Chapter 1

Introduction
1. Introduction

With the advent of scientific and technological resources, we human beings have been successful in overcoming various deadly diseases. Still, we face challenges of newly emerging diseases, whether of infectious type or because of physiological disorders due to sedentary lifestyle. Hence, pharmaceutical companies are always trying hard to bring new drugs to market. Traditionally, there was a trend of using small chemical drug molecules because their physicochemical properties and structural stabilities are more controllable and predictable. But recently, employment of proteins as pharmaceutical agents is increasing as these generally display therapeutically favorable properties such as higher target specificity and pharmacological potency when compared to traditional small molecule drugs. Unfortunately, the structural instability issues generally displayed by this class of molecules still remain one of the biggest challenges to their pharmaceutical employment, as these can negatively impact their final therapeutic efficacies. Also, because of the structural complexity and diversity arising due to the macromolecular nature of proteins, it is more difficult to control and predict their physicochemical properties and structural stability.

Although, the primary structure of a protein is subject to the same chemical instability issues as traditional chemical drug molecules (e.g., acid-base and redox chemistry, chemical fragmentation, etc), the higher levels of protein structure (e.g., secondary, tertiary), often necessary for therapeutic efficacy, are subjected to additional physical instability issues (e.g., irreversible conformational changes, local and global unfolding) as their integrity is mainly maintained by non-covalent interactions. Also, the long-term stabilization of protein pharmaceuticals is hampered with the fact that there is only a marginal difference in energy levels of their folded and unfolded states and this is a major hurdle for the long-term stabilization of protein pharmaceuticals.

Mainly, sensitivity/stability of the proteins to physicochemical environment is of special relevance in a pharmaceutical production setting where proteins can be simultaneously exposed to varied environments during their production, purification, storage and delivery. Hence, much emphasis has been given to increase the long-term stability of protein pharmaceuticals. This could be done either by use of stabilizing excipients (e.g., amino acids, sugars, polyols) and/or by altering the characteristics of the protein itself. The protein can be altered either by chemical modifications of the protein like glycosylation or PEGylation or by engineering of the gene encoding protein.
for increased stability.

To engineer the proteins, factors responsible for their enhanced stability to various denaturing agents like higher temperature, extreme acidic/basic pH range and presence of various salts must be known and are being searched widely. In the present thesis, we have explored the presence of polyproline type II (PPII) fold in an alkaline serine protease from non-streptomyces actinomycetes sp. as a responsible factor for kinetic stability.

1.1 Serine proteases

Proteases, commonly called peptidases, are most ubiquitous and probably one of the most studied enzymes. Still, they continue to amaze us by their varied structures, stabilities and functionalities. They are vital for survival of all the living cells. Some peptidases recycle polypeptides into their constitutive amino acids, whereas others catalyze selective polypeptide cleavage for post-translational modification. Proteases have evolved multiple times, and different classes of protease can perform the same reaction by completely different catalytic mechanisms. Interesting thing is that their sequence, hence structure and stability differ based on their source, yet they catalyze the same function. Proteases represent approximately 2% of the total number of proteins present in all types of organisms [1]. There are about 500 human genes that encode peptidases and their homolog. Many of these enzymes are of medical importance, and are potential drug targets that originate from the human genome or from the genome of the disease causing organisms.

The importance of proteases can be illustrated by growing number of publications about proteases each year (around 8000/annum) [2]. The latest MEROPS database (9.9) contains information about 413 834 peptidases which are grouped into 244 families and 55 clans [3]. The MEROPS database has organized proteases into different families as well as clans. The organization into families on the basis of amino acid sequence homology, and the families are assembled into clans on the basis of similarities in tertiary structure [4].

On the basis of catalytic site residues proteases can be classified into following classes: cysteine, serine, metallo, aspartic acid, threonine, glutamic acid, aspargine, mixed and unknown type of proteases. Along with the difference in active site residues, these proteases also differ on the basis of presence or absence of a covalent acyl-enzyme
intermediate on the reaction pathway. The catalyses of serine and cysteine peptidases involve the covalent intermediate ester and thioester, respectively, whereas the aspartic and the metallopeptidase catalyses do not. During hydrolysis carried out by the latter two groups, the substrate is attacked directly by a water molecule rather than by a serine or cysteine residue.

Serine protease class (E.C. 3.4.21) comprise nearly one-third of all known proteases identified to date and play crucial roles in a wide variety of cellular as well as extracellular functions, including the process of blood clotting, protein digestion, cell signaling, inflammation, and protein processing. The serine protease class uses the classical Ser/His/Asp catalytic triad mechanism, where serine is the nucleophile, histidine is the general base and acid, and the aspartate helps orient the histidine residue and neutralize the charge that develops on the histidine during the transition states. Four distinct three-dimensional protein folds that catalyze hydrolysis of peptide bonds use combination of these three residues with identical structural configuration, suggesting four different evolutionary origins. Examples of these folds are observed in trypsin, subtilisin, prolyl oligopeptidase, and ClpP peptidase [4].

In addition to the Ser/His/Asp serine proteases, there are serine proteases that use catalytic residue arrangements other than the canonical triad. These atypical serine proteases use novel triads such as Ser/His/Glu, Ser/His/His, or Ser/Glu/Asp, dyads such as Ser/Lys or Ser/His, or a single Ser catalytic residue. There are also proteases in which the nucleophilic hydroxyl is derived from threonine rather than a serine residue.

The reason behind different serine/threonine proteases using different active site configuration is being widely studied. One possible reason is that different active site arrangements may allow for activity in a different cellular environment. For instance, a higher pH optimum is observed for the proteases with Ser/Lys active sites than Ser/His/Asp proteases. In contrast, the pH optimum is lower for serine proteases with Ser/Glu/Asp active sites than those with Ser/His/Asp active sites. In addition, variations in the active site architecture of proteases may influence what cellular inhibitor they are susceptible to.
Serine proteases: classification, structure and function (As per MEROPS database)

Clan SB peptidases
Peptidases from this clan are predominantly present in plant and bacterial genomes. However, these proprotein convertases are vital for protein processing in all animals. The classic example of this clan is subtilisin. Interestingly, the catalytic Asp/His/Ser triad exists in the exact geometric organization as observed in clan PA peptidases, yet their tertiary structures are totally different. They have an α/β-twisted open sheet structure. A second family of peptidases S53, the sedolisins, is also part of Clan SB. The His general base is substituted by a Glu residue in these peptidases. And the tetrahedral intermediate is stabilized by a negatively charged carboxyl group from an Asp residue rather than through partial positive charges. The sedolisins are active at acidic pH.

Subtilisins have been very useful in protein engineering studies. Various interesting characteristics such as substrate selectivity, thermal stability, cold adaptation, stability in non-aqueous solvents have all been introduced into subtilisin through rational mutagenesis and directed evolution approaches. Subtilisins have been engineered as improved laundry detergents. Mostly their physiological function is nutrition oriented and they are secreted outside of the cell or localized to the cell membrane.

Clan SC peptidases
Clan SC peptidases are the second largest family of serine peptidases in the human genome. They contain α/β hydrolase-fold in which the parallel β-strands are surrounded by α helices. Prolyl oligopeptidase (PDB ID 1QFS) is a classical example from this clan. In addition to proteolytic activity they can act as an esterase, lipase, dehalogenase, haloperoxidase, lyase, or epoxide hydrolase as the α/β hydrolase-fold provides a versatile catalytic platform. The peptidases from this clan have an identical geometry to the catalytic triad observed in clans PA and SB but the order of residues is different in the polypeptide sequence. Nearly all serine peptidases have activity restricted within the range of neutral to alkaline pH. However, carboxypeptidases from family S10, within clan SC, are unique for their ability to maintain catalytic activity in acidic environments. Clan SC peptidases are thought to be particularly important in cell signaling mechanisms.
Table 1: Classification of serine proteases in MEROPS database.

<table>
<thead>
<tr>
<th>Clan</th>
<th>Families</th>
<th>Representative member</th>
<th>Fold</th>
<th>Catalytic residues</th>
<th>PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA</td>
<td>14</td>
<td>Chymotrypsin</td>
<td>β barrels</td>
<td>His, Asp, Ser</td>
<td>1DPO</td>
</tr>
<tr>
<td>SB</td>
<td>2</td>
<td>Subtilisin, Sedolisin</td>
<td>alpha/beta</td>
<td>Asp, His, Ser</td>
<td>1SCN</td>
</tr>
<tr>
<td>SC</td>
<td>6</td>
<td>Prolyl oligopeptidase</td>
<td>alpha/beta/alpha</td>
<td>Ser, Asp, His</td>
<td>1QFS</td>
</tr>
<tr>
<td>SE</td>
<td>3</td>
<td>D-Ala-D-Ala carboxypeptidase B</td>
<td>Nterminal: helical bundle,Cterminal: α/β/α sandwich</td>
<td>Ser, Lys</td>
<td>3PTE</td>
</tr>
<tr>
<td>SF</td>
<td>2</td>
<td>UmuD protein</td>
<td>coiled β sheets, a β barrel</td>
<td>Ser, Lys (or His)</td>
<td>1UMU</td>
</tr>
<tr>
<td>SH</td>
<td>5</td>
<td>Cytomegalovirus assemblin</td>
<td>β barrel surrounded by helices</td>
<td>His, Ser, His</td>
<td>1LAY</td>
</tr>
<tr>
<td>SJ</td>
<td>3</td>
<td>Lon-A peptidase</td>
<td>α + β</td>
<td>Ser, Lys</td>
<td>1RR9</td>
</tr>
<tr>
<td>SK</td>
<td>3</td>
<td>Clp peptidase (type 1)</td>
<td>α superhelix + β strands</td>
<td>Ser, His, Asp</td>
<td>1TYF</td>
</tr>
<tr>
<td>SP</td>
<td>1</td>
<td>Nucleoporin 145</td>
<td>All β</td>
<td>His, Ser</td>
<td>1KO6</td>
</tr>
<tr>
<td>SO</td>
<td>1</td>
<td>Enterobacteria phage K1F</td>
<td>α / β</td>
<td>Ser, Lys</td>
<td>3GW6</td>
</tr>
<tr>
<td>SR</td>
<td>1</td>
<td>Lactoferrin</td>
<td>αβ sandwich</td>
<td>Lys, Ser</td>
<td>1LCT</td>
</tr>
<tr>
<td>SS</td>
<td>1</td>
<td>Murein tetrapeptidase</td>
<td>β sheet + β barrel</td>
<td>Ser, Glu, His</td>
<td>1ZRS</td>
</tr>
<tr>
<td>ST</td>
<td>1</td>
<td>Rhomboid</td>
<td>α-barrel</td>
<td>Ser, His</td>
<td>2IC8</td>
</tr>
</tbody>
</table>

Clan SF peptidases

Peptidases from clan SF use a dyad of Ser and Lys in prokaryotes or Ser and His in eukaryotic peptidases. They have an endoproteolytic catalytic activity. These enzymes have an all β-strand structure. Most peptidases of clan SF are self-activating and involve considerable conformational change following bond hydrolysis, e.g. LexA repressor.

Clan SJ peptidases

Clan SJ peptidases utilize a catalytic dyad of Ser and Lys for catalysis. Unique property of clan SJ peptidases is the ATP-dependent nature of proteolysis and ability to act as protein-activated ATPases. Lon peptidase from E. coli is perhaps the best-known example of the family. Lon peptidases, along with peptidases from other clans, are responsible for intracellular protein levels in E. coli and other bacteria.
Figure 1: Crystal structures of various proteases from different clans. The respective PDB IDs have been mentioned in table 1.

Clan SK peptidases
Clan SK peptidases are widely distributed in bacteria and play a role in intracellular protein levels and protein processing. **ClpP peptidase** is an important peptidase for protein turnover in *E. coli*, which utilizes a conventional catalytic triad of residues but in a novel arrangement of Ser, His and then Asp in the polypeptide sequence.
Clan SH, SP, SQ, SR peptidases
Family S21 assemblin peptidases of SH clan are present in several viral genomes and utilize a novel triad of His, Ser, His, which is not observed in any other serine peptidase family. The S59 family of clan SP contains the self processing nucleoporins. Nucleoporins present a unique dyad arrangement of His and Ser separated by only a single residue. Autolytic cleavage is similarly observed in the S58 family of clan SQ peptidases such as the DmpA aminopeptidase from Ochrobactrum anthropi. Lactoferrin in clan SR displays proteolytic activity towards a number of proteins from pathogenic organisms including Haemophilus influenzae serine-type IgA endopeptidase and the hap peptidase, as well as the EspB protein from E. coli.

Clan SS and ST peptidases
Clan SS peptidases have a catalytic triad of Ser, Glu and His. LD-Carboxypeptidases in the S66 family of clan SS are capable of peptide bond hydrolysis between L- and D-amino acids in bacterial peptidoglycan and are thought to play a role in peptidoglycan recycling. Clan ST peptidases hydrolyze peptide bonds within a phospholipid bilayer. Metallopeptidases are known to play a role in sterol metabolism and the aspartyl peptidase g-secretase liberates β-amyloid peptides involved in Alzheimer’s disease. Rhomboids are serine peptidases capable of intramembrane proteolysis conserved in all kingdoms of organisms from bacteria to man. Rhomboids have diverse functions including quorum sensing, mitochondrial morphology and dynamics, and intracellular signaling.

Clan PA peptidases
Clan PA proteases with the trypsin fold are the largest family of serine proteases and perhaps the best studied group of enzymes. Peptidases from the S1 family of PA clan have the even distribution of catalytic residues across the entire polypeptide sequence. Two six-stranded β-barrels are arranged perpendicular to each other and the residues of the catalytic triad are present at their interface. Two residues of the triad are from the N-terminal β-barrel with the nucleophilic Ser and oxyanion hole generated from the C-terminal β-barrel. β-strand topology of the fold clearly evidences the classical Greek-key architecture.
Family S1 of clan PA is grouped into two subfamilies, S1A and S1B which are phylogenetically distinct but share a common two β-barrel architecture. The S1A proteases mediate a variety of extracellular processes while the S1B proteases are found in all cellular life and are responsible for intracellular protein turnover. S1A proteases have a limited distribution in plants, prokaryotes, and the archaea. Nearly all clan PA peptidases utilize the canonical catalytic triad, but a few family members of viral origin use an active site thiol from a Cys residue. Many chymotrypsin-like serine peptidases are expressed as inactive zymogen precursor protein and require proteolytic processing for activation.

1.2 Kinetic stability

For proteases to work in different environments different active site arrangements are used in serine proteases across the three kingdoms of life. Also the secreted bacterial proteases, whose biological function is to provide nutrients for the bacterium, are highly evolved to maintain their native states under harsh, highly proteolytic conditions. As they function in extracellular environment, there is no need for their regulated turnover. The native states of these proteases are very stable and are difficult to unfold. These proteases and such other proteins are called as “kinetically stable proteins”.

The protein structure is maintained by various covalent and non covalent interactions, mainly by hydrogen bonds. Using these interactions, which provide protein the structural stability, it must maintain its shape for the length of time for which it is required to perform its designated function(s) in the environment in which it is required to function. Protein stability comes in two flavors: thermodynamic stability, which is related to a low amount of unfolded and partially-unfolded states in equilibrium with the native, functional protein and kinetic stability, which is related to a high free-energy barrier separating the native state from the non-functional forms (unfolded states, irreversibly-denatured protein) [5]. This high energy barrier allows the protein to maintain its biological function at least during a physiologically relevant time-scale; sometimes, even, if the native state is not thermodynamically stable with respect to non-functional forms. For proteins which often work under conditions (harsh extracellular or crowded intracellular environments) in which deleterious alterations (proteolysis, aggregation, undesirable interactions with other macromolecular components) are prone to occur, kinetic stabilization is very important. Also, kinetic stability may provide a
mechanism for the evolution of optimal functional properties. Furthermore, enhancement of kinetic stability is essential for many biotechnological applications of proteins.

1.2.1 Thermodynamic stability versus kinetic stability

Determination of protein stability in vitro is usually done by performing denaturation experiments in simple solvent conditions, involving comparatively short time-scales and using small model protein systems. Many times reversible denaturation is found in these experiments and analysis of the denaturation profiles is done on the basis of equilibrium thermodynamics. In the simplest case, the data can be fitted with the two-state equilibrium model:

\[ \text{N} \leftrightarrow \text{U} \]

Where \( \text{N} \) is the native state and \( \text{U} \) is the unfolded state (actually an ensemble of more or less unfolded conformations) and \( K \) is the unfolding equilibrium constant,

\[ K = \frac{U}{N} \]

Which is related to the standard unfolding free-energy change \( [\Delta G = G(\text{U}) - G(\text{N})] \) through the Lewis equation,

\[ \Delta G = -RT \ln K \]

\( \Delta G \) as a function of an environment parameter (temperature, concentration of a chemical denaturant) is obtained after the fit of the model to the experimental profiles. Extrapolation to physiological conditions (e.g. zero concentration of chemical denaturant, 37 °C), leads typically to a positive (although not too large) value of the unfolding free-energy change, indicating that \( K < 1 \) and that the unfolding equilibrium is shifted towards the folded state. And this explains the concept of thermodynamically stable proteins, in which the biological function of the protein is guaranteed if equilibrium is established between the native and the unfolded states of the protein and the unfolding thermodynamics favors the folded state under physiological conditions. But many a times, even if the native state is not thermodynamically stable with respect to non-functional states, the biological function of the protein can be maintained, at least during a certain physiologically relevant time-scale. This is possible because of a sufficiently high free-energy barrier separating the native state from the non-functional forms.
Kinetic stability is a measure of rate of protein unfolding. It is particularly important for proteins that unfold very slowly or denature irreversibly. In such cases, the free energy difference between the folded and unfolded state is not of much importance as that will only affect the equilibrium and it is not an equilibrium process. The important thing is the free energy difference between the folded and the transition states (activation energy), as it is the magnitude of this difference that determines the rate of unfolding (and hence inactivation).

![Energy Diagram](image)

**Figure 2:** An energy diagram showing difference between a typical protein and a kinetically stable protein; alpha lytic protease. (Adapted from Jaswal et al [37]).

1.2.2 Alpha lytic protease (αLP): A model protein for studying kinetic stability

αLP and SGPB (*Streptomyces griesus* protease B) are well studied kinetically stable bacterial proteases which are of the chymotrypsin superfamily. These proteases display extraordinarily slow folding ($t_{1/2}$, folding= 1800 years for αLP [6] and 3 days for SGPB [7]) and unfolding rates ($t_{1/2}$, unfolding=1.2 years for αLP and 11 days for SGPB. Surprisingly, the native state of αLP is actually less stable than unfolded forms by ~4 kcal/mol [6]. Although a basic principle of protein folding is that the native state of a protein is at the minimum free energy [8], both the I and fully unfolded states of αLP are lower in free energy than the native state. Native αLP is thus metastable, its apparent stability derives from a large barrier to unfolding. Consequently, the evolution of αLP has been distinct from most other proteins; it has not been constrained by the
free-energy difference between the native and unfolded states, but instead by the size of its unfolding barrier.

1.2.3 Pro-region: as a folding catalyst
αLP and virtually all other extracellular bacterial proteases have coevolved a covalently attached pro region that catalyzes their folding [9,10]. This is to facilitate folding on a biological timescale. The pro-region is proteolytically destroyed after folding has completed, yielding the kinetically trapped native protease [11]. This amino-terminal pro-region is essential for its folding in vivo and in vitro [12, 13]. The protease folds to an inactive, partially folded state designated ‘I’, in the absence of the pro-region. The folding of native state of αLP, which is separated from I state by the folding transition state (>26 kcal mol⁻¹), is catalyzed by pro-region [14].

As, both I and U are more stable than the folded N state, the pro-region shifts the energetic thermodynamically to favour the N:pro-region complex over the I:pro-region complex, thereby providing the driving force that brings the protease into its native, active conformation. After proteolytic degradation of the pro-region, αLP gets locked in its metastable active N state. The pro-region actively catalyses folding reaction by stabilizing the folding transition state. Mutants of pro-region have been identified that preferentially alter this transition-state stabilization [15].

1.2.4 Energetics of stabilization of native state of αLP
At 10 °C, the native conformation is favoured enthalpically by 18 ±1.5 kcal mol⁻¹ over I. Therefore, the stability of I must be entropic in origin. The αLP sequence contains 16 % glycine compared to only 9 % in chymotrypsin (a close homolog but thermodynamically stable protein), is the reason of excess entropy in αLP. As glycines lack a side chain, there is increase in the number of conformations accessible to unfolded states. It was predicted that the 10 additional glycines in αLP contribute an additional, 7 kcal mol⁻¹ of configurational entropy to the unfolded state compared to that of chymotrypsin at 4 °C [16]. Also it was suggested that the removal of this entropic source alone would be sufficient to place the N state at the global free-energy minimum for αLP.
1.2.5 Conformational rigidity of αLP

Unusually, low conformational entropy of the αLP N state may also contribute to increasing the unfolding entropy. Although, native states of proteins are often dynamic, the N state of αLP is quite rigid so that the protease is not readily digested by itself or by other proteases. The fact that the loops in αLP are generally shorter and are likely to be less flexible than those found in chymotrypsin, could be the reason of structural rigidity of αLP. Many of the glycines which are unique to αLP are found in these loops and may reduce the entropy of the N state while increasing the configurational entropy of the I state. The rigidity of the native state of αLP may be the result of evolutionary pressures to suppress autolysis and thereby extend the lifetime of the protease, which provide nutrients for its host.

There is a strong correlation between high glycine content and the presence of a pro-region. As the average glycine content of αLP and other chymotrypsin-like proteases made with pro-regions is 18 %, compared with 9 % in family members that do not contain a pro-region. Incorporating more glycines may be one mechanism by which large kinetic barriers and pro-region catalysts have co-evolved in extracellular bacterial proteases to prolong their functional lifetime. Large barriers have been observed in the folding of several other proteins that are not synthesized with pro-regions. The kinetic barriers of influenza virus haemaglutinin (HA2), luciferase, the serpin PAI-1, and of the prion protein PrP, all function to enhance the stability of compact native state [17-21].

1.2.6 Identification of kinetically stable proteins: general approaches

A wide variety of experimental results provide indication of high free-energy barriers and likely kinetic stabilization of proteins: metastability of folded states and/or existence of alternative native-like states [22,23] and low unfolding rates [24,25] under physiological conditions (obtained, for instance, by extrapolation to zero denaturant concentration of the rate constant for urea- or guanidine-induced unfolding).

Historically, evidence for protein kinetic stability came from two kinds of studies: differential scanning calorimetry analyses on irreversible protein denaturation and analysis of stability of alpha lytic protease system. Recent approaches to the determination of kinetic stability in a proteomic scale are the resistance to SDS and proteolysis approach.
Chapter 1: Introduction

A. Differential scanning calorimetry

The usual equilibrium test in DSC is the operational reversibility of the denaturation process: for the denaturation process to be considered reversible, a second (re-heating) scan, performed after cooling to room temperature, should yield a transition comparable to that in the first scan.

Protein thermal denaturation is often found to be calorimetrically irreversible, as no transition is detected in the re-heating run [26, 27]. This suggests that a Lumry–Eyring kind of mechanism applies, with some irreversible alteration step taking the protein to a final state unable to fold back to the native state.

\[ N \leftrightarrow U \rightarrow F \]

Where, N is the native state, U is unfolded state and F is the final state unable to fold back to native state.

When the irreversible alteration step \( U \rightarrow F \) is fast enough, in such a way that any molecule U formed is immediately converted to F, a two-state kinetic process applies [28],

\[ N \rightarrow F \]

This theoretical proposal of the two-state irreversible model allowed researchers to make reliable tests of the possible kinetically-controlled character of the experimental thermal denaturation profiles. Also, scan rate dependence of protein thermal denaturation is a peculiarity of kinetically stable proteins.

Recently, two promising procedures allowing the high-throughput, proteomic-scale analysis of protein kinetic stability have been developed by Colon and coworkers [29] and Marqusee and coworkers [30].

B. The resistance to SDS approach

SDS is known to denature proteins. However, unlike common chemical denaturants, its mode of action seems based on its ability to irreversibly trap the proteins during the time in which they are transiently or partially unfolded. Denaturation by SDS seems, therefore, a potential probe of protein kinetic stability; i.e., kinetically-stable proteins are expected to be highly resistant to denaturation by SDS. The resistance to SDS induced denaturation can be investigated by comparing the migration on polyacrylamide gels of identical boiled and unboiled protein samples containing SDS. Proteins which are not kinetically stable become denatured even in unboiled condition.
and, therefore, migrate the same distance as that of boiled sample in presence of SDS. On the other hand, kinetically-stable proteins are expected to survive in unboiling condition and to denature only after boiling in SDS, therefore, they will migrate a shorter distance in the first case. E.g. avidin, papain, serum amyloid P, streptavidin, transthyretin.

**C. The resistance to proteolysis approach**

Transient access to high-energy, partially-unfolded states in which the cleavable states become exposed leads to proteolysis of compact, folded proteins. For αLP and SGPB, the native states are extremely rigid and hence both global and local unfolding processes are limited. This significantly reduces their susceptibility to proteolytic degradation and increases their functional lifetime [31]. As these proteins function in highly proteolytic conditions, this property appears to be a result of the functional optimization for their survival. Since even subtle subglobal unfolding events can make a protein susceptible to exogenous proteolysis [32-36], the typical breathing and partial unfolding motions [37] observed in native proteins are lacking in αLP and SGPB as they have evolved highly rigid native states. The resistance to proteolysis approach can also be applied to identify kinetically stable proteins in a proteome e.g. this approach has been used to identify kinetically stable proteins in *E. coli* [30].

**1.2.7 The physical basis for kinetic stability**

Current experimental evidences do not show any clear relation between kinetic stability and protein structure/fold. e.g., the studies of Marqusee and coworkers did not highlight any common structural features that could explain proteolytic resistance [30]. Also, the studies of Colon and coworkers on resistance to denaturation by SDS did not show any dramatic trends. Except, the presence of predominantly β-sheet and oligomeric structures emerged as a common characteristic of most of the kinetically stable proteins that they studied.

The physical basis for kinetic stability is poorly understood and no structural consensus has been found to explain this phenomenon. In previous studies, the addition of hydrophobic residues on the protein surface [38], the engineering of disulfide bonds [39], and the introduction of metal-binding sites [40] have been shown to increase kinetic stability. A connection between kinetic stability and oligomeric quaternary
structure has also been proposed [41]. Electrostatic interactions have been suggested to be a major factor in case of slow unfolding of some hyperthermophilic protein [42, 43]. Also, there is evidence that even at low pH, where electrostatic interactions should be significantly weakened some kinetically stable proteins retains their slow unfolding rate [44, 45]. Thus, it appears that no common structural feature exists to explain kinetic stability, and perhaps this property may be achieved by different means, depending on the individual protein.

In the present work, we have studied polyproline type II (PPII) helix as a responsible factor for kinetic stability of a serine protease.

1.3 Polyproline type II (PPII) helix

The biological function of proteins depends on their unique three dimensional structures. Although the overall protein structures are complex, Pauling and Corey had proposed two types of regularities in the local backbone conformation i.e. alpha helices and beta sheets [46-48]. Previously, the secondary structure composition was mainly restricted to α helix, β sheets and a state corresponding to other regions in the backbone, the random coil. Recently, a large number of experimentally determined protein structures are becoming available and it is clear that other backbone conformations are also favored in proteins. One of those secondary structures is the polyproline II (PPII) helix. It appears as local conformation in proteins [49]. The PPII helix was discovered more than 60 years ago in fibrous proteins [50, 51] like collagen. They contributed to the coiled coil supersecondary structures in these proteins.

Previously, the polyproline left handed helical structure was often confused with unordered, disordered, irregular, unstructured, extended, or random coil conformations because it is neither α helical nor β turn or β sheet, i.e., a classical structure. However, later on it was realized that it is a common conformation in the unfolded state and this has led to a large number of studies. Interestingly, this was first suggested in 1968 by Tiffany and Krimm [52] based on the similarities between CD spectra of unfolded proteins and proline polymers. The preference for PPII in denatured peptides has many implications, particularly with respect to modeling the denatured state or understanding the determinants of protein stability [53].
1.3.1 Characteristics of PPII helix

PPII helix is a left-handed helical structure with an overall shape resembling a triangular prism [54, 55]. It forms an extended helix, with a helical pitch of 9.3 Å/turn,-

![Figure 3: Side and top views of two ideal helical conformations of a polyproline peptide](image)

(a) polyproline I helix (PPI), a right-handed helix with backbone dihedral angles of \((\phi, \psi, \omega) = (-75^\circ, 160^\circ, 0^\circ)\); (b) polyproline II helix (PPII, polypro II), a left-handed helix with backbone dihedral angles of \((\phi, \psi, \omega) = (-75^\circ, 146^\circ, 180^\circ)\), respectively. (Adapted from M Moradi et al [56]).

- each turn constituting of 3 residues. This conformation is characterized by recurrent trans isomers of peptide bonds and \((\Phi, \Psi)\) values of -75° and +145° respectively.

Due to steric restrictions, the main chain hydrogen-bond donors and acceptors cannot be easily satisfied in PPII helix. There are no local hydrogen bonds, in a manner similar to that of \(\alpha\) helices; also hydrogen bonding potential of neighboring residues in polyproline conformation is not satisfied in a manner analogous to \(\beta\) strands [57]. And because of this lack of characteristic local main chain hydrogen bonding patterns they are harder to detect directly in NMR spectra than \(\alpha\) helices and \(\beta\) sheets. Also, they cannot be identified in protein structures by hydrogen-bonding patterns, as is commonly done for other secondary structure types [58, 59]. For example, PPII helices are not assigned by the widely used Secondary Structure Assignment Method (SSAM): DSSP [58]. As a consequence, residues in PPII conformations are often inappropriately assigned to “disordered” or “random” conformation classes.

Recently, Mansiaux Y et al have developed the assignment rule for PPII helix which could be coupled with DSSP for further analysis [60]. Previously, the PPII structure has not been studied widely because: (i) It has a low frequency of occurrence, (ii) it was considered as an unstable conformation as it is not stabilized by internal hydrogen
bonding and, (iii) only a few methods for PPII assignment are available and these methods use different assignment parameters resulting in variable assignments.

1.3.2 PPII in absence of proline

The left-handed helix adopted by polyproline in aqueous solution was initially considered to be limited to proline-rich polypeptides such as poly(Pro), poly(Hyp), and the (Pro-X-Gly)_n polypeptides making up the collagen triple helix. But later studies suggest that the PPII helices can be found even if no prolines are present in the sequence [61-65] e.g., short stretches of polyglutamine and polylysine were found to form PPII conformation. And these peptides have been very well studied [66-69].

1.3.3 Isodichroic Point

PPII conformations of variable lengths and lifetimes are formed in solutions of above mentioned polypeptides. Each peptide residue in a random coil ensemble may be considered to participate in a PPII /unordered equilibrium, where unordered denotes all non- PPII conformations [70]. Such equilibria may be represented collectively as the mean fractional population of residues in the PPII conformation. Changes in the ellipticity at the positive maximum in the CD spectrum that generate an isodichroic value are interpreted to indicate a change in the fractional population of residues in the PPII conformation. The fractional population of PPII conformation is decreased by raising the temperature or by addition of neutral salts and increased by lowering the temperature or by addition of denaturants, such as urea or guanidinium chloride [52, 67, 71]. These denaturants are thought to increase the fractional PPII population by hydrogen bonding preferentially with its peptide backbone.

1.3.4 Identification of PPII fold

The PPII conformation is increasingly recognized as an important element in peptide and protein conformation. In spite of the regularity of the PPII structure and well defined dihedral angle values, a typical feature of PPII structure is the absence of any intramolecular hydrogen bonds. And this feature makes the PPII structure indistinguishable from an irregular backbone structure by 1H-NMR spectroscopy. Circular Dichroism (CD) is one of the most useful methods for detecting and characterizing PPII [72]. Also the techniques such as vibrational CD (VCD) and Raman
optical activity (ROA) have been reported for detecting the PII conformation [73, 74].

**Figure 4: CD spectrum of polyproline peptide**

*Figure 4: CD spectrum of polyproline peptide a) PPII conformation in solution [75] b) PPI in solution [76].*

Experimentally the CD instrumentation is much more widely available than ROA and VCD instrumentation and hence is more used. The characteristic spectrum of CD of a polyproline II film cast from water is described by Young MA and Pysh ES [75]. They reported that the CD spectrum of PPII is composed of the positive band at 230 nm and the stronger negative band at 214 nm (Fig. 4a). The maximum ellipticity of the negative band is four times greater than the maximum ellipticity of the positive band, and the crossover between the two bands is at 223 nm. Below the large negative band at 214 nm the ellipticity decreases in magnitude to 192 nm below which it becomes more negative. Sometimes there can be shift in the wavelength of maximum and minimum ellipticity from the above give values. This could be because of the difference between secondary and tertiary amides [77]. However, the aromatic side chains of Phe, Tyr, and Trp amino acid residues also contribute to the electronic absorption in the far UV region. Also the disulphide bond in the protein contributes to the positive maxima at 230 nm. In such cases, the identification of PPII conformation can be done recording the CD spectra in presence of denaturants. For example, CD spectra of the high molecular weight (HMW) [78] subunits of wheat glutenin, containing both PPII conformation and tyrosine residues in 6–7 mol %, contain a positive absorption at about 220 nm, which can be attributed either to PPII conformation or to aromatic residues. In this case, CD spectra
at different concentrations of urea and at different temperatures, which promote the formation of PPII like-structure, were carried out in order to get a clear assignment of the band. The strong increase of the positive band observed in 8 M urea unambiguously confirmed the assignment of the positive band at 220 nm to the PPII conformation.

1.3.5 Lack of internal hydrogen bonding: stability of PPII helix

The PPII helices are highly solvent-exposed [61] and are not stabilized by salt bridges [79]. As, they do not have internal hydrogen bonding in the main chain, they tend to have a regular pattern of hydrogen bonds with water. So, it has been suggested that PPII helices could be stabilized by water mediated main chain hydrogen bonds [80]. Several studies suggest that peptide-solvent interaction is a major determinant of PPII conformation [81-83]. Stapley and Creamer suggested that local side chain to main chain hydrogen bonds is also important in stabilizing PPII helices [63]. Recently, it has been highlighted that PPII helices are stabilized by non-local interactions [57]. As a result of the non local stabilization of hydrogen bond donors and acceptors, PPII conformations are well suited for participating in protein-protein interactions.

1.3.6 Physiological implications of presence of PPII helices in proteins

Recently, PPII structure has been demonstrated to be essential to biological activities such as signal transduction, transcription, cell motility, and immune response [84]. Many protein interactions occur via proline rich sequences that in turn are assumed to have high propensity to form PPII helices. They are also suspected to have a role in amyloid formation [85, 86] and nucleic acid binding [87].

1.3.7 Functions of PPII helix

A. Elastic Function

PPII conformation is widely present in titin, abductin, and elastin, which are elastomeric proteins. Along with other motifs, the giant elastic protein; titin, contains a motif consisting of mainly four amino acid residues, PEVK. The extension of the PEVK segment is important for the elastic response of striated muscle to passive stretch and behaves mechanically as an entropic spring [88]. A PEVK fragment containing 16 PEVK modules and composed of 469 residues has been studied by CD spectroscopy. These studies indicate the likely presence of PPII structure within flexible joints. These
helices are supposed to be flexible and to make significant entropic contribution to elasticity.

**Table 2: Examples of proteins with PPII fold and the technique used for its identification**

<table>
<thead>
<tr>
<th>Name</th>
<th>Source</th>
<th>Function</th>
<th>Technique Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 High molecular weight (HMW) subunits</td>
<td>Wheat Glutenin</td>
<td>Cereal storage protein</td>
<td>CD</td>
</tr>
<tr>
<td>2 Antigenic peptide analog</td>
<td>foot-and-mouth disease virus (FMDV)</td>
<td></td>
<td>CD, X-ray</td>
</tr>
<tr>
<td>3 Amyloidogenic prefibrillar intermediate of human lysozyme</td>
<td>Human lysozyme</td>
<td></td>
<td>ROA</td>
</tr>
<tr>
<td>4 Collagen</td>
<td>Stiff connective fibre</td>
<td></td>
<td>CD, X-ray</td>
</tr>
<tr>
<td>5 PEVK segment of titin</td>
<td>Human</td>
<td>Imparts elasticity to the muscle sarcomere</td>
<td>CD</td>
</tr>
<tr>
<td>6 Abductin</td>
<td>bivalve shellfish “<em>Pecten jacobaeus</em>”</td>
<td>elastic function</td>
<td>CD</td>
</tr>
<tr>
<td>7 Elastin</td>
<td>Vertebrates</td>
<td>elastic function</td>
<td>CD</td>
</tr>
<tr>
<td>8 PXXP seq. peptide of the p85 subunit of P13-kinase</td>
<td>Vertebrates</td>
<td>Interaction with Fyn SH3 domain of nebulin</td>
<td>CD</td>
</tr>
<tr>
<td>9 Systemin</td>
<td>Plants</td>
<td>Oligopeptide hormone-like molecule</td>
<td>CD</td>
</tr>
<tr>
<td>10 Cysteine proteinase</td>
<td>Trypanosoma brucei</td>
<td>Interaction with other proteins, divides two domains</td>
<td>From sequence</td>
</tr>
</tbody>
</table>

**B. Protein protein interaction**

The molecular recognition processes consist of protein–protein interactions usually described in terms of ligand-acceptor complexes. Generally, the acceptor is a large
protein controlling the ligand, which is a small peptide sequence on a loop on the surface of a large protein. The X-ray or NMR data reported in the literature have revealed that the ligands are mostly in extended conformation when they are bound to their acceptors. In this conformation, the backbone atoms of peptide can form hydrogen bonds with the protein receptor at the interface of the peptide–protein complex [89]. Following are some examples of these complexes where the ligand is in PPII conformation:

1) The protein kinase inhibitor peptide that binds the catalytic subunit of the AMP-dependent protein kinase (PKA) [90]

2) The PXXP sequence peptide of the p85 subunit of P13- kinase that interacts with Fyn SH3 domain. Proline-rich sequences are the ligands of SH3 (Src homology 3) [91]

3) Profilin [92], which is an actin-binding protein involved in the dynamic turnover and restructuring of the actin cytoskeleton

4) Class II major histocompatibility complexes [93]

5) The repetitive sequence (VHLPPP)₈, belonging to the N-terminal domain of the γ zein protein of maize. It is responsible for targeting of this protein to the endoplasmic reticulum [94]

6) The PXY motif that is recognized by the WW domain (named for a conserved Trp-Trp motif) [95]

7) Enabled, VASP homology (EVH1, also known as WASP homology 1 or WH1) [96]. Recently the proline rich proteins are being targeted to develop new drugs in immune mediated disorders [97].

Recent studies [98] have demonstrated the ability to tune polyproline helix conformation and cis – trans isomerism in proline-rich sequences using aromatic electronic effects. This could have broad applications in biomolecular design, medicinal chemistry, biomaterials, and engineering.
4. References

193.
Chapter 1: Introduction