., A. Cameron, et al. (2000). "Processing and sorting of the prohormone convertase 2 propeptide." J Biol Chem **275**(50): 39213-22.

Nebes, V. L. and E. W. Jones (1991). "Activation of the proteinase B precursor of the yeast Saccharomyces cerevisiae by autocatalysis and by an internal sequence." J Biol Chem **266**(34): 22851-7.

Nekliudov, A. D., A. N. Ivankin, et al. (2000). "[Preparation and purification of protein hydrolysates]." Prikl Biokhim Mikrobiol **36**(4): 371-9.

Nekliudov, A. D., A. N. Ivankin, et al. (2000). "[Characteristics and use of protein hydrolysates (review)]." Prikl Biokhim Mikrobiol **36**(5): 525-34.

Nishimura C, Prytulla S, et al. (2000). "Conservation of folding pathways in evolutionarily distant globin sequences." Nat Struct Biol. . **7**(8): 679-686.

Northey, J. G., A. A. Di Nardo, et al. (2002). "Hydrophobic core packing in the SH3 domain folding transition state." Nat Struct Biol **9**(2): 126-30.

Ogino, T., T. Kaji, et al. (1999). "Function of the propeptide region in recombinant expression of active procathepsin L in Escherichia coli." J Biochem (Tokyo) **126**(1): 78-83.

Onuchic, J. N. and P. G. Wolynes (2004). "Theory of protein folding." Curr Opin Struct Biol **14**(1): 70-5.

Ozbudak, E. M., M. Thattai, et al. (2002). "Regulation of noise in the expression of a single gene." Nat Genet **31**(1): 69-73.

Pan, J., Q. Huang, et al. (2004). "Gene cloning and expression of an alkaline serine protease with dehairing function from Bacillus pumilus." Curr Microbiol **49**(3): 165-9.

Pappenberger, G., H. Aygun, et al. (2001). "Nonprolyl cis peptide bonds in unfolded proteins cause complex folding kinetics." Nat Struct Biol **8**(5): 452-8.

Parker, D., M. Rivera, et al. (1999). "Role of secondary structure in discrimination between constitutive and inducible activators." Mol Cell Biol **19**(8): 5601-7.

Pavletich, N. P. (1999). "Mechanisms of cyclin-dependent kinase regulation: structures of Cdks, their cyclin activators, and Cip and INK4 inhibitors." J Mol Biol **287**(5): 821-8.

Perea, A., U. Ugalde, et al. (1993). "Preparation and characterization of whey protein hydrolysates: applications in industrial whey bioconversion processes." Enzyme Microb Technol **15**(5): 418-23.

Philipp, M. and M. L. Bender (1983). "Kinetics of subtilisin and thiolsubtilisin." Mol Cell Biochem **51**(1): 5-32.

Plaxco, K. W., S. Larson, et al. (2000). "Evolutionary conservation in protein folding kinetics." J Mol Biol **298**(2): 303-12.

Plaxco, K. W., K. T. Simons, et al. (2000). "Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics." Biochemistry **39**(37): 11177-83.

Ptitsyn, O. B. (1995). "How the molten globule became." Trends Biochem Sci **20**(9): 376-9.

Radisky, E. S., D. S. King, et al. (2003). "The role of the protein core in the inhibitory power of the classic serine protease inhibitor, chymotrypsin inhibitor 2." Biochemistry **42**(21): 6484-92.

Radisky, E. S., G. Kwan, et al. (2004). "Binding, proteolytic, and crystallographic analyses of mutations at the protease-inhibitor interface of the subtilisin BPN'/chymotrypsin inhibitor 2 complex." Biochemistry **43**(43): 13648-56.

Ramos, C., J. R. Winther, et al. (1994). "Requirement of the propeptide for in vivo formation of active yeast carboxypeptidase Y." J Biol Chem **269**(9): 7006-12.

Raschke, T. M. and S. Marqusee (1997). "The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions." Nat Struct Biol **4**(4): 298-304.

Rawlings, N. D., D. P. Tolle, et al. (2004). "MEROPS: the peptidase database." Nucleic Acids Res **32**(Database issue): D160-4.

Rockwell, N. C. and R. S. Fuller (2001). "Differential utilization of enzyme-substrate interactions for acylation but not deacylation during the catalytic cycle of Kex2 protease." J Biol Chem **276**(42): 38394-9.

Rockwell, N. C. and J. W. Thorner (2004). "The kindest cuts of all: crystal structures of Kex2 and furin reveal secrets of precursor processing." Trends Biochem Sci **29**(2): 80-7.

Rovere, C., J. Luis, et al. (1999). "The RGD motif and the C-terminal segment of proprotein convertase 1 are critical for its cellular trafficking but not for its intracellular binding to integrin alpha5beta1." J Biol Chem **274**(18): 12461-7.

Ruan, B., J. Hoskins, et al. (1999). "Rapid folding of calcium-free subtilisin by a stabilized pro-domain mutant." Biochemistry **38**(26): 8562-71.

Ruan, B., J. Hoskins, et al. (1998). "Stabilizing the subtilisin BPN' pro-domain by phage display selection: how restrictive is the amino acid code for maximum protein stability?" Protein Sci **7**(11): 2345-53.

Ruvinov, S., L. Wang, et al. (1997). "Engineering the independent folding of the subtilisin BPN' prodomain: analysis of two-state folding versus protein stability." Biochemistry **36**(34): 10414-21.

Sajid, M., C. Withers-Martinez, et al. (2000). "Maturation and specificity of Plasmodium falciparum subtilisin-like protease-1, a malaria merozoite subtilisin-like serine protease." J Biol Chem **275**(1): 631-41.

Schwede, T., J. Kopp, et al. (2003). "SWISS-MODEL: An automated protein homology-modeling server." Nucleic Acids Res **31**(13): 3381-5.

Segel (1975). Biochemical Calculations.

Seidah, N. G. and M. Chretien (1999). "Proprotein and prohormone convertases: a family of subtilases generating diverse bioactive polypeptides." Brain Res **848**(1-2): 45-62.

Seidah, N. G. and A. Prat (2002). "Precursor convertases in the secretory pathway, cytosol and extracellular milieu." Essays Biochem **38**: 79-94.

Shakhnovich, E., V. Abkevich, et al. (1996). "Conserved residues and the mechanism of protein folding." Nature **379**(6560): 96-8.

Shastry, M. C. and J. B. Udgaonkar (1995). "The folding mechanism of barstar: evidence for multiple pathways and multiple intermediates." J Mol Biol **247**(5): 1013-27.

Shental-Bechor, D., T. Danieli, et al. (2002). "Long-range effects on the binding of the influenza HA to receptors are mediated by changes in the stability of a metastable HA conformation." Biochim Biophys Acta **1565**(1): 81-9.

Shinde, U., X. Fu, et al. (1999). "A pathway for conformational diversity in proteins mediated by intramolecular chaperones." J Biol Chem **274**(22): 15615-21.

Shinde, U. and M. Inouye (1993). "Intramolecular chaperones and protein folding." Trends Biochem Sci **18**(11): 442-6.

Shinde, U. and M. Inouye (1994). "The structural and functional organization of intramolecular chaperones: the N-terminal propeptides which mediate protein folding." J Biochem (Tokyo) **115**(4): 629-36.

Shinde, U. and M. Inouye (1995). "Folding mediated by an intramolecular chaperone: autoprocessing pathway of the precursor resolved via a substrate assisted catalysis mechanism." J Mol Biol **247**(3): 390-5.

Shinde, U. and M. Inouye (1995). "Folding pathway mediated by an intramolecular chaperone: characterization of the structural changes in pro-subtilisin E coincident with autoprocessing." J Mol Biol **252**(1): 25-30.

Shinde, U. and M. Inouye (2000). "Intramolecular chaperones: polypeptide extensions that modulate protein folding." Semin Cell Dev Biol **11**(1): 35-44.

Shinde, U., Y. Li, et al. (1993). "Folding pathway mediated by an intramolecular chaperone." Proc Natl Acad Sci U S A **90**(15): 6924-8.

Shinde, U. P., J. J. Liu, et al. (1997). "Protein memory through altered folding mediated by intramolecular chaperones." Nature **389**(6650): 520-2.

Siezen, R. J. (1996). "Subtilases: subtilisin-like serine proteases." Adv Exp Med Biol **379**: 75-93.

Siezen, R. J. and J. A. Leunissen (1997). "Subtilases: the superfamily of subtilisin-like serine proteases." Protein Sci **6**(3): 501-23.

Silen, J. L. and D. A. Agard (1989). "The alpha-lytic protease pro-region does not require a physical linkage to activate the protease domain in vivo." Nature **341**(6241): 462-4.

Smith, C. A., H. S. Toogood, et al. (1999). "Calcium-mediated thermostability in the subtilisin superfamily: the crystal structure of Bacillus Ak.1 protease at 1.8 A resolution." J Mol Biol **294**(4): 1027-40.

Smith, S. M. and M. M. Gottesman (1989). "Activity and deletion analysis of recombinant human cathepsin L expressed in Escherichia coli." J Biol Chem **264**(34): 20487-95.

Sohl, J. L., S. S. Jaswal, et al. (1998). "Unfolded conformations of alpha-lytic protease are more stable than its native state." Nature **395**(6704): 817-9.

Sohl, J. L., A. K. Shiau, et al. (1997). "Inhibition of alpha-lytic protease by pro region C-terminal steric occlusion of the active site." Biochemistry **36**(13): 3894-902.

Sorenson, P., J. R. Winther, et al. (1993). "The pro region required for folding of carboxypeptidase Y is a partially folded domain with little regular structural core." Biochemistry **32**(45): 12160-6.

Spolar, R. S. and M. T. Record, Jr. (1994). "Coupling of local folding to site-specific binding of proteins to DNA." Science **263**(5148): 777-84.

Steiner, D. F. (1998). "The proprotein convertases." Curr Opin Chem Biol **2**(1): 31-9.

Steiner, D. F. (2004). "The proinsulin C-peptide--a multirole model." Exp Diabesity Res **5**(1): 7-14.

Strausberg, S., P. Alexander, et al. (1993). "Catalysis of a protein folding reaction: thermodynamic and kinetic analysis of subtilisin BPN' interactions with its propeptide fragment." Biochemistry **32**(32): 8112-9.

Strongin, A. Y., L. S. Izotova, et al. (1978). "Intracellular serine protease of Bacillus subtilis: sequence homology with extracellular subtilisins." J Bacteriol **133**(3): 1401-11.

Subbian, E., Y. Yabuta, et al. (2005). "Folding pathway mediated by an Intramolecular chaparone: Intrinsically unstructured propeptide modulates stochastic activation of subtilisin." Journal of Molecular Biology .

Subbian, E., Yabuta, Y. & Shinde, U. (2004). Positive selection dictates the choice between kinetic and thermodynamic protein folding and stability in subtilases. Biochemistry 43, 14348-60.

Suter, U., J. V. Heymach, Jr., et al. (1991). "Two conserved domains in the NGF propeptide are necessary and sufficient for the biosynthesis of correctly processed and biologically active NGF." Embo J **10**(9): 2395-400.

Swain, P. S., M. B. Elowitz, et al. (2002). "Intrinsic and extrinsic contributions to stochasticity in gene expression." Proc Natl Acad Sci U S A **99**(20): 12795-800.

Swalley, S. E., B. M. Baker, et al. (2004). "Full-length influenza hemagglutinin HA2 refolds into the trimeric low-pH-Induced conformation." Biochemistry **43**(19): 5902-11.

Swanson, W. J., Z. Yang, et al. (2001). "Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals." Proc Natl Acad Sci U S A **98**(5): 2509-14.

Takagi, H., Y. Morinaga, et al. (1988). "Mutant subtilisin E with enhanced protease activity obtained by site-directed mutagenesis." J Biol Chem **263**(36): 19592-6.

Takagi, H., T. Takahashi, et al. (1990). "Enhancement of the thermostability of subtilisin E by introduction of a disulfide bond engineered on the basis of structural comparison with a thermophilic serine protease." J Biol Chem **265**(12): 6874-8.

Takei, J., R. A. Chu, et al. (2000). "Absence of stable intermediates on the folding pathway of barnase." Proc Natl Acad Sci U S A **97**(20): 10796-801.

Takeshima, H., M. Sakaguchi, et al. (1995). "Intracellular targeting of lysosomal cathepsin D in COS cells." J Biochem (Tokyo) **118**(5): 981-8.

Takeuchi, Y., Y. Satow, et al. (1991). "Refined crystal structure of the complex of subtilisin BPN' and Streptomyces subtilisin inhibitor at 1.8 A resolution." J Mol Biol **221**(1): 309-25.

Tang, B., S. Nirasawa, et al. (2003). "General function of N-terminal propeptide on assisting protein folding and inhibiting catalytic activity based on observations with a chimeric thermolysin-like protease." Biochem Biophys Res Commun **301**(4): 1093-8.

Thacker, C. and A. M. Rose (2000). "A look at the Caenorhabditis elegans Kex2/Subtilisin-like proprotein convertase family." Bioessays **22**(6): 545-53.

Thomas, G. (2002). "Furin at the cutting edge: from protein traffic to embryogenesis and disease." Nat Rev Mol Cell Biol **3**(10): 753-66.

Thorne, B. A. and G. D. Plowman (1994). "The heparin-binding domain of amphiregulin necessitates the precursor pro-region for growth factor secretion." Mol Cell Biol **14**(3): 1635-46.

Tompa, P. (2002). "Intrinsically unstructured proteins." Trends Biochem Sci **27**(10): 527-33.

Tornero, P., V. Conejero, et al. (1996). "Primary structure and expression of a pathogen-induced protease (PR-P69) in tomato plants: Similarity of functional domains to subtilisin-like endoproteases." Proc Natl Acad Sci U S A **93**(13): 6332-7.

Truhlar, S. M., E. L. Cunningham, et al. (2004). "The folding landscape of Streptomyces griseus protease B reveals the energetic costs and benefits associated with evolving kinetic stability." Protein Sci **13**(2): 381-90.

Tseng, Y. Y. and J. Liang (2004). "Are residues in a protein folding nucleus evolutionarily conserved?" J Mol Biol **335**(4): 869-80.

Valls, L. A., C. P. Hunter, et al. (1987). "Protein sorting in yeast: the localization determinant of yeast vacuolar carboxypeptidase Y resides in the propeptide." Cell **48**(5): 887-97.

van den Hazel, H. B., M. C. Kielland-Brandt, et al. (1993). "The propeptide is required for in vivo formation of stable active yeast proteinase A and can function even when not covalently linked to the mature region." J Biol Chem **268**(24): 18002-7.

van Nuland, N. A., F. Chiti, et al. (1998). "Slow folding of muscle acylphosphatase in the absence of intermediates." J Mol Biol **283**(4): 883-91.

Voorberg, J., R. Fontijn, et al. (1993). "Biogenesis of von Willebrand factor-containing organelles in heterologous transfected CV-1 cells." Embo J **12**(2): 749-58.

Wang, L., B. Ruan, et al. (1998). "Engineering the independent folding of the subtilisin BPN' pro-domain: correlation of pro-domain stability with the rate of subtilisin folding." Biochemistry **37**(9): 3165-71.

Weissman, J. S. and P. S. Kim (1992). "The pro region of BPTI facilitates folding." Cell **71**(5): 841-51.

Wiederanders, B., G. Kaulmann, et al. (2003). "Functions of propeptide parts in cysteine proteases." Curr Protein Pept Sci **4**(5): 309-26.

Winther, J. R. and P. Sorensen (1991). "Propeptide of carboxypeptidase Y provides a chaperone-like function as well as inhibition of the enzymatic activity." Proc Natl Acad Sci U S A **88**(20): 9330-4.

Winther, J. R., P. Sorensen, et al. (1994). "Refolding of a carboxypeptidase Y folding intermediate in vitro by low-affinity binding of the proregion." J Biol Chem **269**(35): 22007-13.

Withers-Martinez, C., L. Jean, et al. (2004). "Subtilisin-like proteases of the malaria parasite." Mol Microbiol **53**(1): 55-63.

Wolff, A. M., M. S. Showell, et al. (1996). "Laundry performance of subtilisin proteases." Adv Exp Med Biol **379**: 113-20.

Wong, S. L. and R. H. Doi (1986). "Determination of the signal peptidase cleavage site in the preprosubtilisin of Bacillus subtilis." J Biol Chem **261**(22): 10176-81.

Yabuta, Y., E. Subbian, et al. (2003). "Folding pathway mediated by an intramolecular chaperone. A functional peptide chaperone designed using sequence databases." J Biol Chem **278**(17): 15246-51.

Yabuta, Y., E. Subbian, et al. (2002). "Folding pathway mediated by an intramolecular chaperone: dissecting conformational changes coincident with autoprocessing and the role of Ca(2+) in subtilisin maturation." J Biochem (Tokyo) **131**(1): 31-7.

Yabuta, Y., H. Takagi, et al. (2001). "Folding pathway mediated by an intramolecular chaperone: propeptide release modulates activation precision of pro-subtilisin." J Biol Chem **276**(48): 44427-34.

Yamagata, H., T. Masuzawa, et al. (1994). "Cucumisin, a serine protease from melon fruits, shares structural homology with subtilisin and is generated from a large precursor." J Biol Chem **269**(52): 32725-31.

Yang, Z. (1997). "PAML: a program package for phylogenetic analysis by maximum likelihood." Comput Appl Biosci **13**(5): 555-6.

Yang, Z. (2002). "Inference of selection from multiple species alignments." Curr Opin Genet Dev **12**(6): 688-94.

Young, J. C., V. R. Agashe, et al. (2004). "Pathways of chaperone-mediated protein folding in the cytosol." Nat Rev Mol Cell Biol **5**(10): 781-91.

Zhong, M., J. S. Munzer, et al. (1999). "The prosegments of furin and PC7 as potent inhibitors of proprotein convertases. In vitro and ex vivo assessment of their efficacy and selectivity." J Biol Chem **274**(48): 33913-20.

Zhou, A., S. Martin, et al. (1998). "Regulatory roles of the P domain of the subtilisin-like prohormone convertases." J Biol Chem **273**(18): 11107-14.

Zhu, H., S. Huang, et al. (2004). "The next step in systems biology: simulating the temporospatial dynamics of molecular network." Bioessays **26**(1): 68-72.

Zhu, X. L., Y. Ohta, et al. (1989). "Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process." Nature **339**(6224): 483-4.