STRUCTURAL AND FUNCTIONAL STUDIES ON KINETICALLY STABLE SERINE PROTEASE FROM *NOCARDIOPSIS SP.* NCIM 5124

Thesis submitted to Savitribai Phule Pune University

For the degree of

DOCTOR OF PHILOSOPHY

IN

BIOTECHNOLOGY

By

SONALI B. ROHAMARE

Under the guidance of

DR. SUSHAMA M. GAIKWAD

DIVISION OF BIOCHEMICAL SCIENCES
NATIONAL CHEMICAL LABORATORY
PUNE -411 008 (INDIA)

SEPTEMBER 2014

For,

Ganesh, my parents and in laws...

CONTENTS

		Page No.
ACKNOWLEDGEMENTS		**i**
CERTIFICATE		**iv**
DECLARATION BY THE CANDIDATE		**v**
ABBREVIATIONS		**vi**
ABSTRACT		**vii**
LIST OF PUBLICATIONS		**x**
Chapter 1:	**Introduction**	**1-25**
	Introduction	1
	Serine proteases	2
	Serine proteases: classification, structure and function	4
	Kinetic stability	8
	Thermodynamic stability versus kinetic stability	9
	Alpha lytic protease (αLP): A model protein for studying kinetic stability	10
	Pro-region: as a folding catalyst	11
	Energetics of stabilitzation of native state of αLP	11
	Conformational rigidity of αLP	12
	Identification of kinetically stable proteins: general approaches	12
	The physical basis for kinetic stability	14
	Polyproline type II (PPII) helix	15
	Characteristics of PPII helix	16
	PPII in absence of proline	17
	Isodichroic Point	17
	Identification of PPII fold	17
	Lack of internal hydrogen bonding: stability of PPII helix	19
	Physiological implications of PPII helices in proteins	19

	Functions of PPII helix	19
	References	21
Chapter 2:	**Biochemical and biophysical characterization of NprotI**	**26-41**
	Summary	26
	Introduction	26
	Materials and Methods	27
	Results and Discussion	31
	References	41
Chapter 3:	**Role of PPII fold in imparting kinetic stability to NprotI**	**42-58**
	Summary	42
	Introduction	42
	Materials and Methods	44
	Results and Discussion	46
	References	57
Chapter 4:	**Acid stability of NprotI and differential modes of thermal denaturation at different pH**	**59-72**
	Summary	59
	Introduction	59
	Materials and Methods	60
	Results and Discussion	62
	References	72
Chapter 5:	**Cloning and expression of a serine protease from *Nocardiopsis sp.* NCIM 5124**	**73-95**
	Summary	73
	Introduction	74
	Materials and Methods	76
	Results and Discussion	80
	References	94

Chapter 6: Discussion **96-99**

Discussion 96

Conclusions 98

References 99

Acknowledgements

The journey becomes lot easier when you travel together. It is difficult to single handedly accomplish difficult tasks. My present work is no exception. I am delighted to have this opportunity to express my sincere thanks to all those who contributed in many ways to the success of this study and make it an unforgettable experience.

It is a joyous opportunity for me to express my gratitude and sincere thanks to my mentor Dr. Sushama M. Gaikwad, for her invaluable guidance, unending support and keen interest during the course of this investigation. She has given me the freedom to think and work; and I shall cherish my learning experience under her. I am very grateful for her patience, motivation and enthusiasm during course of the study. I would always be indebted for her support and guidance.

I am thankful to Dr. Dafydd Jones, Cardiff University, my research supervisor, during one year Commonwealth fellowship, for giving opportunity to work in his laboratory. I would be grateful for his valuable suggestions and help in learning new techniques.

I would also like to thank to Dr. Vaishali Javadekar, A. G. College, Pune, for all the help and valuable guidance at the start of my research work.

I am particularly thankful to Dr. J.K. Pal from the Savitribai Phule Pune University, Pune, for providing me the freedom to use the facilities of molecular biology in his laboratory.

I am grateful to Dr. M. J. Swamy, Dept. of Chemistry, University of Hyderabad for the help in carrying DSC studies under his guidance.

I am thankful to Dr. C.G. Suresh, NCL and Dr. Dr. Musti V. Krishnasastry from National Centre for Cell Science, Pune, for their keen interest and valuable suggestions during my entire work, and also for being members of my PhD committee.

I am thankful to Dr. Neelanjana Sengupta, Physical Chemistry Division, NCL, Pune for valuable discussions during the molecular dynamics work.

I am thankful to Dr. Sushim Kumar Gupta for his valuable suggestions in the cloning work.

I am grateful to Dr. R. Suresh Kumar, Dr. Sumedha S. Deshmukh and Dr. Siddharth H. Bhosale for their kind interest and technical help.

I would also like to extend my sincere thanks to Dr. Moneesha Fernandez, Organic Chemistry Division, NCL for availing the CD facility.

I am extremely thankful to Varsha, Ranu, Prathit and Tulika for their invaluable and timely help in putting my last chapter in proper shape. I have learned many new techniques from them and that experience is invaluable to me.

Special thanks are due to my dear colleagues Dr. Shashidhar, Dr. Ansary, Dr. Avinash, Dr. Madhurima, Sayli, Priya and Ekta for their help, continuous support and cooperation throughout my doctorate course. I am very fortunate to have seniors like Dr. Shashidhar, Dr. Ansary, Dr. Avinash and Dr. Madhurima. I am thankful for all their teaching and support. The time spent with them in lab will always be cherished. I have wonderful memories of the time spent outside the lab premises with Dr. Madhurima, Dr. Suprit, Dr. Poonam, Dr. Sameera, Sayli, Manas, Dr. Trupti and Pallavi.

During the course of PhD, I have worked at many different institutes and I am fortunate to befriend with many people at respective institutes. I take this opportunity to thank Ashwini, Kalpna Madam (TMV, Pune), Pavankumar, Gayathri (University of Hyderabad), Laura, Becky, Lisa, Rose, Sam, Adam, Sarunas, Andrew, Calvince, Dacia, Athraa (Cardiff University), Lucy, Jeny, Tintin, Grace, Roberta, Alyssia, Gemma, Angel, Tess, Nina (my housemates at 85 Colum Road, Cardiff), Narendra, Prajwal, Shashank, Hemant, Khushbu (friends in Cardiff), Dr. Radha, Varsha, Jemmy and Shilpa (Savitribai Phule Pune University). I would like to express my deep gratitude to my friends at NCL, Vrushali, Ruchi, Dr. Shadab, Dr. Asad, Dr. Varsha, Sana, Ravi, Vidya, Nishant, Urvashi, Deepak, Manu, Hemangi, Parth, Neha (3), Yojana, Ruby, Shweta and late Dr. Vishnu for all the help and support.

My old friends, Alpna, Manoj, Vrushali, Preeti, Meenakshi and Ashwini always encouraged me during the difficult times.

I would like to extend my deep gratitude to my school teachers, Sakhare Madam and Sir, Uma Raman Madam, Kalyan Raman Sir and Galgate Madam; who were always in touch and kept faith in me. They have always inspired me.

I also would like to thank scientists, staff of Biochemical Sciences Division, NCL, helping me directly or indirectly during course of my stay at the division.

A special thank and love to my parents (Mummy and Papa), in laws (Bhau and Akka), beloved grandmothers, my sisters (Arti, Puja), nephews, sister in law (Renuka) and my

family who have been constant inspiration for me and this work would not have been possible without their constant support and sacrifices.

I find no words to express gratitude to my beloved husband Ganesh, who has become my best friend now. It is because of his understanding and patient nature, I could achieve this difficult task. Ganesh has been a true and great supporter and has unconditionally loved me during my good and bad times. He was and is my pillar of strength during all the difficult times.

I thank the Director, National Chemical Laboratory, and the Head, Division of Biochemical Sciences, for permitting me to carry my PhD work in the Division of Biochemical Sciences, NCL and to submit this work in the form of the thesis. I am also thankful to the library staff, administrative and technical staff for their help during the course of my study. The financial assistance from the Council of Scientific and Industrial Research, India and Commonwealth Commission, UK is also duly acknowledged.

Finally, I would like to thank to all those who have helped directly or indirectly during my journey to date.

Sonali

CERTIFICATE

Certified that the work incorporated in the thesis entitled **"Structural and functional studies on kinetically stable serine protease from *Nocardiopsis sp*. NCIM 5124"** submitted by Ms. Sonali B. Rohamare was carried out under my supervision. Such materials has been obtained from other sources has been duly acknowledged in the thesis.

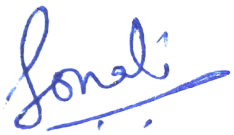
Dr. Sushama M. Gaikwad

Research Guide

Date..Sept.1, 2014

DECLARATION OF THE CANDIDATE

I declare that the thesis entitled **"Structural and functional studies on kinetically stable serine protease from *Nocardiopsis sp.* NCIM 5124"** submitted by me for the degree of Doctor of Philosophy is the record of work carried out by me during the period from **16th July, 2009 to 28th May, 2014** under the guidance of **Dr. Sushama M. Gaikwad** and has not formed the basis for the award of any degree, diploma, associateship, fellowship, titles in this or any other University or other institute of higher learning. I further declare that the material obtained from other sources has been duly acknowledged in the thesis.



Signature of the Candidate

Date .Sept 1, 2014

Sonali B. Rohamare

LIST OF ABBREVIATIONS USED

ANS	1-anilino-8-naphthalenesulfonate
Arg	arginine
BLAST	Basic Local Alignment Sequence Tool
CanPI	*Capsicum annuum* Pin-II Protease Inhibitor
CD	circular dichroism
CTAB	cetyl trimethyl ammonium bromide
cDNA	complementary DNA
EDTA	ethylene diamine tetra acetic acid
GdnHCl	Gaunidine hydrochloride
GuSCN	Gaunidine thiocyante
His	histidine
IPTG	isopropyl-β-D-thiogalactoside
K	Kelvin
K_m	Michaelis-Menten constant
K_s	Static quenching constant
K_{sv}	Stern-Volmer (dynamic) quenching constant
kDa	kilo dalton
LB	Luria Bertani Broth
MG	molten globule
MRE	Mean Residue Ellipticity
NRMSD	normalized root mean square deviation
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
pKa	Acid dissociation constant
PPII	Polyproline II fold
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
Ser	serine
Taq polymerase	*Thermus aquaticus* polymerase
Trp	tryptophan
V_{max}	maximum rate of an enzyme catalyzed reaction

ABSTRACT

Chapter 1: Introduction

Chapter 2: Biochemical and biophysical characterization of NprotI

Presence of polyproline type II (PPII) helix was speculated in NprotI based on characteristic far UV CD spectrum. The microenvironment of single tryptophan residue in Nocardiopsis sp. serine protease (NprotI) was studied using steady state and time-resolved fluorescence. The emission maximum of intrinsic fluorescence was observed at 353 nm with excitation at 295 nm indicating tryptophan to be solvent exposed. Upon denaturation with 6 M guanidinum thiocyanate (GuSCN) the emission maxima was shifted to 360 nm. Solute quenching studies were performed with neutral (acrylamide) and ionic (I- and Cs+) quenchers to probe the exposure and accessibility of tryptophan residue of the protein. Maximum quenching was observed with acrylamide. In native state quenching was not observed with Cs+ indicating presence of positively charged environment surrounding tryptophan. However; in denatured protein, quenching was observed with Cs+, indicating charge reorientation after denaturation. No quenching was observed with Cs+ even at pH 1.0 or 10.0; while at acidic pH, a higher rate of quenching was observed with KI. This indicated presence of more positive charge surrounding tryptophan at acidic pH. In time resolved fluorescence measurements, the fluorescence decay curves could be best fitted to monoexponential pattern with lifetimes of 5.13 ns for NprotI indicating one conformer of the tryptophan. Chemical modification studies with phenyl glyoxal suggested presence of Arg near the active site of the enzyme. No inhibition was seen with soyabean trypsin and limabean inhibitors, while, CanPI uncompetitively inhibited NprotI. Various salts from Hofmeister series were shown to decrease the activity and PPII content of NprotI.

Chapter 3: Role of PPII fold in imparting kinetic stability to NprotI

In the present chapter, the presence of polyproline II (PPII) fold was confirmed in NprotI based on structural transitions of the enzyme in presence of GdnHCl and a distinct isodichroic point (far UV CD spectrum) in chemical and thermal denaturation. Also, it’s role in imparting kinetic stability to NprotI was studied. The functional and conformational transitions of the enzyme were studied under various denaturing

conditions. Enzymatic activity of NprotI was stable in the vicinity of GdnHCl upto 6.0 M concentration, organic solvents viz. methanol, ethanol, propanol (all 90 %v/v), acetonitrile (75 %v/v) and proteases like trypsin, chymotrypsin and proteinase K (NprotI:protease 10:1). The NprotI seems to be a kinetically stable protease with a high energy barrier between folded and unfolded states. Also, an enhancement in the activity of the enzyme was observed in 1 M GdnHCl upto 8 h, in organic solvents (75 %v/v) for 72 h and in presence of proteolytic enzymes. The polyproline fold remained unaltered or became more prominent under the above mentioned conditions. However, it gradually diminished during thermal denaturation above 60 °C. Thermal transition studies by differential scanning calorimetry (DSC) showed scan rate dependence as well as irreversibility of denaturation, the properties characteristic of kinetically stable proteins.

Chapter 4: Acid stability of NprotI and Differential modes of thermal denaturation at different pH conditions

The kinetically stable alkaline serine protease from Nocardiopsis sp.; NprotI, possessing polyproline II fold (PPII) was characterized for its pH stability using proteolytic assay, fluorescence and CD spectroscopy, and DSC. NprotI is optimally active at pH 10.0 and most stable at pH 5.0. However, the enzyme was functionally stable when incubated alone at pH 1.0; even after 24 h, while immediate and drastic loss of the activity was observed at pH 10.0. The activity of NprotI incubated at pH 1.0 and 3.0 was enhanced at higher temperature (50 °C- 60 °C). NprotI maintained the overall PPII fold in broad pH range as seen using far UV CD spectroscopy. The PPII fold of NprotI incubated at pH 1.0 remained fairly intact up to 70 °C. Based on the isodichroic point and Tm values revealed by secondary structural transitions, different modes of thermal denaturation at pH 1.0, 5.0 and 10.0 were observed. DSC studies of NprotI in acidic pH range showed Tm values in the range of 74-76 °C while significant decrease in Tm (63.8 °C) was observed at pH 10.0. NprotI could be chemically denatured at pH 5.0 only with guanidine thiocynate. Acid tolerant and thermostable NprotI can serve as a potential candidate for biotechnological applications.

Chapter 5: Cloning and expression of serine proteases from *Nocardiopsis sp.* NCIM 5124

A serine protease (named as ‘N.protease’), from a marine isolate of Nocardiopsis sp., was cloned, expressed in E.coli and investigated for its potential kinetic stability. Expression was carried out using two vectors, pET-22b (+) and pET-39b (+) and comparisons were made based on proper folding and soluble expression of the protein. Studies into the role of incubation temperature for E.coli growth and inducer concentration on solubility of expressed protein were undertaken. pET-39b (+) was found to be a better vector for the soluble expression of this disulfide bonds containing protease. In silico studies were carried out for N.protease. Homology modelling of N.protease suggested it to be a part of PA clan of proteases. High temperature MD simulations were performed on N.protease to study its unfolding behavior and comparisons were made with alpha lytic protease (αLP), a model kinetically stable protein. A novel approach to study cooperativity of protein unfolding was undertaken, wherein ‘P’ value analysis based on φ and ψ values of the protein was performed. Gradual transition in P value analysis for N.protease suggested it to be comparatively less kinetically stable than αLP. Present study holds significance as the non-strepetomycete actinomycetes group is least explored and holds a promise for industrially important enzymes with exceptional stabilities.

Chapter 6: Discussion

LIST OF PUBLICATIONS:

1. **Sonali B. Rohamare**, Vaishali Dixit, Pavan Kumar Nareddy, D. Sivaramakrishna, Musti J. Swamy, Sushama M. Gaikwad. (2013) Polyproline fold – in imparting kinetic stability to an alkaline serine endopeptidase. Biochimica Biophysica Acta (proteins and Proteomics) 1834:708-716.

2. **Sonali B. Rohamare,** Sushama M. Gaikwad. Tryptophan environment and functional characterization of a kinetically stable serine protease containing polyproline II fold. Journal of Fluorescence. (Accepted for publication) DOI: 10.1007/s10895-014-1445-5.

3. **Sonali B. Rohamare**, Vaishali Dixit, Pavan Kumar Nareddy, Musti J. Swamy, Sushama M. Gaikwad. Acid stability of the kinetically stable alkaline serine protease possessing polyproline II fold. (Manuscript under review)

4. **Sonali B. Rohamare**, Sushama M. Gaikwad, Dafydd Jones, Varsha Bhavnani, Jayanta Pal, Ranu Rawat, Prathit Chaterjee. Cloning, expression and *in silico* studies of a serine protease from *Nocardiopsis sp.* NCIM 5124 (Manuscript communicated)

Chapter 1

Introduction

1. Introduction

With the advent 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 life style. Hence, pharmaceutical companies are always trying hard to bring new drugs in market. Traditionally, there was 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-streptomycetes 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.*

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

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.

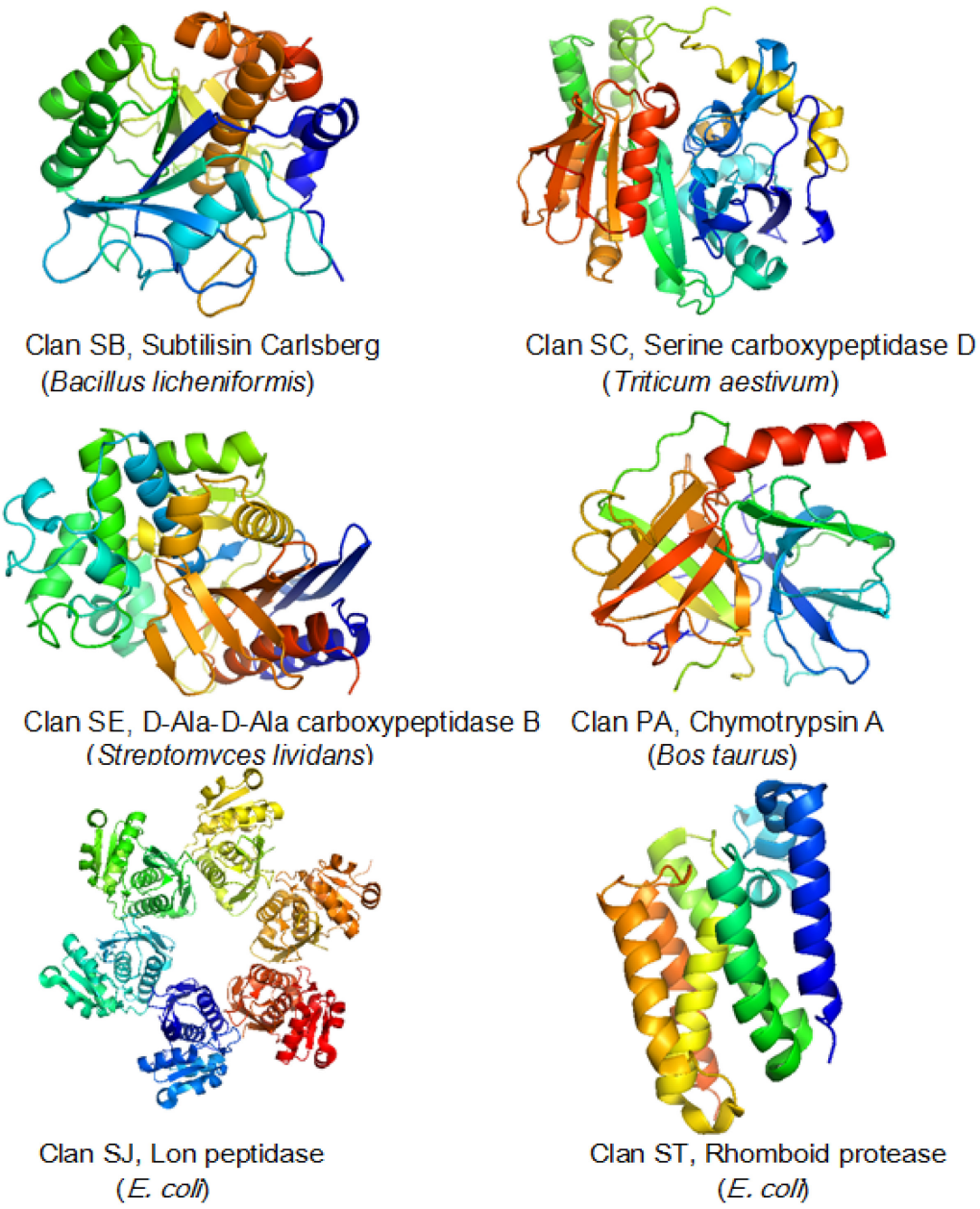


Figure 1**: **Crystal structures of various proteases from different clans. The respective PDB IDs have been mentioned in table 1.

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:

$N \leftrightarrow U$

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

$K = U/N$

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

$\Delta G = -RT \ln K$

ΔG 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).

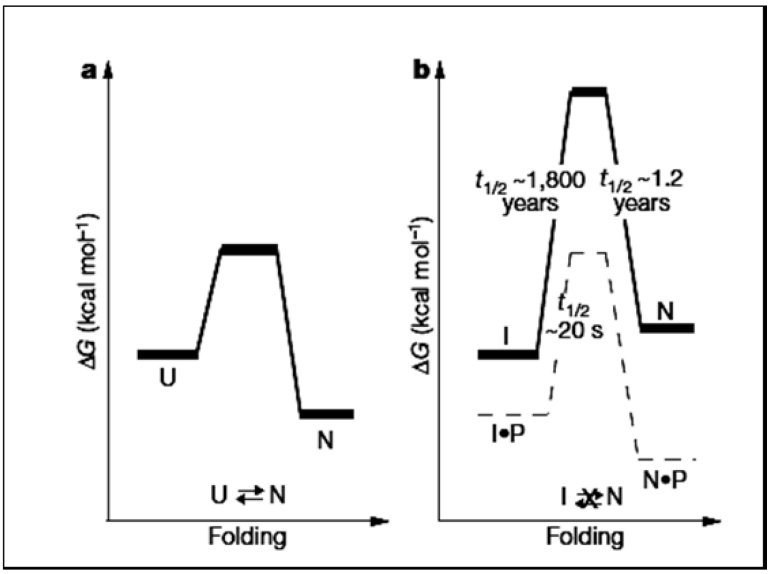


***Figure 2**: An energy diagram showing difference between a typical protein and a kinetically stable protein; alpha lytic protease. (Adapted from Jaswal et al [37]).*

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 stabilitzation 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.

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.

$$N \longleftrightarrow U \longrightarrow F$$

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 $U \longrightarrow F$ is fast enough, in such a way that any molecule U formed is immediately converted to F, a two-state kinetic process applies [28],

$$N \longrightarrow F$$

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, -

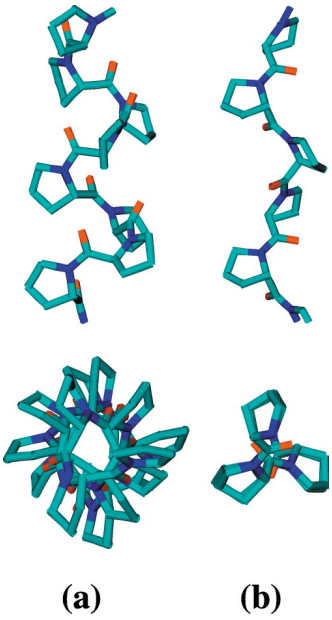


Figure 3: ***Side and top views of two ideal helical conformations of a polyproline peptide*** (a) polyproline I helix (PPI), a right-handed helix with backbone dihedral angles of (φ,ψ,ω)=(−75°,160°,0°) ; (b) polyproline II helix (PPII, polypro II), a left-handed helix with backbone dihedral angles of (φ,ψ,ω)=(−75°, 146°, 180°), 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 (Φ, Ψ) 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 α helices; also hydrogen bonding potential of neighboring residues in polyproline conformation is not satisfied in a manner analogous to β 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 α helices and β 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].

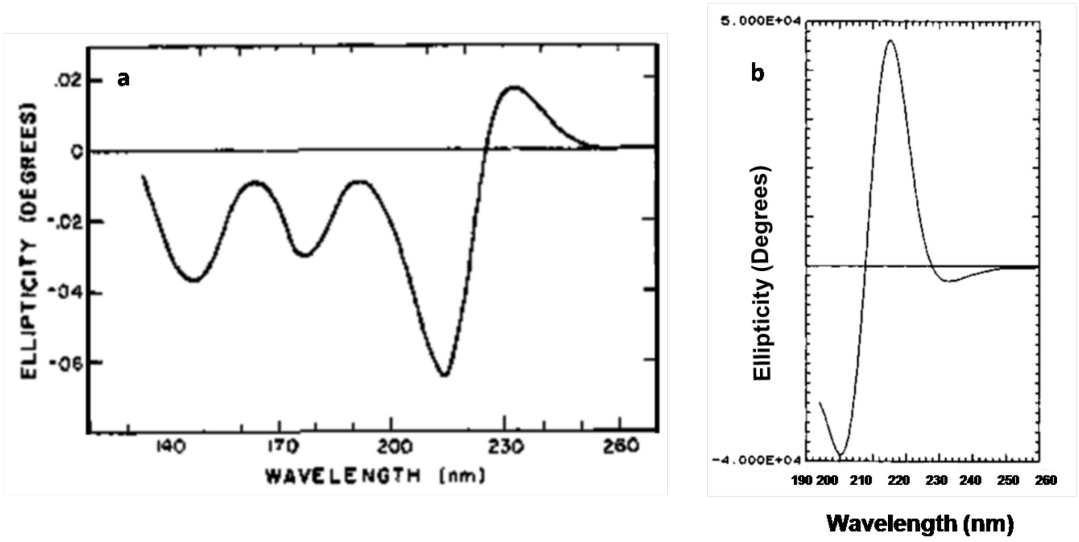


Figure 4: CD spectrum of polyproline peptide **a)** PPII conformation in solution [75] **b)** PPI in solution [76].

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].

1.3.7Functions 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*

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

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

1. N D Rawlings, A J Barrett. *Nucleic Acids Res.* **1999**, 27(1): 325–331.
2. A J Barrett, N D Rawlings, E A O'Brien. *J Struc Biol.* **2001**, 134: 95–102.
3. N D Rawlings, M Waller, A J Barrett, A Bateman. *Nucleic Acids Res.* **2014**, 42:D503–D509.
4. M J Page, E Di Cera. *Cell Mol Life Sci.* **2008**, 65:1220 – 1236.
5. J M Sanchez-Ruiz. *Biophys Chem.* **2010**, 148: 1–15.
6. J L Sohl, S S Jaswal, and D A Agard. *Nature.* **1998,** 395: 817–819.
7. S M Truhlar, E L Cunningham, D A Agard. *Protein Sci.* **2004**, 13: 381–390.
8. G I Makhatadze, P L Privalov. *Adv Protein Chem.* **1995,** 47: 307–425.
9. D Baker, A K Shiau, D A Agard. *Curr Opin Cell Biol.* **1993,** 5: 966–970.
10. J Eder, A R Fersht. *Mol Microbiol*. **1995**, 16: 609–614.
11. E L Cunningham, D A Agard. *Protein Sci.* **2004,** 13: 325–331.
12. D Baker, J L Sohl, D A Agard. *Nature.* **1992,** 356: 263– 265.
13. J L Silen, D A Agard. *Nature.* **1989,** 341: 462–464.
14. M Fujinaga, L T Delbaere, G D Brayer, M N James. *J. Mol. Biol.* **1985,** 184: 479-502.
15. R J Peters, A K Shiau, J L Sohl, D E Anderson, G Tang, J L Silen, D A Agard *Biochemistry*. **1998,** 37: 12058–12067.
16. J A D'Aquino, J Gomez, V J Hilser, K H Lee, L M Amzel, E Freire. *Proteins*. **1996,** 25:143–156.
17. A C Clark, S W Raso, J F Sinclair, M M Ziegler, A F Chaffotte, T O Baldwin. *Biochemistry*. **1997,** 36: 1891–1899.
18. J Chen, S A Wharton, W Weissenhorn, L J Calder, F M Hughson J J Skehel, D C Wiley. *Proc Natl Acad Sci.***1995**, 92: 12205–12209.
19. P M Harrison, P Bamborough, V Daggett, S B Prusiner, F E Cohen. *Curr Opin Struct Biol.* **1997,** 7: 53–59.
20. I Z Wang, J Mottonen, E J Goldsmith. *Biochemistry*. **1996,** 35: 16443–16448.
21. J L Sohl, S S Jaswal, D A Agard. *Nature.* **1998,** 395 817-819.
22. J F Sinclair, M M Ziegler, T O Baldwin. *Nat Struct Biol.* **1994**, 1: 320–326.
23. I V Baskakov, G Legname, S B Prusiner, F E Cohen. *J Biol Chem.* **2001**, 276: 19687–19690.
24. P Forrer, C Chang, D Ott, A Wlodawer, A Plückthun. *J Mol Biol.* **2004**, 344: 179–

193.

25. S L Flaugh, I A Mills, J King. *J Biol Chem.* **2006**, 281: 30782–30793.
26. I M P del Pino, B Ibarra-Molero, J M Sanchez-Ruiz. *Proteins.* **2000**, 40: 58–70.
27. E. Freire, W W van Osdol, O L Mayorga, J M Sanchez-Ruiz. *Annu Rev Biophys Biophys Chem.* **1990,** 19:159–188.
28. J M Sanchez-Ruiz. *Biophys J.* **1992,** 61: 921–935.
29. K Xia, M Manning, H Hesham, Q Lin, C Bystroff, W Colon. *Proc Natl Acad Sci.* 2007, 104: 17329–17334.
30. C Park, S Zhou, J Gilmore and S Marqusee. *J Mol Biol.* **2007**, 368: 1426–1437.
31. B A Kelch, K P Eagen, F P Erciyas, E L Humphris, A R Thomason, S Mitsuiki, D A Agard. *J Mol Biol.* **2007,** 368: 870–883.
32. K Linderstrom-Lang. *Nature.* **1938**, 142: 996.
33. K. Linderstrom-Lang, *Cold Spring Harbor Symp Quant Biol.* **1950**, 14: 117–126.
34. M Ottesen. *Annu Rev Biochem.* **1967**, 36: 55–76.
35. C B Anfinsen, H A Scheraga. *Advan Protein Chem.* **1975**, 29:205–300.
36. A Fontana, P Polverino de Laureto, V De Filippis, E Scaramella and M Zambonin. *Fold Des.* **1997**, 2: R17–R26.
37. S S Jaswal, J L Sohl, J H Davis, D A Agard. *Nature.* **2002**, 415:343–346.
38. M Machius, N Declerck, R Huber, G Wiegand. *J Biol Chem.* **2003**, 278: 11546-11553.
39. J Mansfeld, G Vriend, B W Dijkstra, O R Veltman, B Van den Burg, G Venema, R Ulbrich-Hofmann, G H Eijsink. *J Biol Chem.* **1997**, 272: 11152-11156.
40. I Pozdnyakova, J Guidry, P Wittung-Stafshede. *Arch Biochem Biophys.* **2001**, 390: 146-148.
41. A W Rietveld, S T Ferreira. *Biochemistry.* **1998**, 37: 933-937.
42. S Solis-Mendiola, L H Gutierrez-Gonzalez, A Arroyo-Reyna, J Padilla-Zuniga, A Rojo-Dominguez, A Hernandez-Arana. *Biochim Biophys Acta.* **1998**, 1388: 363-372.
43. R Jaenicke, G Bojm. *Curr Opin Struct Biol.* **1998**, 8: 738-748.
44. J K Kaushik, K Ogasahara, K Yutani. *J Mol Biol.* **2002**, 316: 991-1003.
45. S Cavagnero, D A Debe, Z H Zhou, M W W Adams, S I Chan. *Biochemistry.* **1998**, 37: 3369-3376.
46. L Pauling, R B Corey. *J Am Chem Soc.* **1950**, 72: 5349.

47. L Pauling, R B Corey, H R Branson. *Proc Natl Acad Sci.* **1951**, 37: 205–211.
48. L Pauling, R B Corey *Proc Natl Acad Sci.* **1951**, 37: 251–256.
49. J Makowska, S Rodziewicz-Motowidlo, K Baginska, J A Vila, A Liwo, L Chmurzyński, H A Scheraga. *Proc Natl Acad Sci.* **2006**, 103: 1744–1749.
50. L Pauling, RB Corey. *Proc Natl Acad Sci.* **1951**, 37: 272–281.
51. P M Cowan, S McGavin, A C North. *Nature.* **1955**, 176: 1062–1064.
52. M L Tiffany, S Krimm. *Biopolymers*. **1968**, 6:1379–1382.
53. J C Ferreon, V J Hilser. *Protein Sci.* **2003**, 12:447–457.
54. S Arnott, S D Dover. *Acta Crystallogr B*. **1968**, 24: 599–601.
55. V Sasisekharan. *Acta Crystallogr.* **1959**, 12: 897–903.
56. M Moradi, V Babin, C Roland, C Sagui. *J of Chem Phys.* **2010**, 133: 125104-18.
57. M V Cubellis, F Caillez, T L Blundell, S C Lovell. *Proteins.* **2005**, 58: 880–892.
58. W Kabsch, C Sander. *Biopolymers*. **1983**, 22:2577–2637.
59. D Frishman, P Argos. *Proteins*. **1995**, 23:566–579.
60. Y Mansiaux, A P Joseph, J-C Gelly, A G de Brevern. *PLoS ONE.* **2011**, 6(3): e18401.
61. A A Adzhubei, M J Sternberg. *J Mol Biol.* **1993**, 229: 472–493.
62. T P Creamer. *Proteins.* **1998**, 33: 218–226.
63. B J Stapley, T P Creamer. *Protein Sci.* **1999**, 8: 587–595.
64. T P Creamer, M N Campbell. *Adv Protein Chem.* **2002**, 62: 263–282.
65. B W Chellgren, T P Creamer. *Biochemistry.* **2004**, 43: 5864–5869.
66. B W Chellgren, A F Miller, T P Creamer. *J Mol Biol*. **2006**, 361: 362–371.
67. M L Tiffany, S Krimm. *Biopolymers*. **1969**, 8:347-359.
68. W A Hiltner, A J Hopfinger, A G Walton. *J Am Chem Soc.* **1972**, 94:4324-4327.
69. M G Paterlini, T B Freedman, L A Nafie. *Biopolymers.* **1986**, 25: 1751- 1765.
70. S-H Park, W Shalongo, E Stellwagen. *Protein Sci.* **1997**, 6:1694-1700.
71. A F Drake, G Siligardi, W A Gibbons. *Biophy Chem.* **1988**, 31: 143-146.
72. R W Woody. *J Am Chem Soc*. **2009**, 131: 8234–8245
73. F Zhu, J Kapitan, G E Tranter, P D A Pudney, N W Isaacs, L Hecht, L D Barron. *Proteins-Struct Funct Bioinform.* **2008**, 70: 823–833.
74. P Bour, J Kubelka, T A Keiderling. *Biopolymers.* **2002**, 65: 45–59.
75. M A Young, E S Pysh. *J Am Chem Soc.* **1975**, 97(18): 5100-5103.
76. F Rabanal, M D Ludevid, M Pons, E Glralt. *Biopolymers.* **1993**, 33: 1019-1028.

77. R W Woody. *J Am Chem Soc.* **2009**, 131: 8234–8245.
78. S M Gilbert, N Wellner, P S Belton, J A Greenfield, G Siligardi, P R Shewry, A S Tatham. *Biochim Biophys Acta.* **2000**, 1479:135–146.
79. S J Whittington, T P Creamer. *Biochemistry.* **2003**, 42: 14690–14695.
80. Z Liu, K Chen, A Ng, Z Shi, R W Woody, et al. *J Am Chem Soc.* **2004**, 126:15141–15150.
81. A Kentsis, M Mezei, T Gindin, R Osman. *Proteins.* **2004**, 55: 493–501.
82. M Mezei, P J Fleming, R Srinivasan, G D Rose. *Proteins*. **2004**, 55: 502–507.
83. N Sreerama, R W Woody. *Proteins.* **1999**, 36: 400–406.
84. B K Kay, M P Williamson, M Sudol. *FASEB J.* **2000**, 14:231–241.
85. E W Blanch, L A Morozova-Roche, D A Cochran, A J Doig, L Hecht, et al. *J Mol Biol*. **2000**, 301: 553–563.
86. F Eker, K Griebenow, R Schweitzer-Stenner. *Biochemistry.* **2004**, 43: 6893–6898.
87. J M Hicks, V L Hsu. *Proteins.* **2004**, 55: 330–338.
88. G Gutierrez-Cruz, A H Van Heerden, K Wang. *J Biol Chem*. **2001**, 276:7442–7449.
89. G Siligardi, A F Drake. *Biopolymers*. **1995**, 37:281–292.
90. D R Knighton, J Zheng, L F Ten Eyck, N Xoung, S S Taylor, I M Sowadski. *Science.* **1991**, 253:414–420.
91. B J Mayer, M J Eck. *Curr Biol.* **1995**, 5: 364–367.
92. M Tanaka, H Shibata. *Eur J Biochem*. **1985**, 151:291–297.
93. T S Jardetsky, J H Brown, J C Gorga, L J Stern, R G Urban, J L Strominger, D C Wiley. *Proc Natl Acad Sci*. **1996**, 93:734–738.
94. M Torrent, M I Geli, M Ruiz-Avila, J M Canals, P Puigdomenech, M D Ludevid. *Planta*. **1994**, 192:512–518.
95. M J Macias, M Hyvonen, E Baraldi, J Schultz, M Sudol, M Saraste, H Oschkinat. *Nature.* **1996**, 382: 646–649.
96. K Niebuhr, F Ebel, R Frank, M Reinhard, E Domann, U D Carl, U Walter, F B Gertler, J Wehland, T Chakraborty. *EMBO J.* **1997**, 16: 5433–5444.
97. M Srinivasan, A K Dunker, *Int J Pep.* **2012**, 2012: 14 pages.
98. A K Pandey, K M Thomas, C R Forbes, N J Zondlo. *Biochemistry.* **2014**, 53: 5307−5314.

Chapter 2

Biochemical and biophysical characterization of NprotI

Tryptophan environment and functional characterization of a kinetically stable serine protease containing polyproline II fold. Journal of Fluorescence. DOI: 10.1007/s10895-014-1445-5.

Summary

Presence of polyproline type II (PPII) helix was speculated in NprotI based on characteristic far UV CD spectrum. The microenvironment of single tryptophan residue in *Nocardiopsis sp.* serine protease (NprotI) was studied using steady state and time-resolved fluorescence. The emission maximum of intrinsic fluorescence was observed at 353 nm with excitation at 295 nm indicating tryptophan to be solvent exposed. Upon denaturation with 6 M guanidinum thiocyanate (GuSCN) the emission maxima was shifted to 360 nm. Solute quenching studies were performed with neutral (acrylamide) and ionic (I^- and Cs^+) quenchers to probe the exposure and accessibility of tryptophan residue of the protein. Maximum quenching was observed with acrylamide. In native state quenching was not observed with Cs^+ indicating presence of positively charged environment surrounding tryptophan. However; in denatured protein, quenching was observed with Cs^+, indicating charge reorientation after denaturation. No quenching was observed with Cs^+ even at pH 1.0 or 10.0; while at acidic pH, a higher rate of quenching was observed with KI. This indicated presence of more positive charge surrounding tryptophan at acidic pH. In time resolved fluorescence measurements, the fluorescence decay curves could be best fitted to monoexponential pattern with lifetimes of 5.13 ns for NprotI indicating one conformer of the tryptophan. Chemical modification studies with phenyl glyoxal suggested presence of Arg near the active site of the enzyme. No inhibition was seen with soyabean trypsin and limabean inhibitors, while, CanPI 7 uncompetitively inhibited NprotI. Various salts from Hofmeister series were shown to decrease the activity and PPII content of NprotI.

1. Introduction

Proteases are the enzymes needed and synthesized by almost all forms of life. They have evolved multiple times, and a completely different catalytic mechanism is used by different class of proteases to perform the same reaction. Serine proteases are extensively studied proteases among different classes of proteases. The interesting property for these enzymes is that although they use same catalytic triad for catalysis, their sequences can be totally different. The serine proteases do vary on the basis of their substrate binding pocket. So, some have broad substrate specificity while some can cleave only particular peptide bond in particular protein.

Extracellular bacterial proteases can have unique structural features in order to survive in

harsh environments. These proteases are generally very stable to various denaturing conditions and are called as kinetically stable proteases [1, 2]. In the present study, we have explored biochemical and biophysical properties of a kinetically stable serine protease from *Nocardiopsis sp.* NCIM 5124 (NprotI). Protease I from *Nocardiopsis* sp. (NprotI) is an alkaline serine protease with a molecular mass of 21 kDa, optimum temperature of 60 °C and optimum pH of 10.0 for activity [3].

Fluorescence quenching of indole by adding solutes have provided valuable information regarding the structure and dynamics of proteins in solution [4, 5]. We report the structural studies of the NprotI by using steady-state and time resolved fluorescence and CD spectroscopy. The protein has been characterized with respect to tryptophan environment in native, denatured and different pH conditions. Determination of pKa of the amino acids present at active site has been done. Also, amino acid residues probably involved at the active site were studied performing chemical modification studies. Effect of different salts from Hofmeister series was studied with respect to activity and PPII content of NprotI.

2. Materials and Methods

2.1 Materials

Soyabean trypsin and lima bean inhibitor were procured from Sigma (USA). CanPI 7 was a gift from Dr. Ashok Giri, NCL, Pune. All other reagents, buffer compounds used were of analytical grade. Solutions prepared for spectroscopic measurements were in MilliQ water.

2.2 Production of the enzyme

The organism was isolated from an oil contaminated marine site near Mumbai harbor (India). The protocol used for the production and purification of NprotI was as described earlier [6]. Briefly, the culture broth of *Nocardiopsis sp.* NCIM 5124 was obtained by fermentation in a medium containing 1 % starch, 1 % casein, 0.1 % K_2HPO_4, 1 % Na_2CO_3, 0.2 % glucose, pH 10.0 after incubation for 108 h at 30 °C and 200 rpm. Cell pellet was separated after centrifugation of crude culture broth at 10,000 rpm for 10 min and the cell free supernatant was subjected to purification procedures. Protease activity was assayed using casein as a substrate. Protein concentration in column eluents was monitored in terms

of absorbance at 280 nm and total protein in pooled fractions was determined by the method of Lowry et al with bovine serum albumin as a standard.

2.3 Purification of NprotI

The pH of the crude culture filtrate was adjusted to 5.0 with 1 M acetic acid. 1 M NaCl was added to the filtrate, before adjusting the pH, to minimize the co-precipitation of the protease. The supernatant obtained after centrifuging this precipitate was concentrated to 1/10 of its original volume by ultra-filtration under reducing atmosphere, using a YM-3 membrane (Amicon). The concentrate was dialyzed against 20 mM Na-acetate buffer, pH 5.0 and loaded on a CM-Sephadex column equilibrated with the same buffer. NprotI was found to retain on the column. The column was washed with 20 mM Na-acetate buffer, pH 5.0 containing 0.2 M NaCl and the adsorbed protein was eluted with a linear gradient of 0.2 - 1.0 M NaCl. The active fractions were pooled and dialyzed against 20 mM Na-carbonate buffer, pH 9.0 and re-chromatographed on CM- Sephadex at pH 9.0. The adsorbed protein eluted as a single peak in 50 mM NaCl in the same buffer.

2.4 Enzyme assay

Protease activity was determined by incubating 3 µg of the enzyme in 300 µl of 1 % casein (substrate) at pH 10.0 at 60 °C for 20 min as described by Dixit *et al* [6]. One unit of protease activity is defined as the amount of enzyme which releases 1 µmol of tyrosine per minute in the assay conditions.

2.5 Steady state fluorescence study

Intrinsic fluorescence of the enzyme was measured using a Perkin-Elmer Luminescence spectrometer LS50B connected to a Julabo F20 water bath. The protein solution was excited at 295 nm and the emission was recorded in the range of wavelength 300-400 nm at 30 °C. The slit widths for the excitation and emission were set at 7.0 nm, and the spectra were recorded at 100 nm/min. To eliminate the background emission, the signal produced by the buffer solution was subtracted.

2.6 Steady state fluorescence quenching

Fluorescence titrations were carried out by adding 3–5 µl of acrylamide (5 M), potassium iodide (5 M) and cesium chloride (5 M) to the protein sample. Fluorescence intensity was recorded after each addition. The iodide solution contained sodium thiosulfate (200 µM) to suppress triiodate formation. The excitation wavelength was set at 295 nm; the emission spectra were recorded in the range 300 to 400 nm with both the slit widths as 7 nm at a scan speed of 100 nm/min. To eliminate contribution from background emission, the signal produced by buffer solution was subtracted. The steady state fluorescence quenching was done for native (pH 5.0); GuSCN denatured NprotI, as well as, with NprotI at pH 1.0 and 10.0. Following buffers were used in 20 mM concentration; pH 1.0: glycine-HCl, pH 5.0: sodium acetate, pH 10.0: sodium carbonate-bicarbonate.

2.7 Time-resolved fluorescence study

Fluorescence lifetime measurements were carried out on Edinburgh Instruments' FLS-920 single photon counting spectrofluorimeter. A pico second pulsed light emitting diode of wavelength 296.8 nm, pulse width 747.8 ps and bandwidth 10.4 nm was used as excitation source and a Synchronization photomultiplier was used to detect the fluorescence. The diluted Ludox solution was used for measuring Instrument Response Function (IRF). NprotI (1 mg/ml) was excited at 295 nm and emission was recorded at 353 nm. Slit widths of 15 nm each were used on the excitation and emission monochromators. The resultant decay curves were analyzed by a reconvolution fitting program supplied by Edinburgh instruments.

2.8 Circular dichroism (CD) measurements

The CD spectra of the enzyme were recorded on a J-715 Spectropolarimeter with a PTC 343 Peltier unit (Jasco, Tokyo, Japan) at 25 °C in a quartz cuvette. Each CD spectrum was accumulated from five scans at 100 nm/min with a 1 nm slit width and a time constant of 1 s for a nominal resolution of 0.5 nm. Far UV CD spectra of the enzyme (250 µg/ml) were collected in the wavelength range of 200–250 nm using a cell of path length 0.1 cm for monitoring the secondary structure. The tertiary structure of the enzyme (800 µg/ml) was monitored with near UV CD spectra in the wavelength range of 250–300 nm using cell of

path length 1 cm. All spectra were corrected for buffer contributions and observed values were converted to mean residue ellipticity (MRE) in deg cm^2 $dmol^{-1}$ defined as

$MRE = M\theta_\lambda/10dcr$

Where, M is the molecular weight of the protein, θ_λ is CD in millidegree, d is the path length in cm, c is the protein concentration in mg/ ml and r is the average number of amino acid residues in the protein. The relative content of various secondary structure elements was calculated by using CD pro software (http://lamar.colostate.edu/~sreeram/CDPro/main.html). Low NRMSD values were observed for analysis with CONTINLL and SELCON.

2.9 Modification of serine residues with phenylmethylsulfonyl fluoride (PMSF)

The reaction mixture containing enzyme in 50 mM sodium acetate buffer, pH 5.0 and 25 to 150 µM of PMSF was incubated at 25 °C for 2 h. Aliquots were removed at different time intervals and the residual activities were determined under standard assay conditions. Enzyme sample incubated in the absence of PMSF served as control.

2.10 Determination of pKa of amino acids at the active site residues

Activity for the enzyme was checked at various pH (pH 7-12) conditions and substrate concentrations. K_m and V_{max} values were calculated and a plot of log (V_{max}/K_m) vs pH concentration. The pKa of the amino acids at the active site was obtained from the tangents of bell shaped curve.

2.11 Chemical modification with phenylglyoxal (PG)

The enzyme in 50 mM Tris-HCl buffer, pH 8.0, was incubated with varying concentrations of PG at 25 °C. Aliquots were removed at suitable intervals and the residual activity was determined under standard assay conditions. Substrate protection and time dependence was also studied in presence of phenyl glyoxal.

2.12 Interaction with soyabean trypsin, lima bean and CanPI inhibitors

The enzyme was preincubated with different concentrations of inhibitors for 20 min at pH 8.0, Tris-HCl at 25 °C and then the substrate was added and checked for residual activity.

3. Results and discussion

Biophysical characterization of NprotI was carried out. There was evidence of presence of PPII fold in the enzyme. Fluorescence quenching experiments were carried out on NprotI using acrylamide (neutral), I^- (negatively charged) and Cs^+ (positively charged) quenchers. Also action of various protease inhibitors was studied.

3.1 Intrinsic fluorescence

NprotI contains a single tryptophan residue [6]. Emission maximum at 353 nm was observed with excitation at 295 nm, indicating that the tryptophan residue is exposed to the solvent. The decomposition analysis of trp fluorescence spectra was carried out using PFAST program (http://pfast.phys.uri.edu/pfast/) [7] and it indicated presence of type III tryptophan, i.e. tryptohan is solvent exposed. Emission maximum did not shift in presence of GdnHCl as NprotI is resistant to GdnHCl denaturation [3]. Emission maximum was red shifted by 6 nm in presence of more powerful denaturant GuSCN indicating denaturation of the enzyme (Fig. 1).

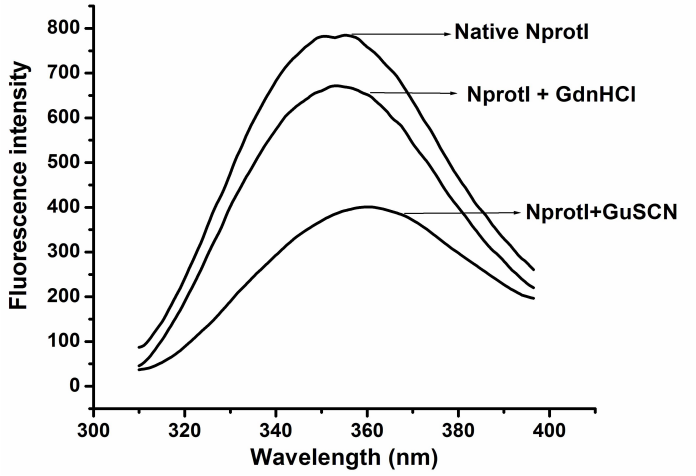


Figure 1: *Intrinsic fluorescence of NprotI under different conditions.*

3.2 CD spectroscopy: Evidence of polyproline II fold in NprotI

The native protein showed a classic far UV CD spectrum of polyproline II structure with a positive band at 230 nm and a negative band at 212 nm (Fig. 2a). Below the large negative band at 212 nm, the ellipticity decreased in magnitude up to 195 nm below which it is again negative. The cross over between two bands at 212 nm and 230 nm is at about 224

nm. The maximum ellipticity of the negative band is four times greater than the maximum ellipticity of the positive band. All these characteristics of the CD spectrum matched very well with the VCUD (vacuum UV CD) of polyproline II film cast from water [8].

NprotI constitutes 10 % proline, 8 % alanine, 5 % arginine, 8 % glutamic acid/glutamine, 11 % aspartic acid/asparagine, 11 % threonine residues [6], all of which have high propensities to be involved in the polyproline like structure. CD pro analysis of the far UV CD spectrum of native NprotI yielded the values of the secondary structure elements as: α-helix-4.2 %, β-sheet-41.5 %, turns-22.4 % and unordered-33.0 % (NRMSD 0.025). However, the software does not detect the polyproline fold. The near UV CD spectrum indicated ordered tertiary structure (Fig. 2b).

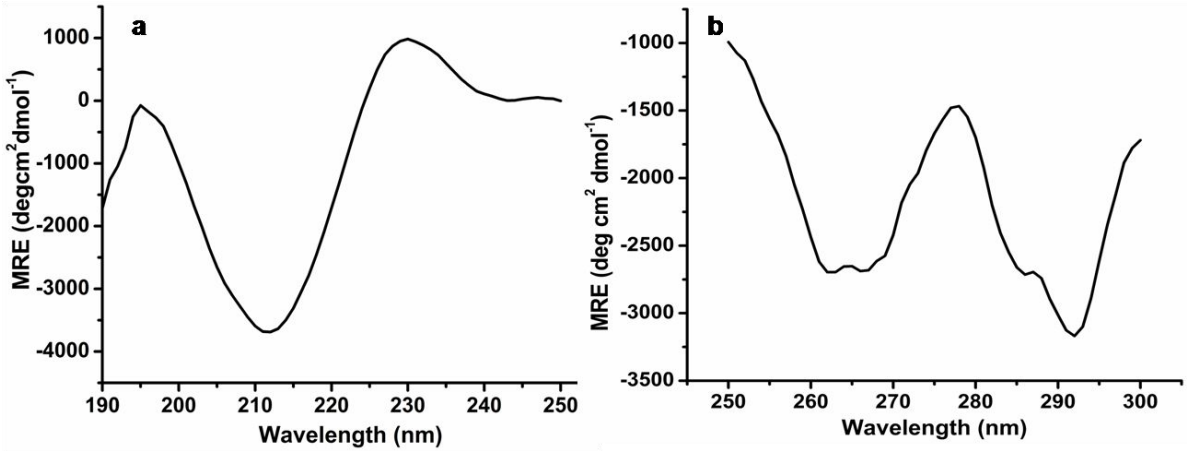


Figure 2**: **CD spectroscopic features of NprotI **a)** Far UV CD spectrum (0.25 mg/ml) **b)** Near UV CD spectrum (0.8 mg/ml).

3.3 Solute quenching studies

Quenching data for all the quenchers used in these studies were analysed by the Stern–Volmer equation (1) as well as by the modified Stern–Volmer equation (2) [9].

$$F_o / F_c = 1 + K_{sv}[Q] \quad (1)$$

$$F / \Delta F = f_a^{(-1)} + (K_a f_a)^{(-1)}[Q]^{(-1)} \quad (2)$$

Where F_o and F_c are the respective fluorescence intensities corrected for dilution, in the absence and presence of quencher, [Q] is the resultant quencher concentration. K_{sv} is the Stern–Volmer quenching constant, f_a refers to the fraction of the total a fluorescence that is accessible to the quencher and K_a is the corresponding quenching constant. Slopes of

Stern–Volmer plots yield K_{sv} values, whereas the slopes of modified Stern–Volmer plots give $(K_a\ f_a)^{-1}$ and their ordinates give values of f_a^{-1}. Stern-Volmer and modified Stern-Volmer plots of quenching of native NprotI, NprotI denatured with GuSCN are shown in Fig. 3 and the various constants are summarized in table 1.

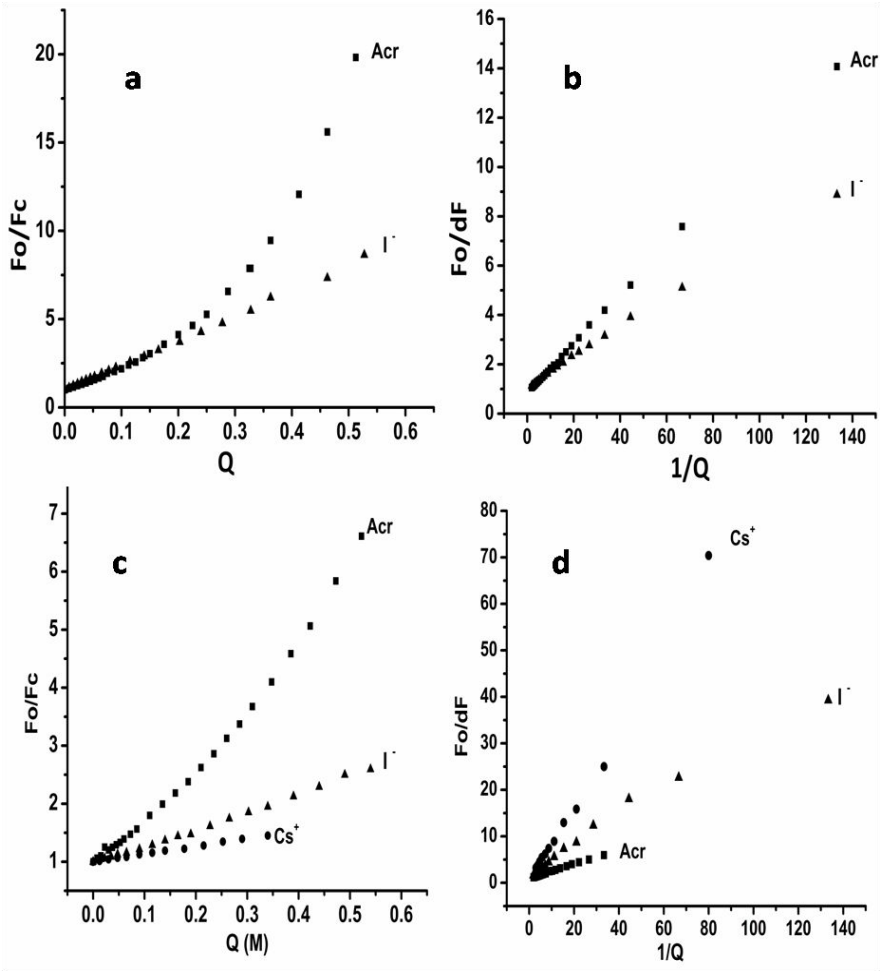


Figure 3: a) and c) *Stern-Volmer plot for native and denatured NprotI respectively* ***b) and d)*** *Modified Stern-Volmer plot for native and denatured NprotI respectively.*

3.4 Solute quenching studies of native NprotI

The upward curvature in case of fluorescence quenching of native NprotI with acrylamide indicated presence of both static and dynamic components. The static and dynamic components were then resolved by time dependent fluorescence quenching. Fluorescence quenching was not observed with CsCl in native conditions (Fig. 3a). A very high rate of

quenching with K_{sv} of 14.02 M^{-1} was observed for $I^{(-)}$ in native conditions. This indicated that tryptophan residue is present in positively charged environment and there are hardly any negatively charged amino acid residues around it.

The modified Stern-Volmer plot provides the information about fraction accessibility of tryptophan to the quencher. For native protein, the single tryptophan, which is surface exposed, was 100 % accessible to neutral quencher acrylamide, while for $I^{(-)}$, it was about 90 % accessible (Fig. 3b).

Table 1: *Quenching parameters for NprotI*

Quencher and Condition	K_{sv} (M^{-1})	K_s (M^{-1})	f_a
Acrylamide			
Native	4.80	5.51	1.19
Native+GdnHCl	-		1.594
Native+GuSCN	10		1.039
KI			
Native	14.017		0.899
Native+GdnHCl	10.067		1.111
Native+GuSCN	2.98		0.548
CsCl			
Native	-		-
Native+GdnHCl	-		-
Native+GuSCN	1.34		0.838

3.5 Solute quenching studies of denatured NprotI

As already mentioned, NprotI does not get denatured with GdnHCl; the quenching parameters were almost constant for NprotI treated with 6 M GdnHCl. For the protein denatured with GuSCN only dynamic component was observed with acrylamide and a high K_{sv} obtained indicated higher rate of quenching in the denatured protein. Quenching with CsCl with K_{sv} of 1.34 M^{-1} in denatured protein indicated presence of negative charge environment surrounding tryptophan (Fig. 3c). The K_{sv} (2.98 M^{-1}) for KI was decreased in denatured protein compared to that of native. This suggested charge reorientation in the denatured protein. The fraction accesiblity for acrylamide in the denatured protein remained same around 100 %. For $I^{(-)}$ it decreased to about 55 % while for $Cs^{(+)}$ it was about 84 % (Fig. 3d). It can be seen here that the accessibility to $I^{(-)}$ was decreased to large

extent along with decrease in the quenching rate. But for $Cs^{(+)}$ the quenching rate was not much higher but the accessibility was higher.

3.6 Solute quenching studies of NprotI in different pH conditions

The enzyme was structurally stable in all the pH conditions; while it is most stable in acidic pH (pH 1.0), it is most active at pH 10.0 (discussed in chapter 4). To investigate the structural changes at these pHs, solute quenching studies with ionic quenchers were carried out at pH 1.0, 5.0 and 10.0 (Fig. 4a, b). No quenching with $Cs^{(+)}$ was observed at all the pH, which indicated that still there was no negative charge surrounding tryptophan, which could be due to maintenance of native structure at all the pH conditions and this supported our previous reports. The rate of quenching with KI increased with decreasing pH indicating that in acidic pH the positive charge around tryptophan had increased. Also, at pH 1.0 with increasing additions of KI, the enzyme was found to get denatured. This could be due to presence of very high positive charge on the enzyme at pH 1.0. Although, enzyme at pH 1.0 was as thermostable as that at pH 5.0, it was found to be susceptible to chemical denaturation (described in next chapters). The observation that NprotI at pH 1.0 gets denatured with addition of KI supported the earlier study.

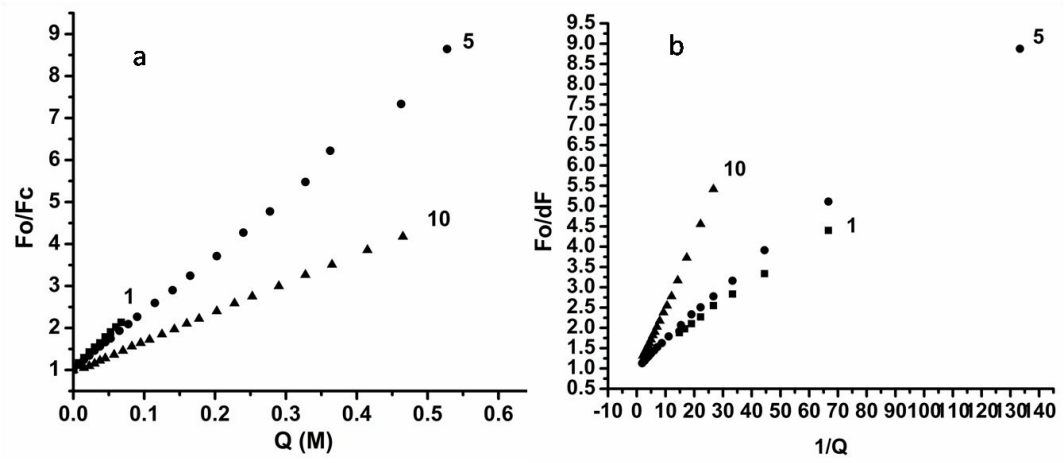


Figure 4: Quenching with KI pH 1.0, 5.0 and 10.0 **a)** Stern-Volmer plot **b)** Modified Stern-Volmer plot. Number on each plot indicates pH of the study.

3.7 Time-resolved fluorescence study

Events that occur during the lifetime of the excited singlet state can be monitored using time-resolved fluorescence spectroscopy. This time scale can range from a few picoseconds to hundreds of nanoseconds [10].

The fluorescent decay of the tryptophan residue on a nanosecond time scale for NprotI, obtained from time resolved measurements are presented in Fig. 5. Monoexponential curve could be fitted ($\chi 2$=1.13) to time resolved fluorescence profile indicating one conformer of tryptophan with value of 5.13 ns of decay time for tryptophan. This could be due to presence of tryptophan in a stable environment with fewer fluctuations so that the electron transfer processes with neighboring solvent molecules or amino acid residues are lower [11]. Such monoexponential decay has been observed in case of ribonuclease T1 [12].

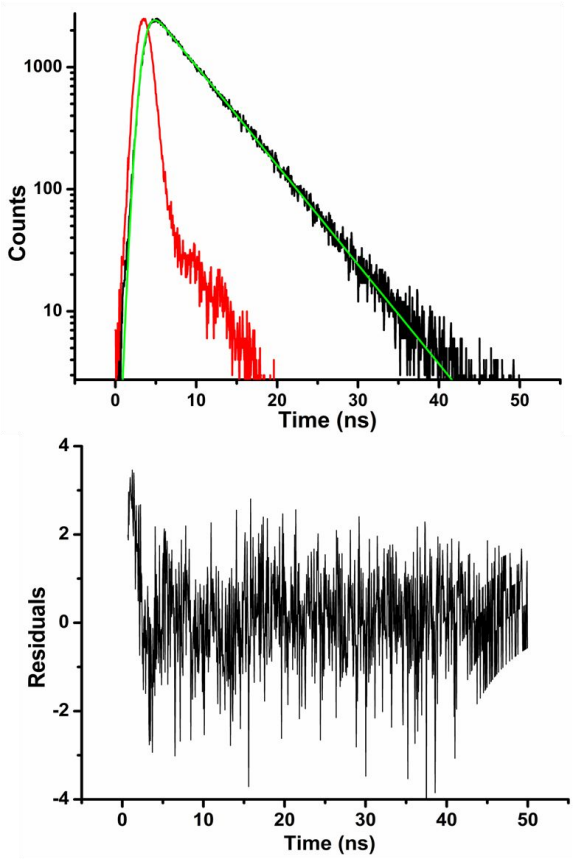


Figure 5: Time resolved fluorescence decay profile of NprotI. The red line corresponds to the instrument response, the black line corresponds to the experimental data and the green line corresponds to the non-linear monoexponential fit of the experimental data. The lower panel represents the residuals.

3.8 Time-resolved fluorescence quenching of native NprotI with acrylamide

The profile obtained for acrylamide quenching in native conditions showed an upward curvature, indicating dynamic and static components of quenching. The static mechanism

involves complex formation, while dynamic mechanism involves collisions with acrylamide during the lifetime of tryptophan in excited state. In such a case, the data can be analyzed using the following eqn. and the dynamic and the static components can be resolved. The parameters obtained for the time resolved quenching with acrylamide are given in table 2.

$F_o/F_c = (1+K_{sv}[Q])(1+K_s[Q])$ (3)

Where K_{sv} is the Stern–Volmer (dynamic) quenching constant, K_s is the static quenching constant and [Q] is the quencher concentration. The dynamic quenching constant reflects the degree to which the quencher achieves the encounter distance of the fluorophore and can be determined by the fluorescence lifetime measurements according to the equation:

$\tau_o/\tau = (1+K_{sv}[Q])$ (4)

Where τo is the average lifetime in absence of the quencher and τ is the average lifetime in presence of a quencher at a concentration [Q]. The value of K_{sv} obtained for acrylamide quenching of NprotI was 4.8 M^{-1} (Fig. 6a). Putting this value in eq. (3) and plotting a graph of $(F_o/F_c)/(1+K_{sv}[Q])$ against [Q], the value of the static quenching constant Ks was obtained as 5.51 M^{-1} (Fig. 6b) and the bimolecular quenching constant, kq was calculated as $kq=K_{sv}/\tau$, and was found out to be 0.9375×10^9 $M^{-1}s^{-1}$. Incorporating the values of K_{sv} and K_s in the expression $(1+K_{sv}[Q])(1+K_s[Q])$, the values obtained were plotted against [Q]. It was observed that the values of F_o/F_c and $(1+K_{sv}[Q])(1+K_s[Q])$ match very well (Fig 4b).

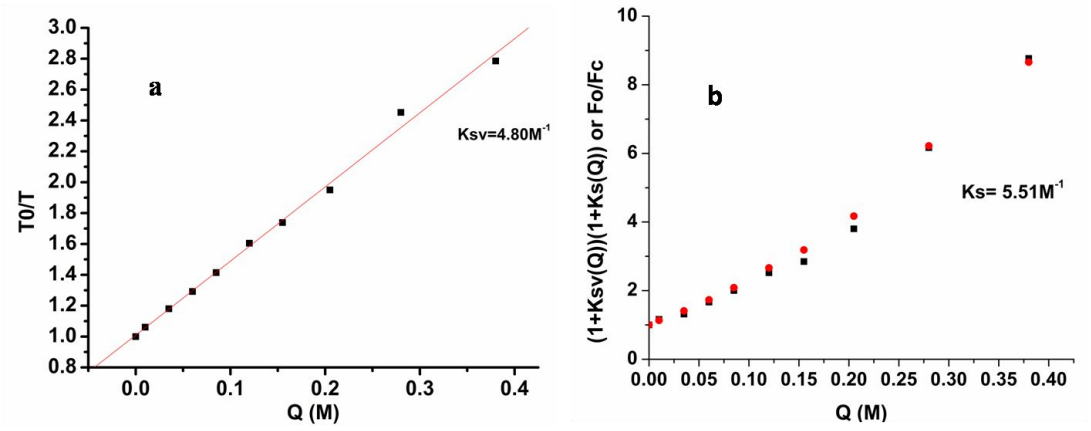


Figure 6: Quenching of NprotI fluorescence with acrylamide studied on time resolved spectrofluorimeter **a)** The plot of τ_o/τ for the quenching data of NprotI with acrylamide **b)** The plot of Fo/Fc and (1+Ksv[Q])(1+Ks[Q]) against [Q].

***Table 2**: Parameters for time-resolved quenching of tryptophan fluorescence of NprotI with acrylamide*

Q(M)	T1 (ns)	A1	T2 (ns)	A2	T (ns)	chi^2	T0/T
0	5.1299	0.028	0	0	5.1299	1.135	1
0.01	4.8358	0.028	0	0	4.8358	1.108	1.06082
0.035	4.345	0.029	0	0	4.345	1.091	1.18064
0.06	3.9705	0.034	0	0	3.9705	1.099	1.292
0.085	3.6267	0.035	0	0	3.6267	1.073	1.41448
0.12	3.1967	0.034	0	0	3.1967	1.182	1.60475
0.155	2.9504	0.039	0	0	2.9504	1.213	1.73871
0.205	2.6308	0.036	0	0	2.6308	1.36	1.94994
0.28	1.8707	0.033	4.5357	0.003	2.0918	1.187	2.45239
0.38	1.5874	0.038	5.1941	0.003	1.8414	1.159	2.78587

3.9 Kinetics of modification of NprotI by PMSF

Although nature of the NprotI as a serine protease was known, to characterize this trait concentration dependent and time dependent inactivation study (Fig. 7a) was performed in presence of PMSF which is an irreversible inhibitor of serine proteases. Slope of a double logarithmic plot showed that a single serine residue is present at the active site (Fig. 7b). Substrate protection was also observed against the inactivation by PMSF. K_m and V_{max} were also calculated of the enzyme inactivated with PMSF.

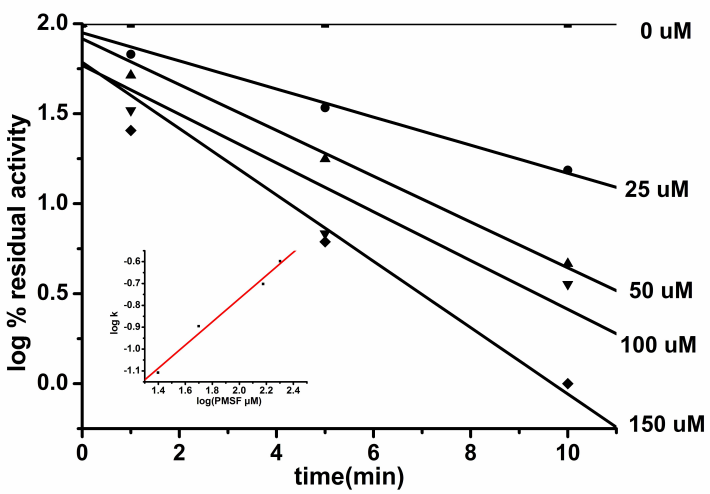


***Figure 7:** Kinetics of modification of NprotI by PMSF.*

3.10 pKa of the amino acid residues at active site and chemical modification with phenylglyoxal (PG)

Although the present enzyme is serine protease and the catalytic triad is known, being an alkaline enzyme, we explored the active site residues. The pH activity profile (Fig. 8a) showed that pKa of the amino acids at active site was in the range of 9.4 and 10.3 i.e. on the basic side. Hence investigation of basic amino acid residue at active site (other than histidine) was carried out with chemical modification studies. Inhibition was observed with phenyl glyoxal (Fig. 8b) which modifies the Arg residue [13]. Then the inhibition was checked for time dependence and substrate protection which was not observed indicating that the Arg residue might not be involved in catalysis but might be responsible for holding the active conformation.

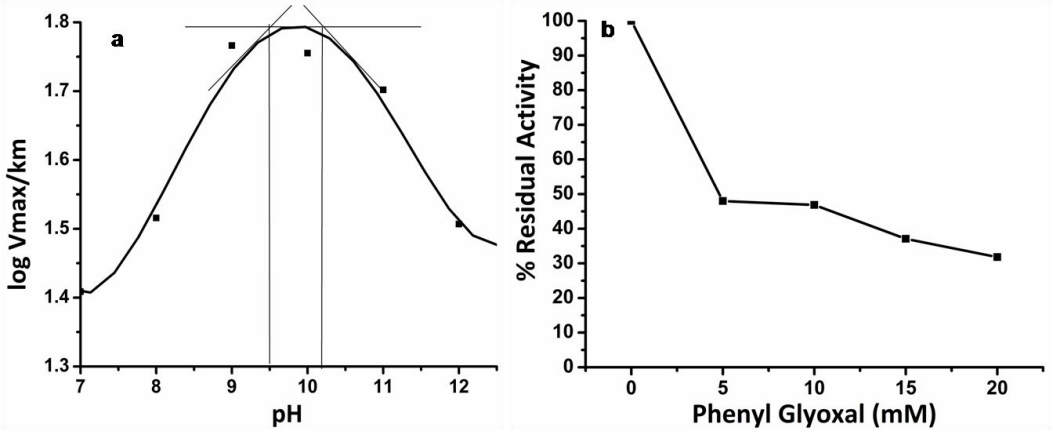


Figure 8: a) *Determination of pKa from pH-activity profile* ***b)*** *Inhibition profile with phenyl glyoxal.*

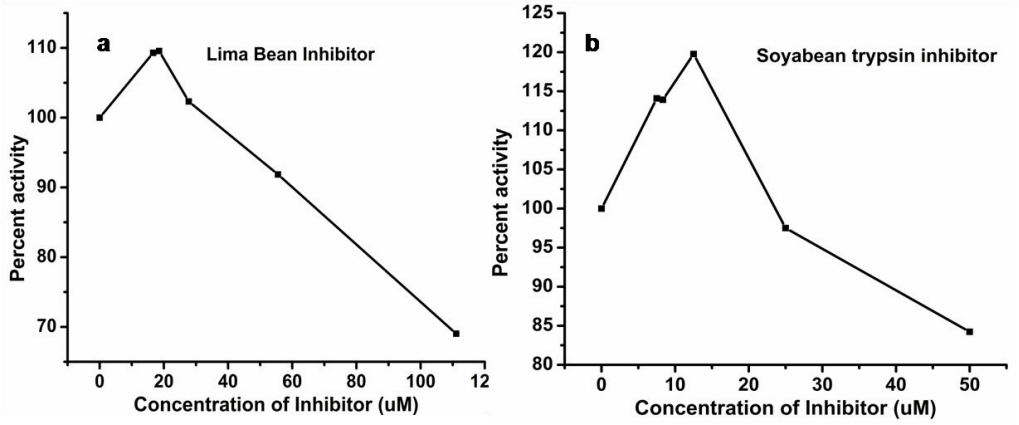


Figure 9: Activity profile in presence of **a)** Lima bean inhibitor **b)** Soyabean trypsin inhibitor.

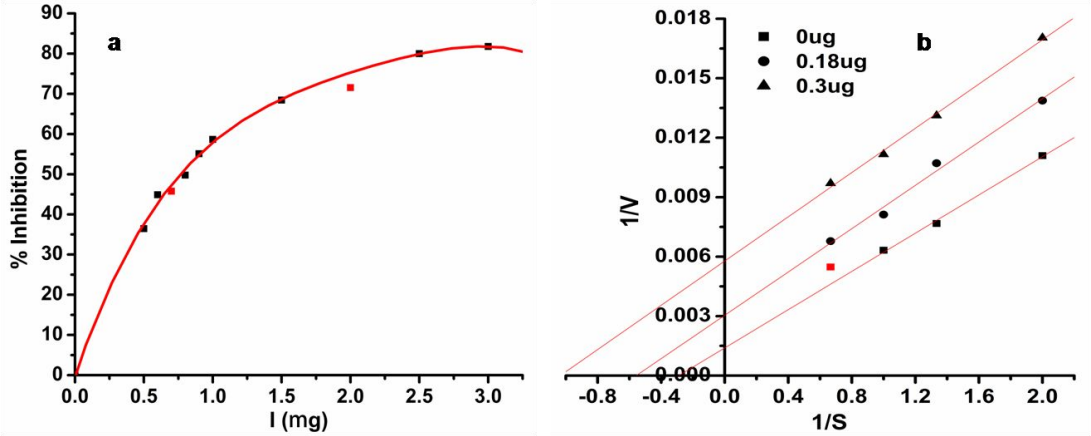


Figure 10: Inhibition profile with Can PI **a)** IC50 determination **b)** Kinetics of inhibition with CanPI 7 showing uncompetitive inhibition.

3.11 Activity profile of NprotI in presence of soyabean trypsin, lima bean and CanPI 7 inhibitor

The soyabean trypsin, lima bean inhibitors are commercially available serine protease inhibitors. Interestingly, activation of the enzyme activity was observed at lower concentrations of these inhibitors. At higher concentrations (above 60 µM), the inhibition was achieved (Fig. 9 a,b). Another inhibitor from plant origin *Capsicum annuum* Pin-II PI (CanPI-7) was used in studies. This inhibitor showed inhibition at much lower concentrations. The IC50 value for this inhibitor was 0.8 µg (32.91 nM) (Fig. 10a). The Line-Weaver-Burk plot of the data showed an uncompetitive inhibition (Fig. 10b), hence further kinetics could not be studied [14]. The uncompetitive inhibition indicated that the inhibitor binds to the site other than the active site.

4. References

1. M Manning, W Colon. *Biochemistry*. **2004**, 43:11248-11254.
2. J M Sanchez-Ruiz. *Biophys Chem.* **2010,** 148:1–15.
3. S B Rohamare, V Dixit, P K Nareddy, D Sivaramakrishna, M J Swamy, S M Gaikwad. *Biochim Biophys Acta.* **2013**, 1834:708-716.
4. S S Lehrer, P C Leavis. *Methods Enzymol.* **1978,** 49:222–236.
5. E M Lakowicz, G Weber. *Biochemistry.* **1973,** 12:4171–4179.
6. V S Dixit, A Pant. *Biochim Biophys Acta.* **2000,** 1523**:**261-268.
7. E A Burstein, S M Abornev, Y K Reshetnyak. *Biophys J.***2001,** 81: 1699–1709.
8. M A Young, E S Pysh. *J Am Chem Soc.* **1975,** 97: 5100-5103.
9. S S Lehrer. *Biochemistry.* **1971***,* 10:3254–3263.
10. J M Beechem, L Brand. *Ann Rev Biochem*. **1985,** 54:43-71.
11. A Ababou, E Bombarda. *Prot Sci.* **2001,** 10:2102–2113.
12. D R James, D R Demmer, R P Steer, R E Verrall. *Biochemistry.* **1985,** 24: 5517-5526.
13. D P Malinowski, I Fridovich. *Biochemistry.* **1979**, 18:5909-5917.
14. M Mishra, R S Joshi, S Gaikwad, V S Gupta, A P Giri. *Biochem Biophys Res Commun.* **2013,** 430(3):1060-1065.

Chapter 3

Role of PPII fold in imparting kinetic stability to NprotI

Polyproline fold – in imparting kinetic stability to an alkaline serine endopeptidase. Biochimica Biophysica Acta (proteins and Proteomics). 2013, 1834:708-716.

Summary

In the present chapter, the presence of polyproline II (PPII) fold was confirmed in NprotI based on structural transitions of the enzyme in presence of GdnHCl and a distinct isodichroic point (far UV CD spectrum) in chemical and thermal denaturation. Also, it's role in imparting kinetic stability to NprotI was studied. The functional and conformational transitions of the enzyme were studied under various denaturing conditions. Enzymatic activity of NprotI was stable in the vicinity of GdnHCl upto 6.0 M concentration, organic solvents viz. methanol, ethanol, propanol (all 90 %v/v), acetonitrile (75 %v/v) and proteases like trypsin, chymotrypsin and proteinase K (NprotI:protease 10:1). The NprotI seems to be a kinetically stable protease with a high energy barrier between folded and unfolded states. Also, an enhancement in the activity of the enzyme was observed in 1 M GdnHCl upto 8 h, in organic solvents (75 %v/v) for 72 h and in presence of proteolytic enzymes. The polyproline fold remained unaltered or became more prominent under the above mentioned conditions. However, it gradually diminished during thermal denaturation above 60 °C. Thermal transition studies by differential scanning calorimetry (DSC) showed scan rate dependence as well as irreversibility of denaturation, the properties characteristic of kinetically stable proteins.

1. Introduction

In previous chapter, based on CD spectrum it was proposed that PPII helix is a major component of NprotI's secondary structure. In the present chapter, this proposal has been confirmed.

Nature has bestowed some uniqueness to all living things in order to survive in their respective environments. Extracellular proteases produced by microorganisms help them by breaking the complex proteins to simple peptides, a more accessible form of food for the microbes. As these proteases are destined to work in harsh environments they have remarkable stability in various denaturing conditions. Some of these proteases have high free-energy kinetic barrier separating the folded and unfolded states and hence prefer to remain in the folded state. These proteases and other difficult to unfold proteins are called kinetically stable proteins. Kinetically stable proteins are difficult to unfold compared to thermodynamically stable proteins [1]. Kinetic stability also imparts SDS and proteolytic resistance to such proteins [2]. There are no evidences of any standard structural pattern responsible for kinetic stability. The physical basis of

kinetic stability differs from protein to protein on individual basis. Manning and Colon have reported high content of β sheet structure in kinetically stable proteins [2].

The proteases from actinomycetes, especially non-streptomyces actinomycete, have not received much attention compared to the proteases from bacteria [4-7]. Extensive structural characterization of the acid resistant protease A from *Nocardiopsis alba* has been carried out by Kelch *et al* [8]. Structural studies of the enzymes from actinomycetes will undoubtedly improve the understanding of the factors responsible for their varied structural stability. In the present chapter, structure-function studies on NprotI; a serine protease from *Nocardiopsis sp.*, were carried out [9].

The structural and functional transitions of NprotI were studied in the presence of chaotropic agents, alcohols, proteases and at high temperature by biophysical and assay based techniques. The CD spectrum of the native protein, structural transitions of the enzyme in presence of GdnHCl and a distinct isodichroic point in various denaturing conditions indicated convincing presence of polyproline II helix (PPII) in NprotI.

In general, the PPII structure is encountered in proteins only locally where sequential proline residues are present, except for collagen and related structures [10]. The PPII structure is present locally in RNA polymerase II C-terminal domain, the high molecular weight (HMW) subunits of wheat glutenin, PEVK segment of the giant elastic protein titin and in dehydrins which are stress related proteins in plants [11-14]. Also, various serine proteases like bovine trypsin, *Streptocmyces griseus* trypsin are known to contain the PPII helix [15]. In bovine trypsin, the PPII-helices form exposed structural elements as they are mainly positioned at the molecule surface. Some PPII-helices can be associated with a specific structural role. E.g. in bovine trypsin, the PPII-helix F (Ile 162-Asp 165), which is also exposed, serves as a connecting segment between the β strand and the α helix in the second domain. The PPII-helices in the *Streptomyces griseus* trypsin has a different role. It is involved in the formation, with some interruptions, of the whole block connecting 2 domains in the molecule. The whole block lies closely to the molecule surface and has high degree of exposure. In all these cases the PPII helix appears as a local structural element.

Here, we propose that this is the first report of PPII helix being the global conformation of a non structural protein from microbial source. In the present study, the role of PPII conformation in imparting kinetic stability to the protein has been reported.

2. Materials and methods

2.1 Materials

Guanidine hydrochloride (GdnHCl) was obtained from Sigma Aldrich Ltd., USA. Trypsin, chymotrypsin and proteinase K were obtained from SRL, India. All other reagents, buffer compounds, organic solvents used were of analytical grade. Solutions prepared for spectroscopic measurements were in MilliQ water.

2.2 Production and purification of NprotI

The protocol for production and purification of NprotI was used as described chapter 2. The purified enzyme was stored at pH 5.0, as it is the pH of maximum stability for NprotI at 2-8 °C.

2.3 Enzyme assay

NprotI was assayed for proteolytic activity as described in chapter 2.

2.4 Circular dichroism (CD) and fluorescence measurements

The CD and fluorescence studies were performed as per protocol mentioned in Chapter 2.

2.5 Treatment of the enzyme with GdnHCl

To study the effect of GdnHCl on activity of the enzyme, 50 µg of NprotI was incubated with GdnHCl in the concentration range of 1.0 M-6.0 M, in 20 mM sodium acetate buffer, pH 5.0 for 15 days at 25 °C. Suitable aliquots were removed at regular time intervals and assayed for enzyme activity.

For fluorescence studies, 30 µg/ml and for CD studies 250 µg/ml of NprotI was incubated in 1-6 M GdnHCl, in 20 mM sodium acetate buffer, pH 5.0 for overnight at 25 °C different concentrations of GdnHCl and the samples were read for regular time intervals. The readings were corrected for blank readings.

2.6 Treatment of the enzyme with solvents

For the assay based studies NprotI (30 µg) was incubated in 75 %(v/v) of methanol, ethanol, propanol, acetonitrile (ACN) and dimethyl sulphoxide (DMSO) at pH 5.0 for 72 h. Aliquots were removed at regular time intervals and assayed for enzyme activity. The CD and fluorescence studies were performed after incubating NprotI in 50 %(v/v)

of above solvents at pH 5.0, 72 h, overnight. The readings were corrected for blank readings.

2.7 Structural and functional studies in proteolytic environment

Trypsin, chymotrypsin, and proteinase K were used for proteolytic digestion of NprotI at 25 °C and 37 °C in 20 mM Tris-HCl buffer at pH 8.0. NprotI and each protease were incubated at 10:1 molar ratio for 24 h. Aliquots were removed at regular time intervals and checked for protease activity. There was no interference in the assay readings from the activities of the proteolytic mixture, as the activity of NprotI was checked at 60 °C, pH 10.0. In these conditions, other proteases were inactive, which was confirmed by assaying suitable controls for the other proteolytic enzymes. CD spectrum was taken after incubating NprotI: trypsin (1:10) at pH 8.0, 25 °C.

2.8 Autocatalysis

NprotI, at different concentrations in the range of 0.01– 0.5 mg/ml, was incubated in 20 mM Tris–HCl, buffer pH 8.0 at 25 °C and 37 °C, for 36 h. An aliquot of the enzyme was used for the residual proteolytic activity measurement. Activity of the enzyme at same concentration after 30 min of incubation under similar conditions was taken as 100 % for the calculation of the residual activity.

2.9 Differential scanning calorimetry

DSC measurements were made on a MicroCal VP-DSC differential scanning calorimeter (MicroCal LLC, Northampton, MA, USA) equipped with two fixed cells, a reference cell and a sample cell. DSC experiments were carried out as a function of pH and scan rate. Sample was dialyzed extensively against 20 mM acetate buffer of pH 5.0, before recording the thermograms. Buffer and protein solutions were degassed before loading. All the data were analyzed by using the Origin DSC software provided by the manufacturer. The data was fitted in non two state transition model. The rate dependence was determined using the methods described by Sanchez-Ruiz et al [16]. The scan-rate-dependent shift in T_m for denaturation was fitted to the equation:

$$\text{Scan rate}/T_m^2 = AR/Ea * e^{-Ea/RT_m}$$

The plot of ln (scan rate/T_m 2) against $1/T_m$ yields a slope $-Ea/R$, where Ea is the energy for denaturation, R the gas constant and A the pre-exponential factor in the Arrhenius equation.

3. Results and Discussion

Structural and functional transitions of a serine protease from *Nocardiopsis* sp. NCIM 5124 (NprotI) were studied in presence of chemical denaturants, organic solvents, proteolytic environment and at different temperature values.

3.1 Resistance of NprotI towards GdnHCl

A. GdnHCl treatment - Activity profile

NprotI was not only stable but also showed 1.25 times enhanced caseinolytic activity in presence of 1 M GdnHCl till 8 h of incubation at 25 °C (Fig. 1a). The enzyme retained about 70 % activity in the vicinity of 6 M GdnHCl up to 24 h and about 30 % activity even after fifteen days (Fig. 1b). The structural basis of these transitions was investigated further.

B. GdnHCl treatment- Fluorescence measurements

When the spectra were recorded in presence of various concentrations of GdnHCl after incubation for 24 h, no shift in λmax of the protein (353 nm, excitation wavelength 295 nm) was observed (Fig. 2a). The change in the fluorescence intensity was also not indicative of any conformational changes suggesting retention of the overall conformation of the NprotI in the presence of GdnHCl.

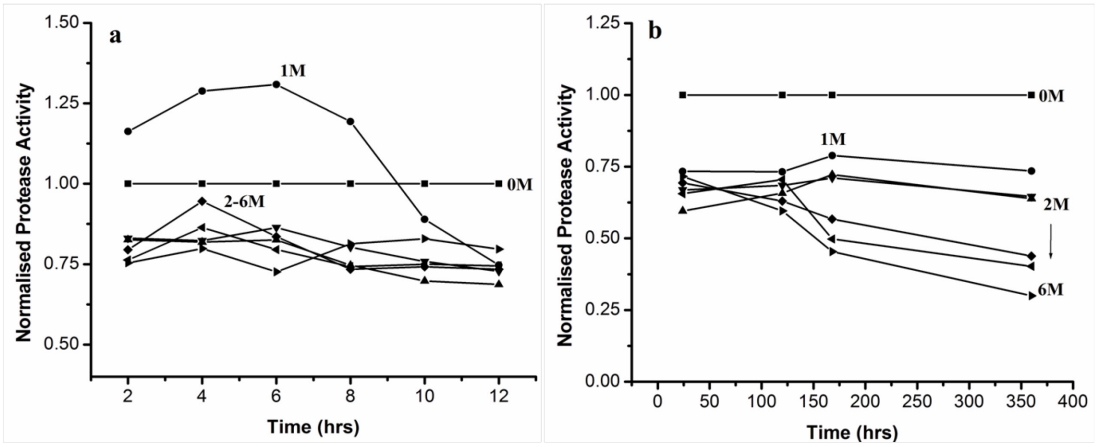


Figure 1: Effect of GdnHCl on NprotI. Normalized protease activity profile of NprotI incubated in different concentrations of GdnHCl at pH 5.0, 25 °C, for different time intervals. Residual activity upto **a)** 12 h and **b)** 24 h to 360 h.

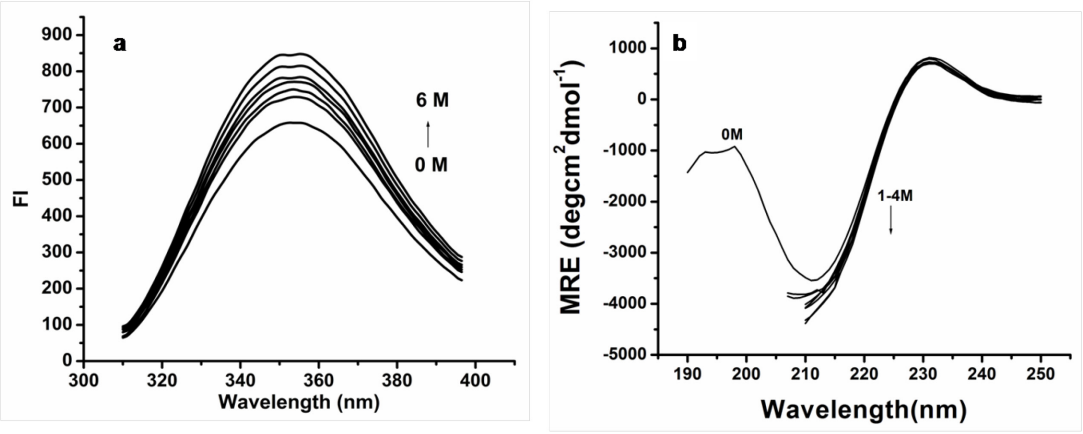


***Figure 2: a) Fluorescence profile and b) Far UV CD spectrum of NprotI*;** the pH of the buffer used was 20 mM sodium acetate buffer of pH 5.0 and the samples were scanned at 25 °C.

C. GdnHCl enhances PPII content - Circular dichroism studies

The MRE value of the positive band at 230 nm remained constant and that of the negative band at 212 nm increased when NprotI was incubated with increasing concentrations of GdnHCl from 1 M to 4 M (Fig. 2b), indicating stability of the polyproline fold. The presence of an isoelliptic point near 222 nm in GdnHCl profile and increase in the structural content with an increase in denaturant concentration confirmed the presence of PPII structure in NprotI.

The activity profile of the enzyme in the presence of GdnHCl correlated well with the changes in secondary structure revealed by far UV CD spectra. The activation observed in the vicinity of 1 M GdnHCl could be due to the pronounced PPII fold as indicated by increased negative ellipticity. The activity of the enzyme decreased to some extent with increasing concentrations of GdnHCl and was maintained for the rest of the period. The more compact, stable structure formed at higher concentration could have had less access to the substrate. There are reports of stabilization of proteins due to charge screening effect of Gdn^+ or Cl^- ions at subdenaturing concentrations [17, 18]. In case of ferrocytochrome c, protein stiffening was the possible cause of stabilization of the protein through lowering of conformational entropy [19]. The present enzyme is unique in its ability to maintain the activity even in presence of denaturing concentrations of GdnHCl.

Presence of an isoelliptic point in GdnHCl treatment of NprotI at 222 nm suggested an equilibrium between unordered and PPII conformation. Such an isoelliptic point has

been observed in PEVK peptide at 218 nm in the presence of urea [17].

A very interesting feature of PPII structure is the absence of any intramolecular and intermolecular hydrogen bonds and because of that the PPII structure, as elucidated by ^{1}HNMR spectroscopy, is nearly identical to an irregular backbone structure. Water or other solvent molecules might be helping to stabilize PPII structure by hydrogen bonding to its backbone [10]. Denaturants like urea or GdnHCl are known to increase the PPII content by hydrogen bonding preferentially to peptide backbone [13, 17, 20], e.g. in polyproline peptides, in PEVK segment of titin, TP1 and C- terminal domain of RNA polymerase. But, in none of the cases the increase in PPII content has been correlated with activity of the protein. The retention of activity and stability of the enzyme in the presence of GdnHCl could be a combined effect of decrease in entropy of protein due to repulsion of Gdn^{+} ions from the surface and the stability of PPII content of the protein.

The rearrangement of the structure, apparently seen as increased compactness and maintenance of the polyproline fold could be the reason of 50 % residual activity in presence of 4 M GdnHCl even after 15 days. Initially, we speculated that the high positive charge density (pI of NprotI, 8.3) at pH 5.0 on the surface of the enzyme could be repelling the positively charged guanidinium ions. This in turn might be reducing the structural flexibility of the enzyme leading to a compact structure and retention of the activity. However, the detection of PPII fold in NprotI led us to correlate the stability of the protein with the unusual structure as also reported in few other cases [14, 21]. The observation of unusual stability of NprotI also led us to suspect the protein to be a kinetically stable one. The PPII fold of the protein or peptide increases in presence of chaotropic agents like urea and GdnHCl. NprotI was found to be more stable as compared to another kinetically stable protein milin, which retained 70 % activity in 3 M GdnHCl while NprotI retained same amount of activity in the vicinity of 6 M GdnHCl [22].

3.2 Studies in presence of organic solvent

A. Organic solvents enhance NprotI activity

The activity of NprotI was enhanced about two fold in presence of alcohol (90 %v/v) concentration (methanol and ethanol) when assayed after 48 h. The protein remained activated in other solvents like 75 %v/v DMSO till 24 h and same concentration of ACN till 48 h (Fig. 3a). The activity decreased with time, however, total loss of activity

was not observed even after 72 h. DMSO and ACN (90 %v/v) caused total loss of NprotI activity.

B. Stability of PPII fold in organic solvents

The fluorescence profile of NprotI was studied in the presence of different organic solvents. The λmax of the intrinsic fluorescence of the protein got blue shifted by 2 to 6 nm in presence the solvents (Fig. 3c) indicating that the protein structure could be attaining a more compact shape due to which the exposed tryptophan was slightly sheltered.

After incubation of NprotI in methanol, propanol and acetonitrile (all 50 %v/v) for 24 h, the PPII fold was not only stable, but was more pronounced (Fig.3b) which can be correlated with the enhanced activity.

The protein structure and its activity are dependent on nature of the solvent used as ultimately it affects the hydrogen bonding pattern (non-covalent interactions) in the protein. Also, water is the necessary component in many hydrolytic reactions like in serine protease driven catalysis, water molecule acts as a nucleophile.

NprotI was found to be more stable than kinetically stable serine protease milin to the solvent induced denaturation. Milin retained 100 % activity in 40 % methanol compared to enhanced activity observed in case of NprotI in 90 % methanol. Also, milin retained total activity in presence of 60 % ACN while NprotI was activated in presence of 75 % ACN [22].

Activity enhancement has been observed in fish gut trypsin in presence of 1-propanol (6.25 % v/v), the activation effect was amplified with increasing hydrophobicity of the alcohol [23]**.** However, for NprotI, the activation effect decreased with increasing hydrophobicity of alcohols. Rate enhancement in the presence of an organic solvent octane was extensively studied by Klibanov *et al* in case of α-chymotrypsin and subtilisin activity. The ability of enzymes to remain active in organic solvents has been attributed to structural compactness attained which might result in high kinetic barrier for the protein to unfold [24].

Alcohols are known to stabilize the helical secondary structure of the protein and disturb the tertiary structure [25-27]. Addition of alcohols in cytochrome C, myoglobin, lysozyme tend to increase the helical content of those proteins [28-30]. Except for Ervatamin C and VlsE, the alcohol induced β sheet structures are not yet studied well [25, 31]. Several proteins like VlsE from *Borrelia burgdorferi* and cpn10 from human-

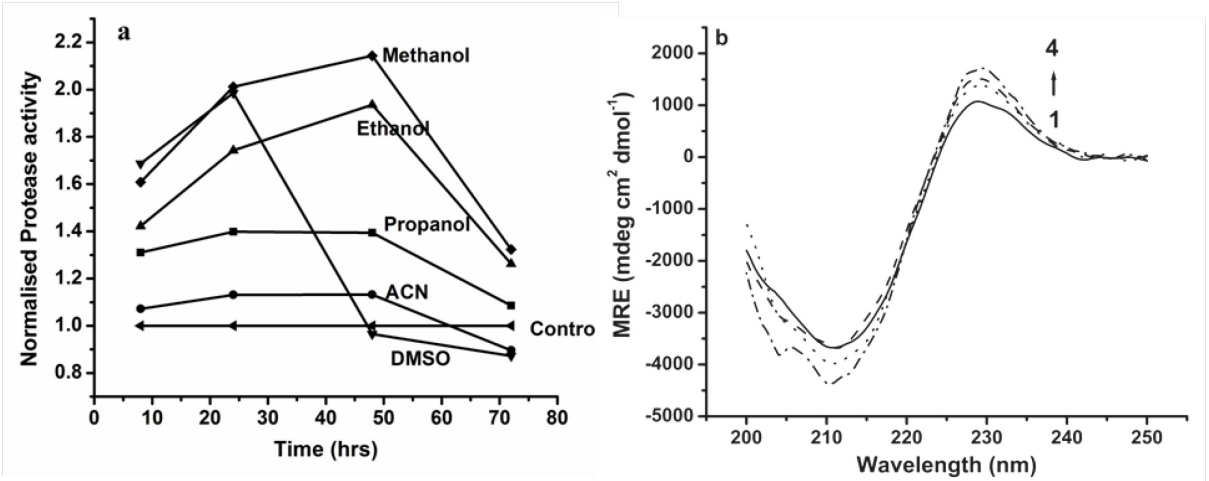


***Figure 3: Effect of Organic solvents on NprotI*: a)** Normalized protease activity profile of NprotI incubated in different solvents 75 %(v/v) at pH 5.0, 25 °C **b)** Far UV CD spectra of NprotI incubated in solvents incubated for 24 h, 50 % (v/v). 1, native protein; 3, methanol; 2, acetonitrile; 4, propanol **c)** Fluorescence profile; λ max in presence of different solvents.

-mitochondria are shown to attain altered secondary structure in presence of alcohol. Addition of methanol in VlsE, an α helical protein induces superhelical structure initially followed by induction of β structure, whereas, in cpn10, a β sheet containing protein, addition of methanol causes initial unfolding followed by induction of non-native β structure [25]. However, the interpretations of conversion of one element of secondary structure into another are based on the visual difference in the spectra and not on the quantitative estimation based on the softwares available for calculations.

The isodichroic point observed in far UV CD spectra at 223 nm indicated equilibration between unordered and PPII conformation. Generally, there is switching of PPII helix to PPI helix in presence of pure aliphatic alcohols like methanol, propanol etc [32-34]. The PPI structure is characterized by CD with a negative band at 198-200 nm, a strong positive band at 214-215 nm, and a weak negative band at 231-232 nm [35].

Presence of such bands was not observed in NprotI in presence of methanol or propanol. This showed the PPII conformation of the enzyme is highly stable. The

peptide with PPII structure, capped at both C and N terminal, has been reported to be more stable to switch to PPI structure in alcohols than the uncapped peptide [36].

3.3 Resistance of NprotI towards other proteases

A. Resistance to proteolysis - Activity profile

The investigation of resistance of NprotI to GdnHCl and organic solvents had already shown the unusual structural property of the protein. The enzyme was checked for its susceptibility towards other proteases. Again, the activity of NprotI was enhanced (upto 2 fold) in presence of trypsin, chymotrypsin and proteinase K, when incubated in 10:1 (NprotI: other protease) ratio at 25 °C (Fig. 4a).

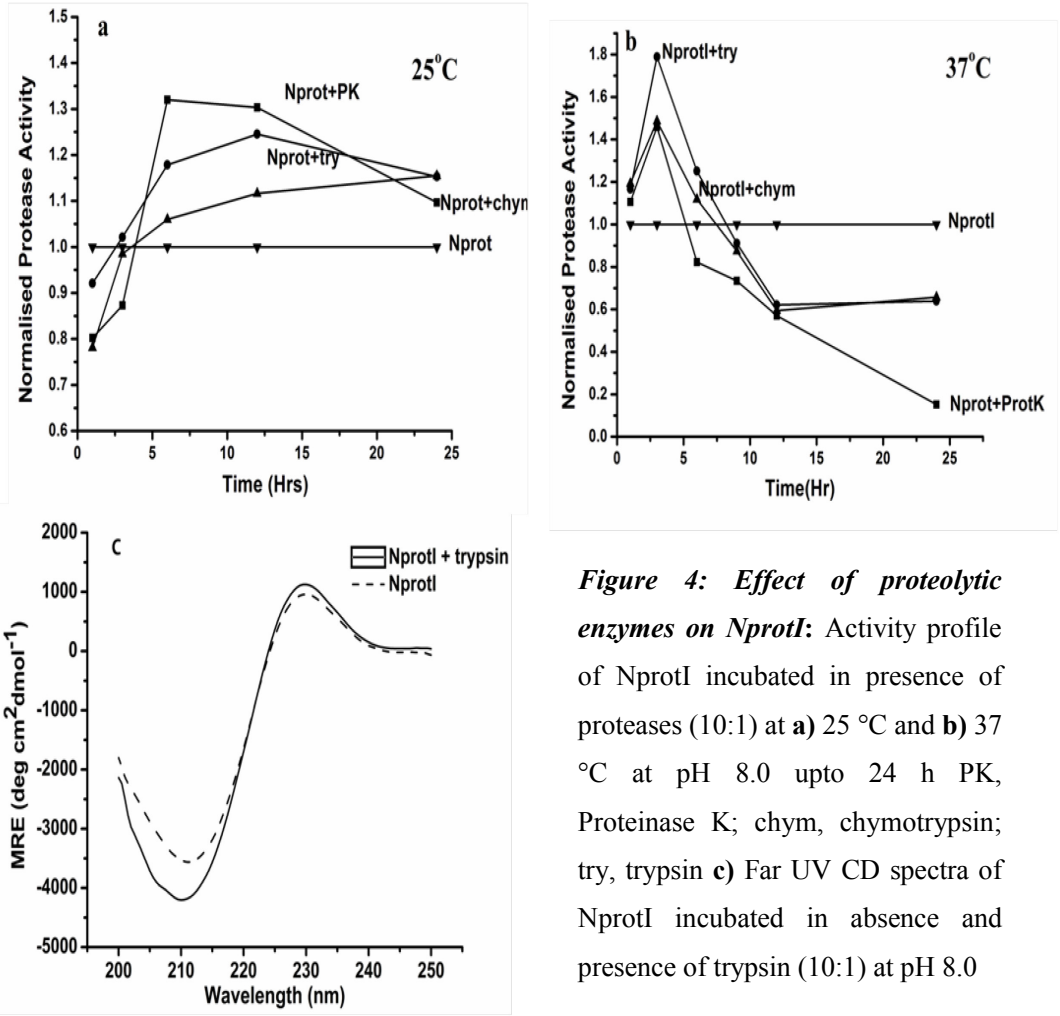


***Figure 4: Effect of proteolytic enzymes on NprotI*:** Activity profile of NprotI incubated in presence of proteases (10:1) at **a)** 25 °C and **b)** 37 °C at pH 8.0 upto 24 h PK, Proteinase K; chym, chymotrypsin; try, trypsin **c)** Far UV CD spectra of NprotI incubated in absence and presence of trypsin (10:1) at pH 8.0

The enzyme was found to be stable to proteolysis at 25 °C for 24 h. Initial enhancement in the activity was observed when the reaction mixture was incubated at 37 °C and there was decrease in activity of NprotI after 8 h in the proteolytic mixture (Fig. 4b).

B. Resistance to proteolysis- CD studies; Structural enhancement

CD spectrum of the mixture of NprotI and trypsin (representing other proteases) in 10:1 ratio showed compactness in the structure i.e. polyproline fold compared to the structure of NprotI alone (Fig. 4c).

The enhancement in activity of NprotI observed in the proteolytic mixture is unusual. The enzyme, being extracellular, must have adapted its structure for survival in harsh conditions. Proteins from extremophilic organisms have evolved to remain stable in harsh conditions like hot, acidic and proteolytic environments. Increased structural compactness has been adapted by many thermophilic enzymes for survival [37, 38]. Here, increased structural compactness was adapted by NprotI to survive in proteolytic conditions. NprotI was more prone to proteolysis at 37 °C, which could be due to some structural changes and/or exposure of proteolytic cleavage sites. Milin was found to be more stable to proteolysis than NprotI. It was stable in presence of other proteases even at 37 °C for 24 h, this could be due to the presence of extensive glycosylation in milin rendering protection to the proteolytic cleavage sites [39].

NprotI might be resisting interaction with other proteases by enhancing structural compactness as well as by maintenance of the polyproline fold. There are many reports on the involvement of PPII fold in protein-protein interactions [9]. The PXXP sequence peptide of the p85 subunit of P13-kinase interacts with Fyn SH3 domain and it goes from unstructured form to PPII-type helix upon binding. Also the PEVK module, with polyproline fold, of titin interacts with SH3 domain of the nebulin during IZI assembly in muscles [40, 41]. M. P. Williamson in his review on structure and function of proline rich regions in proteins has emphasized that the binding of the polyproline to ligands is not specific. This allows binding to wider range of ligands eg. Salivary PRPs binds to wide range of polyphenols and other substrates [42].This shows that the polyproline structure senses the surrounding environment and adapts the structure accordingly.

3.4 Autocatalysis

Most proteases undergo autodigestion, which is dependent on protein concentration, temperature and other experimental conditions. The autocatalytic nature of NprotI was

monitored in the protein concentration range of 0.01–0.5 mg/ml at 25 °C and 37 °C. The extent of autodigestion was expressed in terms of remaining activity as shown in Fig.5. The magnitude of autodigestion of the protein was found to be inversely concentration and directly temperature dependent. At very low protein concentrations (0.01–0.08 mg/ml), autolysis is prominent while a marked reduction in autolysis was observed at higher protein concentrations (0.08 mg/ml and above). It is apparent that the protein is less susceptible to autodigestion at concentrations above 0.1 mg/ml. The increased autolysis at higher temperature could be the reason of NprotI being prone to proteolysis at higher temperature.

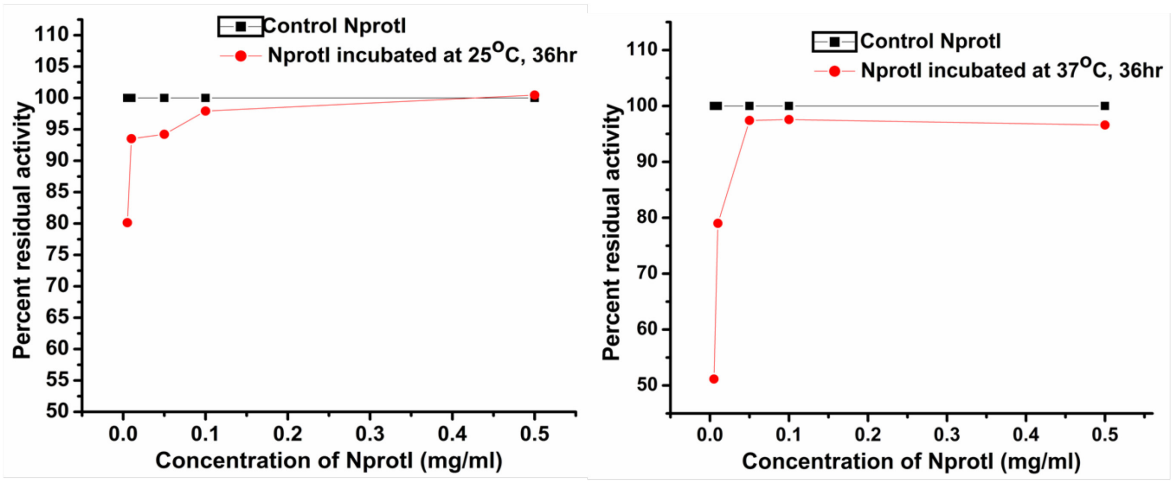


Figure 5: Autolysis study of NprotI at 25 °C and 37 °C for 36 h.

3.5 Thermal denaturation of NprotI: Sensitivity of PPII helix to high temperature

A. Thermal unfolding of NprotI-Activity profile

NprotI retains 60 % activity when incubated alone at 50 °C and 60 °C, pH 5.0 for upto 9 h, while; it retains 50 % activity at 50 °C, pH 5.0 even after 24 h (Fig. 6a). The enzyme loses activity at 70 °C within 1 h, pH 5.0.

B. Thermal unfolding of NprotI-Fluorescence profile

The λmax of NprotI remained almost unchanged with increasing temperature (Fig. 6b). The intrinsic fluorescence of NprotI decreased gradually with increasing temperature, which might be because of deactivation of the excited state by non radiative process (Fig. 6c). The loss in intensity was not regained after cooling the protein suggesting

irreversible changes during the temperature unfolding process.

C. Thermal unfolding of NprotI-CD profile

The Far UV CD spectra of NprotI at different temperatures indicated gradual decrease in the MRE at 230 nm and 200 nm at and above 65 °C, indicating loss in the secondary structure (Fig.7a). An isodichroic point could be seen near 218 nm for the thermal unfolding of NprotI. These points are attributed to the conformational equilibrium between unordered and PPII conformation [17]. Such isodichroic points have been seen in temperature dependent spectra of many PPII containing peptides, e. g., in the consensus sequence CTD eight repeat (R8), such point was observed at 213 nm in water as the temperature increased from 2 °C to 60 °C, in aqueous solutions for poly (Lys) at 203 nm [17], poly (Glu) at 209 nm [43], and systemin, a natural octadecapeptide, at 209 nm [44].

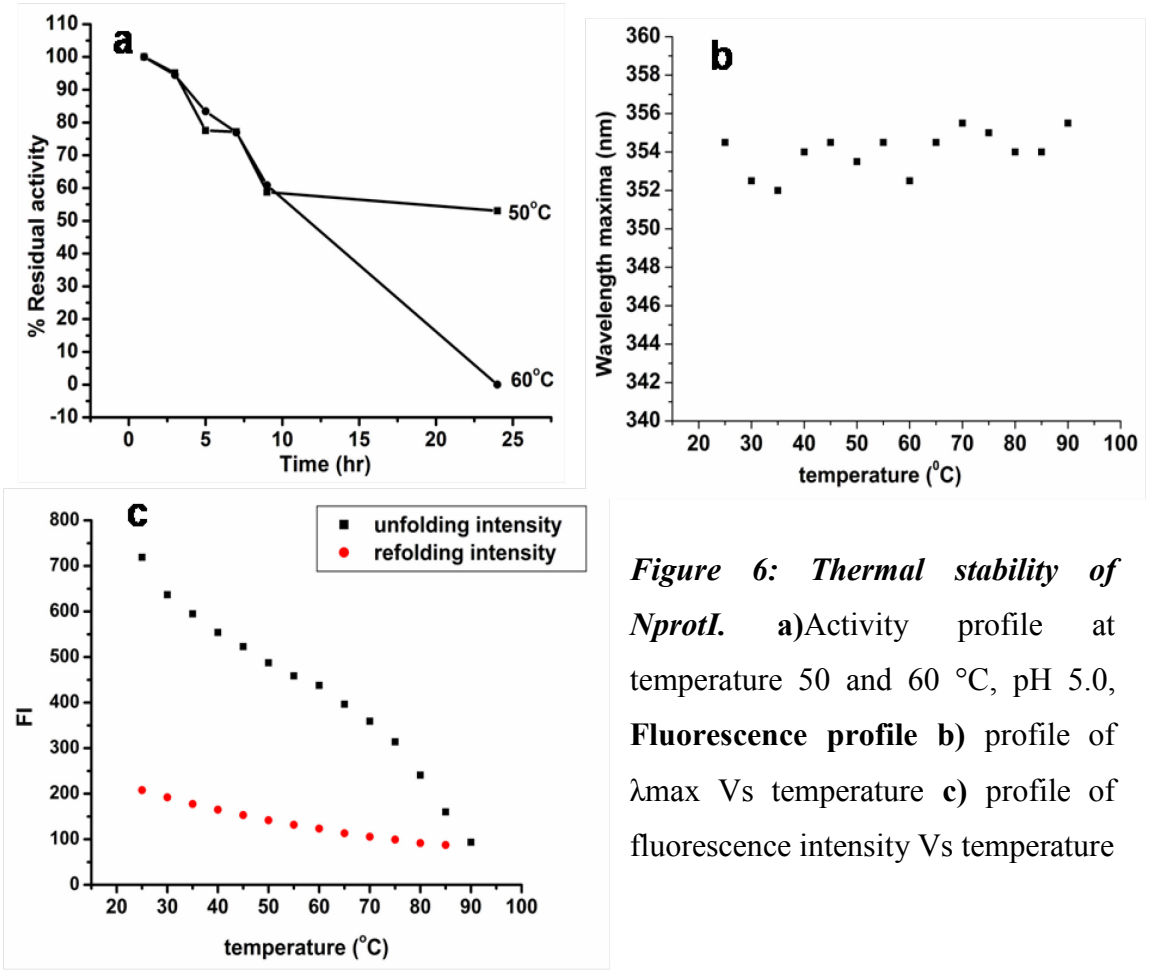


Figure 6: Thermal stability of NprotI. **a)**Activity profile at temperature 50 and 60 °C, pH 5.0, **Fluorescence profile b)** profile of λmax Vs temperature **c)** profile of fluorescence intensity Vs temperature

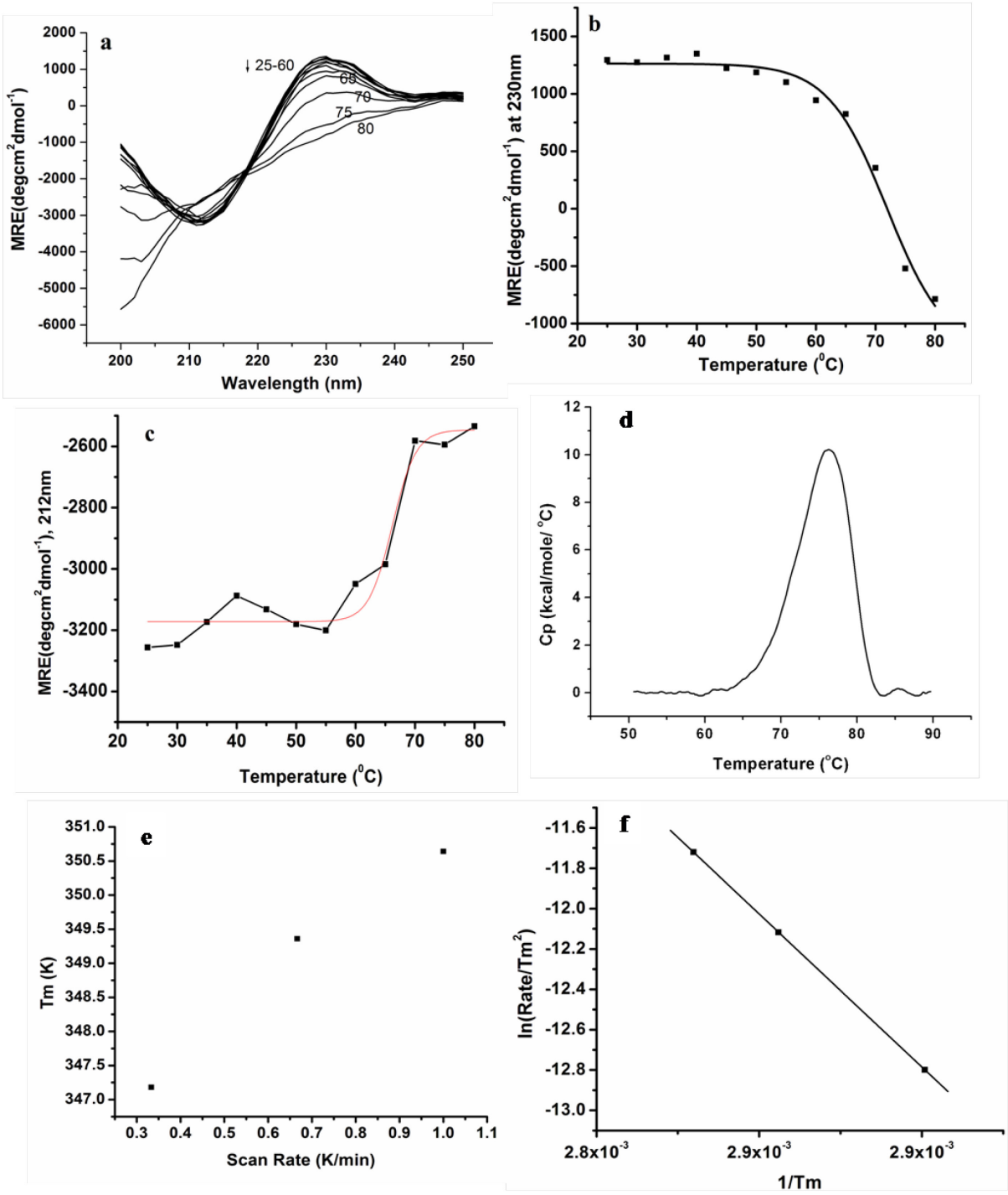


Figure 7: Thermal denaturation of NprotI. **a**) Far UV CD spectra of NprotI (0.25 mg/ml, pH 5.0) after incubation at increasing temperature from 25 to 90 °C with the interval of 5 °C; for 10 min at each temperature. Numbers on line indicate temperature; sigmoidal fits of MRE at **b)** 230 nm and **c)** 212 nm Vs temperature **d)** DSC scan of NprotI in 20 mM acetate buffer of pH 5.0, Analysis of DSC profiles of NprotI **e)** Scan rate dependence **f)** Activation energy plot for NprotI.

The sigmoidal fits for the plots of MRE 212 nm and MRE 230 nm versus temperature showed the T_m values 66.17 °C and 77.93 °C, respectively indicating possible existence of different structural domains in NprotI (Fig. 7 b,c). CD studies indicated presence of two structural domains in NprotI which is also reported in other proteases, e.g., a serine protease from *N. alba* contains a domain bridge and a large β hairpin structure [8]. The alkaline serine protease, KP-43 from Bacillus sp. KSM-KP43 also consists of 2 domains [45]. The scan rate dependence as observed in NprotI has also been studied in kinetically stable proteins such as the lipase and its variant from *Thermomyces lanuginose* [46].

D. Thermal unfolding of NprotI–DSC studies

Thermal unfolding of NprotI was investigated by differential scanning calorimetry at pH 5, where the protein exhibits maximum stability. A representative thermogram recorded at a scan rate of 40 Kh^{-1}, corrected for buffer base line is shown in Fig. 7d.

From the thermogram the Tm of the thermal unfolding of NprotI was determined as 76.4 °C, whereas the area under the endotherm yielded the corresponding denaturation enthalpy (ΔH) as 90.2 kcal/mol. DSC experiments performed at different scan rates clearly showed that the unfolding temperature increases significantly and linearly with increase in the scan rate (Fig. 7e). Further, for each sample, no transition was observed in the second scan clearly showing that the thermal unfolding of NprotI is irreversible. These observations clearly indicated that the thermal unfolding of NprotI is kinetically controlled. Kinetic activation energy for irreversible denaturation was derived from the scan rate dependence of the DSC transition of NprotI using Arrhenius plot (Fig. 6f). From the slope of this plot the activation energy for the thermal unfolding of NprotI was estimated to be 67.8 kcal/mol. The lower transition temperature observed in CD analysis of NprotI could be due to the lower scan rate.

4. References

1. J M Sanchez-Ruiz. *Biophys Chem.* **2010**, 148: 1–15.
2. M Manning and W Colon. *Biochemistry*. **2004**, 43: 11248-11254.
3. S S Jaswal, M E T Stephanie, K A Dill, A A David. *J Mol Biol.* **2005**, 347: 355–366.
4. V J Mehta, J T Thumar, S P Singh. *Bioresour Technol.* **2006**, 97: 1650–1654.
5. W N Hozzein, Li Wen-Jun, I A A Mohammed, O Hammouda, S M Ahmed, Xu Li-Hua, J Cheng-Lin. *Int J Syst Evol Microbiol.* **2004**, 54: 247–252.
6. S Mitsuiki, M Sakai, Y Moriyama, M Goto, K Furukawa. *Biosci Biotechnol Biochem.* **2002**, 66: 164–167.
7. S D Gohel, S P Singh. *J Chromatogr.* **2012**, B 889– 890 61– 68.
8. B A Kelch, K P Eagen, F P Erciyas, E L Humphris, A R Thomason, S Mitsuiki, D A Agard. *J Mol Biol.* **2007**, 368: 870–883.
9. V S Dixit, A Pant. *Biochim Biophys Acta.* **2000**, 1523: 261-268.
10. F Rabanal, M D Ludevid, M Pons, E Glralt. *Biopolymers.* **1993**, 33: 1019-1028.
11. E A Bienkiewicz, Y A M Woody, R W Woody. *J Mol Biol.* **2000**, 297: 119-133.
12. J M Mouillon, P Gustafsson, P Harryson. *Plant Physiol.* **2006**, 141: 638–650.
13. G Gutierrez-Cruz, A H V Heerden, K Wang. *J Biol Chem.* **2001**, 276: 7442–7449.
14. S M Gilbert, N Wellner, P S Belton, J A Greenfield, G Siligardi, P R Shewry, A S Tatham. *Biochim Biophys Acta.* **2000**, 1479: 135-146.
15. A A Adzhubei, M J E Sternberg. *Protein Science.* **1994**, 3:2395-2410.
16. J M Sanchez-Ruiz, J L Lopez-Lacomba, M Cortijo, P L Mateo. *Biochemistry.* **1988**, 27: 1648-1652.
17. L M Mayr, F X Schmid. *Biochemistry*. **1993**, 32: 7994–7998.
18. D Monera, C M Kay, R S Hodges. *Protein Sci.* **1994**, 3: 1984–1991.
19. R Kumar, N P Prabhu, M Yadaiah, A K Bhuyan. *Biophy. J.* **2004**, 87: 2656–2662.
20. P Soon-Ho, W Shalongo, E Stellwagen. *Protein Sci.* **1997**, 6: 1694-1700.
21. M Kan, W Kuan. *Biochem. J.* **2003**, 374: 687–695.
22. S C Yadav, M Pande, M V Jagannadham. *Phytochemistry.* **2006**, 67: 1414–1426.
23. S Harpaz, A Eshel, P Lindner. *J Agric Food Chem.* **1994**, 42: 49-52.
24. A Zaks, A Klibanov. *J Biol Chem.* **1988**, 263: 3194-3201.
25. M Perham, J Liao, P Wittung-Stafshede. *Biochemistry.* **2006,** 45:7740-7749.

26. J W Nelson, N R Kallenbach. *Biochemistry.* **1989**, 28: 5256–5261.

27. A L Fink, B Painter. *Biochemistry.* **1987**, 26: 1665–1671.

28. V E Bychkova, A E Dujsekina, S I Klenin, E I Tiktopulo, V N Uversky, O B Ptitsyn. *Biochemistry.* **1996**, 35: 6058- 6063.

29. Y O Kamatari, T Konno, M Kataoka, K Akasaka. *Protein Sci.* **1998**, 7: 681-688.

30. K R Babu, D J Douglas. *Biochemistry.* **2000**, 39: 14702-14710.

31. M Sundd, S Kundu, M V Jagannadham. *J Protein Chem.* **2000**, 19: 169-176.

32. C Yi-Chun, L Yu-Ju, H Jia-Cherng. *Protein Sci.* **2009**, 18:1967-1977.

33. S Knof, J Engel. *Isr J Chem.* **1974**, 12:165–177.

34. M Mutter, T Wohr, S Gioria, M Keller. *Biopolymers.* **1999**, 51:121–128.

35. S Kakinoki, Y Hirano, M Oka. *Polym Bull.* **2005**, 53: 109–115.

36. M Kuemin, S Schweizer, C Ochsenfeld, H Wennemers. *J Am Chem Soc.* **2009**, 131: 15474–15482.

37. R Scandurra, V Consalvi, R Chiaraluce, L Politi, P C Engel. *Front Biosci.* **2000**, 5: d787–d795.

38. R Jaenicke. *Eur J Biochem.* **1991**, 202: 715–728.

39. S C Yadav, M V Jagannadham. *Eur Biophys J.* **2009**, 38: 981–99.

40. D A Renzoni, D J R Pugh, G Siligardi, P Das, C J Morton, C Rossi, M D Waterfield, I D Campbell, J E Ladbury. *Biochemistry*. **1996**, 35: 15646-15653.

41. M Kan, W Kuan. *FEBS Lett.* **2002**, 532: 273-278.

42. M P Williamson. *Biochem J.* **1994**, 297: 249-260.

43. M L Tiffany, S Krimm. *Biopolymers*. **1972**, 11: 2309-2316.

44. A Toumadje, W C Johnson Jr. *J Am Chem Soc.* **1995**, 117: 7023-7024.

45. T Nonaka, M Fujihash, A Kita, K Saeki, S Ito, K Horikoshi, K Miki. *J Biol Chem.* **2004**, 279: 47344–47351.

46. D Rodriguez-Larrea, S Minning, T V Borchert, J M Sanchez-Ruiz. *J Mol Biol.* **2006**, 360: 715–724.

Chapter 4

Acid stability of NprotI and differential modes of thermal denaturation at different pH

Acid stability of the kinetically stable alkaline serine protease possessing polyproline II fold. (Manuscript under review)

Summary

The kinetically stable alkaline serine protease from *Nocardiopsis sp*.; NprotI, possessing polyproline II fold (PPII) was characterized for its pH stability using proteolytic assay, fluorescence and CD spectroscopy, and DSC. NprotI is optimally active at pH 10.0 and most stable at pH 5.0. However, the enzyme was functionally stable when incubated alone at pH 1.0; even after 24 h, while immediate and drastic loss of the activity was observed at pH 10.0. The activity of NprotI incubated at pH 1.0 and 3.0 was enhanced at higher temperature (50 °C- 60 °C). NprotI maintained the overall PPII fold in broad pH range as seen using far UV CD spectroscopy. The PPII fold of NprotI incubated at pH 1.0 remained fairly intact up to 70 °C. Based on the isodichroic point and T_m values revealed by secondary structural transitions, different modes of thermal denaturation at pH 1.0, 5.0 and 10.0 were observed. DSC studies of NprotI in acidic pH range showed T_m values in the range of 74-76 °C while significant decrease in T_m (63.8 °C) was observed at pH 10.0. NprotI could be chemically denatured at pH 5.0 only with guanidine thiocynate. Acid tolerant and thermostable NprotI can serve as a potential candidate for biotechnological applications.

1. Introduction

Many of the extracellular proteases of microbial and plant origin e.g. alpha lytic protease [1], milin [2] have evolved to remain stable and functional under various harsh environmental conditions such as extremely hot, acidic, or proteolytic environments. These proteases are called as kinetically stable proteases. Studying the mechanisms or structural factors underlying the stability of these proteins is of great importance to understand protein folding and to develop protein engineering strategies. In our previous studies [3], we have reported the presence of polyproline fold (PPII) in a serine protease (NprotI) from *Nocardiopsis sp.* NCIM 5124. The PPII fold was implicated in assigning resistance to the protein against GdnHCl denaturation, proteolytic digestion and solvent denaturation.

In the present chapter, studies on unusual stability of NprotI in the extreme acidic pH range (pH 1.0-3.0) at 50 - 60 °C has been observed using proteolytic assay and biophysical techniques. Most acidophilic proteins are evolved to function solely in low pH conditions and hence do not need to survive and/ or function in neutral and alkaline conditions. The optimum conditions for growth of *Nocardiopsis sp.* NCIM 5124, for production and

stability of NprotI require diverse pH conditions; i.e. pH 10.0 and pH 5.0, respectively. Also, NprotI shows maximum activity at pH 10.0. The same trend (optimum activity in alkaline range, stability in acidic range) was also observed in a serine protease from *N. alba* [4]. This might be because the *Nocardiopsis sp.* can greatly acidify their environment in response to certain nutrients. Thus, the formation of a low pH environment (may be temporary) might necessitate development of secreted proteins with stability and activity at low pH**.**

pH of the medium is a major factor influencing protein stability as it affects the ionic interactions. Also, the thermal denaturation of proteins can be affected by changes in pH, by changing the conformational energy differences between the native and denatured states [5].

The effect of pH on the stability of PPII fold has not been studied extensively as yet. In the present report, the conformational (w.r.t. PPII fold) and functional stability of NprotI was studied over a broad pH range. The overall structure of NprotI was found to be stable in the pH range of 1.0-12.0, while the enzyme showed functional stability for longer time only in acidic pH range. Also, a pH dependence of thermal unfolding was studied using Circular Dichroism (CD) and Differential Scanning Calorimetry (DSC) technique.

2. Materials and Methods

2.1 Materials

Guanidine hydrochloride (GdnHCl) and guanidine thiocyante (GuSCN) were obtained from Sigma Aldrich Ltd., USA. All other reagents, buffer compounds, used were of analytical grade. Solutions prepared for spectroscopic measurements were in MilliQ water.

2.2 Production and purification of NprotI

The protocol for production and purification of NprotI was used as described Chapter 2. The purified enzyme was stored at pH 5.0, as it is the pH of maximum stability for NprotI at 2-8 °C.

2.3 Enzyme assay

NprotI was assayed for proteolytic activity as described in Chapter 2.

2.4 Circular dichroism (CD) and fluorescence measurements

The CD and fluorescence studies were performed as per protocol mentioned in Chapter 2.

2.5 ANS binding studies

The intermediate states of NprotI incubated in 20 mM buffers of different pH (1.0–12.0) were analyzed by the hydrophobic dye, 1-anilinonaphthalene 8-sulphonic acid (ANS) binding. The final ANS concentration used was 50 μM, excitation wavelength was 375 nm and total fluorescence emission was monitored between 400 and 550 nm.

2.7 Effect of pH

Samples of NprotI (250 μg/ml) incubated in an appropriate buffer over the pH range of 1.0–12.0 for 6 h at 25 °C were used for CD measurements as described in Chapter 2. For activity studies, aliquots of enzyme were removed after suitable time interval and assayed at optimum pH and temperature. The following buffers (20 mM) were used for these studies: glycine–HCl for pH 1–3, acetate for pH 4–5, phosphate for pH 6–7, Tris–HCl for pH 8–9 and glycine–NaOH for pH 10–12. pH of the reaction remained stable even after 24 h. ANS binding studies were performed as described in section 2.6.

2.8 Differential scanning calorimetry

DSC measurements were carried out as a function of pH at scan rate of 40 K/h. Samples were dialyzed extensively against buffers of desired pH before recording the thermograms. Same buffers mentioned earlier in section 2.7 were used. Buffer and protein solutions were degassed before loading. All the data were analyzed using the Origin DSC software provided by the manufacturer.

2.9 Effect of GuSCN

NprotI (50 μg) was incubated with various concentrations of GuSCN (0-6 M) at pH 5.0, 25 °C and the enzyme assay was performed after regular time interval.

2.10 Effect of various salts on NprotI

To study the effect on function of the enzyme, NprotI was incubated with NaCl, NH_4Cl, and $MgCl_2$ (0-6 M) in 20 mM sodium acetate buffer, pH 5.0 for 1 h at 25 °C. Suitable aliquot was

removed and assayed for enzyme activity. The readings were corrected for blank readings. For CD measurements, NprotI in final concentration of 250 µg/ml was incubated in 1 M NaCl, NH_4Cl and $MgCl_2$ in 20 mM sodium acetate buffer at pH 5.0 for 1 h at 25 °C and the spectra were recorded.

3. Results and Discussion

The serine protease from *Nocardiopsis sp*. (NprotI) exhibits optimum activity at pH 10.0 and is stored best at pH 5.0 [3, 6]. The enzyme possesses PPII fold and is unusually stable in the vicinity of GdnHCl (24 h in 6 M), organic solvents like methanol, ethanol and DMSO; all (75 % v/v for 24 h), and also in proteolytic environment. Here, we have studied the functional and structural stability of NprotI over wide range of pH for extended time period. Also, a correlation between thermal stability at different pH with special emphasis on PPII fold has been established.

3.1 Stability of NprotI in wide pH range

3.1.1 pH dependent activity profile of NprotI

The pH activity profile of NprotI is shown in Fig.1a. The activity of enzyme incubated at pH 5.0 was assumed to be 100 %, as the enzyme is most stable at this pH for long term storage. Interestingly, the enzyme was found to be more stable below pH 5.0; at extreme acidic pH, retaining full activity in the pH range of 1-3. Incubation of NprotI alone at alkaline pH led to drastic loss in the activity (Fig.1a). Hence, the optimum activity expressed by the enzyme at pH 10.0 at 60 °C (routine assay conditions), must be due the vicinity of substrate casein.

3.1.2 pH dependent structural transitions in NprotI

A. Intrinsic fluorescence measurements

The fluorescence emission maximum wavelength, λmax (353 nm) was not altered when NprotI was incubated in the pH range of 1-12 indicating apparent conformational stability of the protein in wide pH range (Fig. 1b and c). Unaltered fluorescence intensity observed at pH 1.0 indicated no change in the tryptophan microenvironment even after possible protonation of the surface amino acid residues. The deprotonation of surface amino acids of the protein at alkaline pH could have resulted in quenching of the fluorescence intensity.

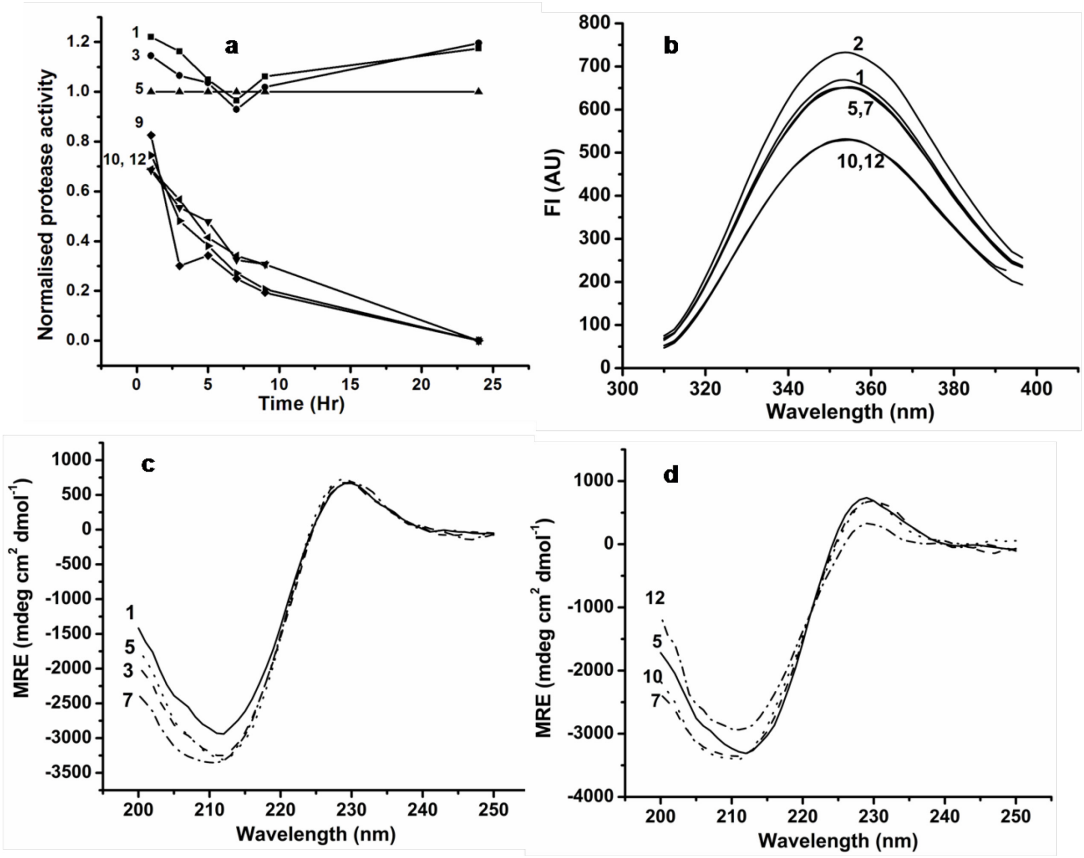


Figure 1: Stability of NprotI at different pH, at 25 °C **a)** Functional stability profile, **b)** Intrinsic fluorescence profile, **c)** and **d)** Far UV CD spectra. Numbers near each spectrum indicates pH.

B. CD spectroscopy

The unique structural property of NprotI i.e. PPII fold as shown by positive ellipticity at 230 nm and negative ellipticity below 200 nm [3] is seen in far UV CD spectra in fig1c. This structural element, especially positive ellipticity around 230 nm remained stable while some alterations, w.r.t. negative minima, were seen at different pH. At pH 1.0, the negative ellipticity is reduced as compared to the one at pH 5.0, although the minima at 212 nm is sharp. While in the alkaline conditions, a trough rather than a sharp minimum was observed indicating some rearrangement in the structure (Fig.1d), which led to comparatively unstable but catalytically active structure. Similar type of rearrangement i.e. shift in the minima by 4 nm in β – casein was observed at alkaline pH (10.5) compared to that at pH

6.75 [7].

Among the very few reports available on correlation of PPII fold transitions with pH, Rucker *et al* [8] have reported the effect of pH on homopolymer lysine. The lysine peptide maintained its PPII structure in varied pH conditions even when side chain charges were heavily screened or neutralized. The plausible explanation given was that the peptide backbone favored the PPII structure to have favorable interaction with solvent. In another study by Shimizu *et al* [9], the type IV collagen has been shown to maintain its PPII fold at acidic as well as neutral pH.

The other kinetically stable proteins like milin and *Nocardiopsis alba* protease (Napase) are also found to be stable over a wide range of pH. Napase has been very well studied for the high acid stability and the reorientation of surface charge residues has been found to be the reason of high acid stability [4]. Milin has been reported to be stable due to heavy glycosylation [2].

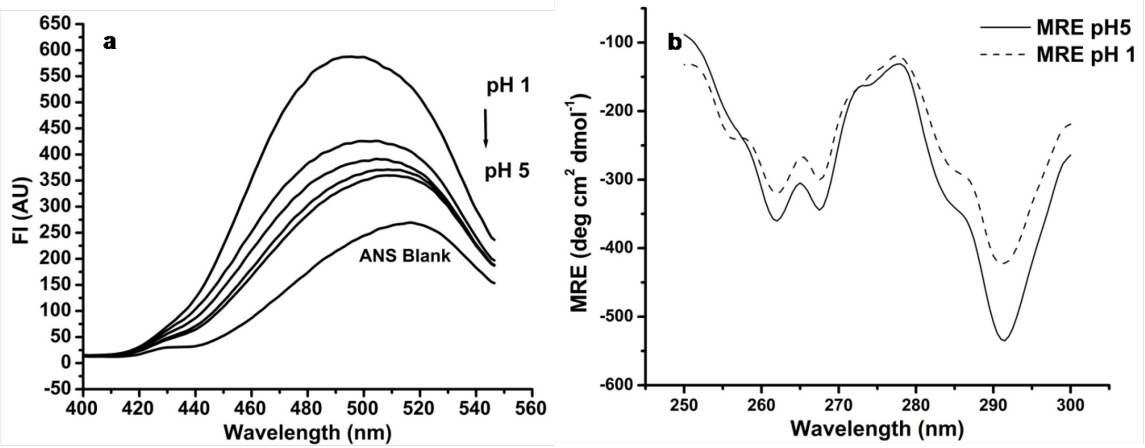


***Figure 2**: **a)** ANS binding to NprotI at different pH and **b)** Near UV CD spectra (800 µg/ml, 6 h incubation) at pH 1 and 5.*

C. ANS binding studies

1-anilinonaphthalene 8-sulphonic acid (ANS) is a dye which shows increased fluorescence intensity when bound to hydrophobic regions in a protein. This phenomenon has been widely used to detect the molten globule state of different proteins. Fluorescence intensity of ANS increased almost two fold with a blue shift in the λmax from 520 nm to 485 nm at

pH 5.0 (Fig. 2a) suggesting exposed hydrophobic amino acids even under native conditions which is not commonly observed. The ANS binding to NprotI was found to further increase in acidic conditions; maximum at pH 1.0, indicating the possible presence of a molten globule like structure.

D. Near UV CD of NprotI

The near UV CD spectrum of NprotI, at pH 5.0 shown in Fig 2b indicates significant ordered tertiary structure of the protein, which is also retained at pH 1.0. The acid induced state appears to retain a significant amount of native-like secondary and tertiary structure, but slight amounts of exposed hydrophobic side chains (Fig. 2a). This suggested the formation of a compact ‘‘molten-globule’’- like intermediate at low pH possessing persistent secondary and compact tertiary structure. The common molten globule conformation reported in the literature [10] is a partly folded one. It is nearly as compact as the native conformation, although with less secondary structure, more hydrated hydrophobic residues, and apparently no defined tertiary structure [11]. Here, NprotI retains the tertiary structure at pH 1.0. The immunoglobulin MAK33 at acidic pH assumes a conformation that possesses all of the characteristics of a molten globule but, in addition, also appears to be stabilized by tertiary interactions [12]. Glucose oxidase shows a refolded conformation (A-state) with 90 % of native secondary structure and native-like near-UV CD spectral features at pH 1.4 [13]. Similarly, the state of NprotI at pH 1.0 could be a molten globule state with stabilized tertiary interactions.

3.2 Thermal stability of NprotI at different pH

A. Activity studies

The unusual stability of NprotI at extreme acidic pH was studied further at higher temperature. Interestingly, NprotI showed 1.8-3.0 fold higher activity at pH 1 and 3 as compared to that at pH 5.0, even after incubation at 50 °C for 6 h (Fig. 3a). The enzyme remained active at pH 1.0 for 24 h at 50 °C. For the sample incubated at pH 3.0, the enhancement in activity was maintained for 10 h at 60 °C (Fig. 3b). Immediate total loss of activity was observed in the alkaline pH range at 50 °C.

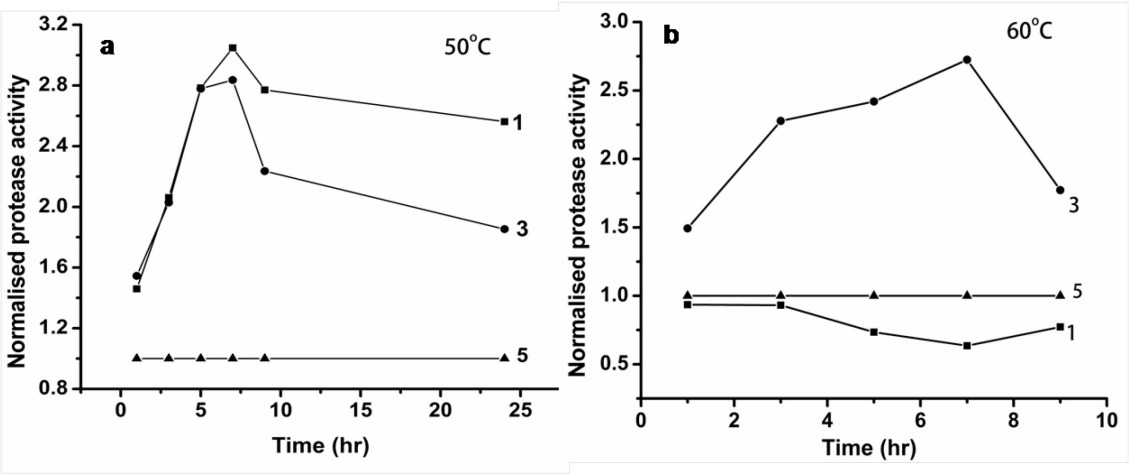


Figure 3: Stability profile of NprotI in acidic pH range at temperatures **a)** 50 °C and **b)** 60 °C.

In spite of being an alkaline serine endopeptidase, NprotI was found to be more stable and showed enhanced activity when incubated at acidic pH. pORF2 (protein from Open Reading Frame 2) of Hepatitis E Virus [14] is reported to acquired more secondary structure and its thermal stability was increased at acidic pH, which could be due to the presence of one or more pair(s) of carboxylate groups which would prevent destabilization due to repulsion between positive charges. For NAPase, a novel mechanism of acid resistance through charge migration has been proposed based on structural and mutational studies [4].

Protonation while going towards lower pH confers unusual properties to NprotI. The structure of the protein at pH 1.0 has stable tertiary structure with few hydrophobic side chains exposed. The structure showed enhanced catalytic rate when checked under standard assay conditions.

B. CD spectroscopy of NprotI at different pH and temperatures

Thermal denaturation of NprotI at pH 1.0 and pH 10.0 (pH for optimum activity) was monitored by recording far UV CD spectra at increasing temperature. At pH 1.0, the PPII fold was fairly intact, upto 70 °C. The isodichroic point at 221 nm was observed till 70 °C (Fig.4a). However, at pH 5.0, the native pH of the enzyme, PPII fold melts above 65 °C [3] (Fig. 4b). And at pH 10.0, the protein started melting above 60 °C (Fig. 4c).

Isodichroic point was not observed at pH 10.0, indicating more than two conformational forms involved in thermal denaturation process. Around 65 °C induction of helical-like structure was observed which got diminished with further increase in temperature. This was interesting as the optimum temperature and pH for the enzyme activity is 60 °C and pH 10.0.

Different modes of thermal denaturation were observed under three pH conditions. At extreme acidic pH i.e. at pH 1.0, the negative ellipticity at 212 nm, remained fairly intact at all the temperature conditions. While at pH 10.0 and pH 5.0, different response was observed. The positive ellipticity at 230 nm was decreased under all the pH conditions with increase in temperature. T_m values were calculated by sigmoidal fit of the MRE at 230 nm, as, 73 °C at pH 1.0, 71 °C at pH 5.0 and 61 °C at pH 10.0 (Fig. 4d).

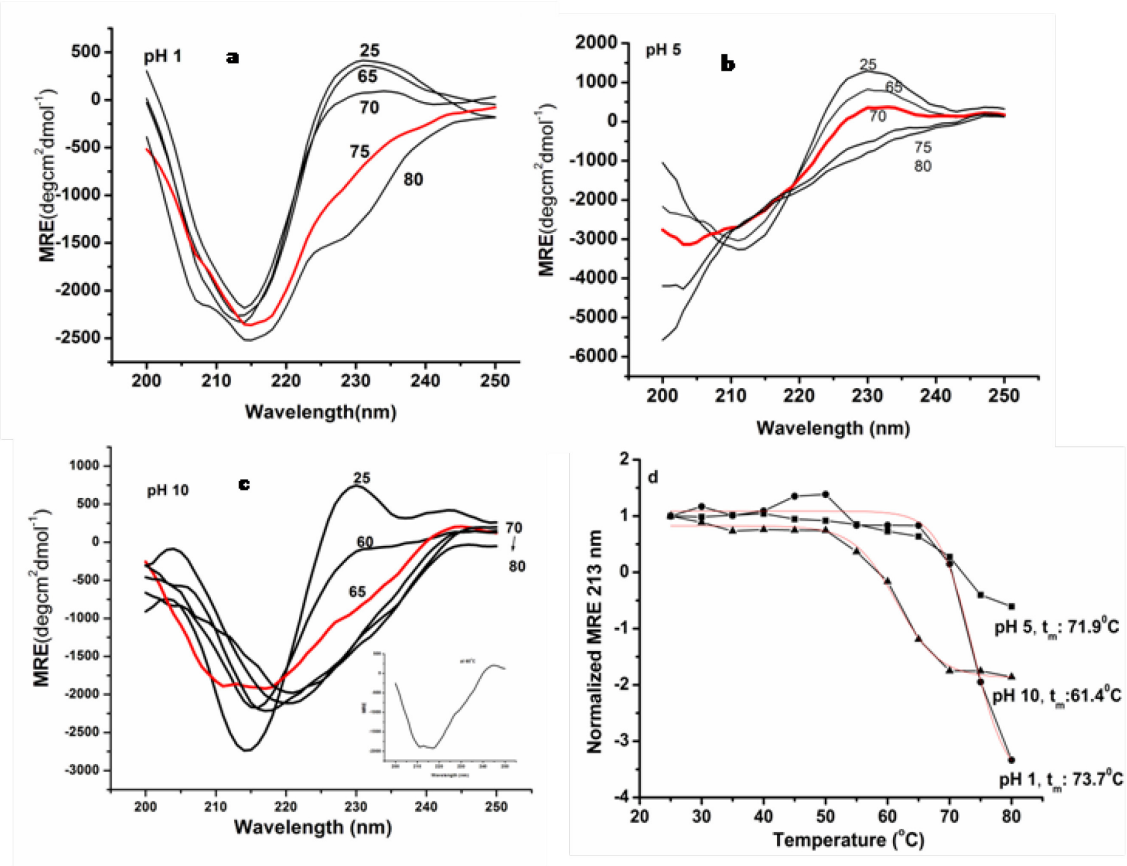


***Figure 4**: **Comparison of Far UV CD spectra of thermal denaturation of NprotI at pH** **a)** 1.0, **b)** 5.0 and **c)** 10.0 (inset shows structural transition at 65 °C), **d)** Sigmoidal fits of denaturation curves based on MRE 230 nm.*

C. Differential Scanning Calorimetry of NProtI at different pH

The DSC scans for NprotI were run at different pH (Fig. 5a and b). The scan at pH 5.0, the native pH, showed a T_m of 76.4 °C [4]. At pH 1.0, T_m of 72.7 °C was observed (Fig. 5a) indicating that the entity formed at pH 1.0 is comparatively stable. At pH 10.0, the optimum pH for NprotI activity, the T_m (63.8 °C) was much lower than that at pH 5.0 and pH 1.0 (Fig. 5b). The increase in enthalpy and a decrease in T_m could be due to change in structure of the protein (Fig. 5c).

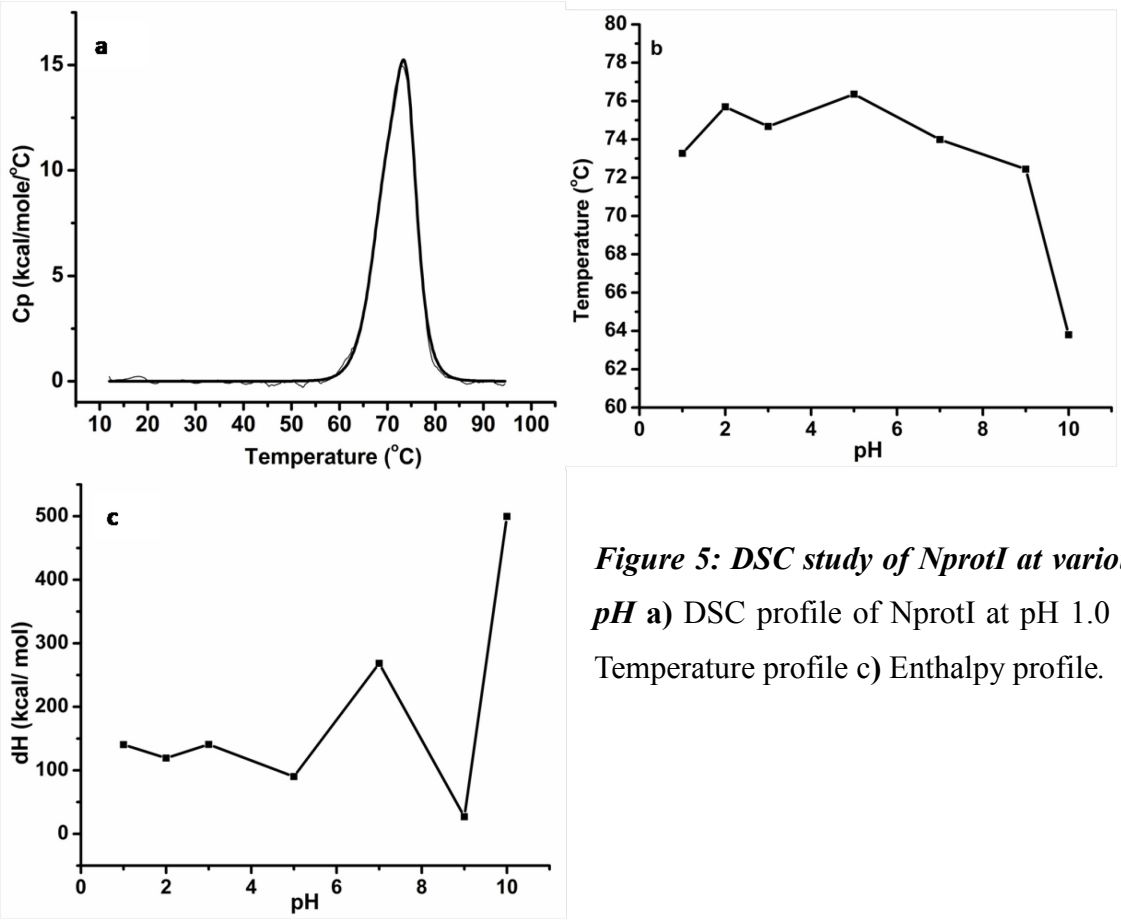


Figure 5: DSC study of NprotI at various pH **a)** DSC profile of NprotI at pH 1.0 **b)** Temperature profile c**)** Enthalpy profile.

The same trend was also followed by thermolysin, a neutral metalloproteinase from *Bacillus thermoproteolyticus rokko* [15]. For thermolysin, the enthalpy value increased and T_m decreased from pH 5.0 to pH 9.0. Possible deprotonation of amino acid side chains of NprotI at pH 10.0 could lead to changes in the active site geometry leading to most favorable structure for catalysis.

When the enzyme was incubated in presence of PMSF prior to DSC scan, the T_m value was

slightly increased from 76.4 °C to 78.2 °C indicating modification or blocking of the active site serine residue increases the stability of the protein.

3.4 Effect of GdnHCl on NprotI at pH 1.0

High stability of NprotI in the vicinity of 6 M GdnHCl at pH 5.0 and the role of the special structural feature of PPII fold therein have been established by us [3]. Here, the protein stability at pH 1.0 in presence of GdnHCl was examined (Fig. 6). Unfolding of the protein at pH 1.0 was observed at and above 3 M GdnHCl after incubation for 24 h (red shift in the λmax from 353 nm to 358 nm) which was not evident at pH 5.0 even in 6 M GdnHCl after incubation for 48 h [3].

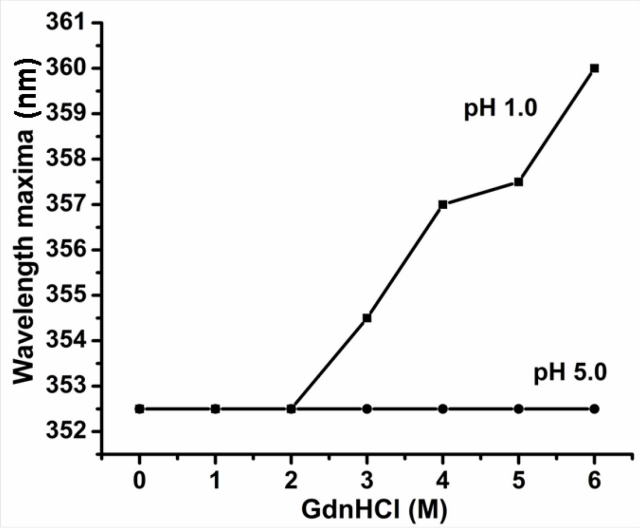


Figure 6: *Denaturation profile of NprotI treated with 0-6 M GdnHCl, at 25 °C, for 24 h.*
pH 1.0 (■); pH 5.0(●)

Fink et al have classified acid denatured states of proteins in three major types [16]. **Type I** proteins, when titrated with HC1 in the absence of salts, show two transitions, initially unfolding in the vicinity of pH 3.0-4.0 and then refolding to a molten globule-like conformation, the A state at lower pH. **Type II** proteins, upon acid titration, do not fully unfold but directly transform to the molten globule state, typically in the vicinity of pH 3.0. **Type III** proteins show no significant unfolding to pH as low as 1.0, but may behave similarly to type I in the presence of urea. Based on our observations, NprotI could be of type III protein. Other proteins of this type are T4 lysozyme, ubiquitin, chicken lysozyme, chymotrypsinogen, protein A, lactoglobulin and concanavalin A.

3.5 Effect of Guanidine thiocyanate on structure and activity of NprotI

As NprotI was resistant to GdnHCl at pH 5.0 [3]; GuSCN, a more powerful denaturant was used to unfold the enzyme. As seen in the activity profile (Fig.7a), initially there was activation in presence of GuSCN, later the activity dropped to 50 % in 20 min at 6 M concentration. After 24 h there was complete drop in activity in the vicinity of 5 M and 6 M concentration, 50 % activity was still retained in 2 M GuSCN after 7 days.

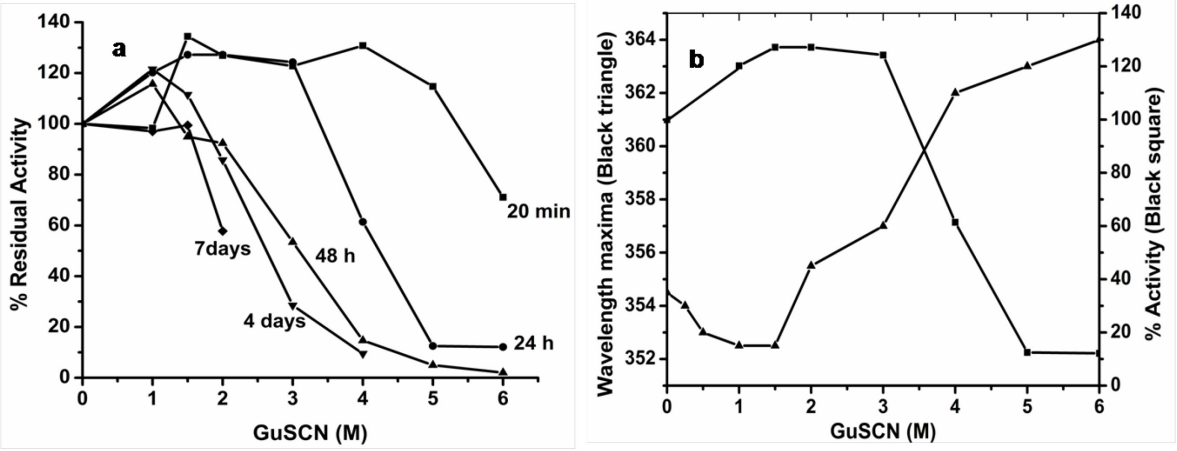


Figure 7**: **Effect of GuSCN on NprotI a) Activity profile of NprotI incubated with GuSCN (0-6 M) b) Fluorescence and activity based denaturation profile of NprotI incubated in GuSCN (0-6 M) at 25 °C.

Interestingly, the fluorescence studies in presence of GuSCN showed blue shift in λmax at lower concentration, with concomitant increase in activity (Fig.7b). As far as chemical denaturation is considered, NprotI could be denatured only with GuSCN at pH 5.0 and with GdnHCl at pH 1.0.

3.6 Effect of salts on NprotI and comparison to effect of GdnHCl

The enzyme was found to follow the reverse Hofmeister series (Fig. 8a) when the activity was studied in presence of various salts ($MgCl_2$, NaCl and NH_4Cl) from the series. The activity was found to decrease in the order; Mg^{2+} >Na^+ > NH_4^+. Long term stability in presence of Gdn^+, which is at the extreme end of the series, has been studied previously [3]. It has been studied that protein follow different Hofmeister series at pH above and below its pI [17, 18]. The pI of NprotI is about 8.3 [6] and the protein was incubated with different pI

pH 5.0 as this is the optimum pH.

The content of PPII fold seemed to be decreased in presence of all the cations (Fig. 8b), although the CD spectrum still possessed the PPII character. It was shown in a study by Drake el al that the PPII structure can be disrupted somewhat by the addition of sodium chloride, while still retaining the PPII character [19]. This was also shown by Rucker et al, in homopolymers of lysine [20]. When the lysine peptide was titrated with sodium chloride the PPII content was decreased but the characteristic PPII CD signal was still there. It was speculated that short lysine peptides adopt PPII helical structure as a result of the nature of the backbone rather than as a consequence of electrostatic interactions between side chains.

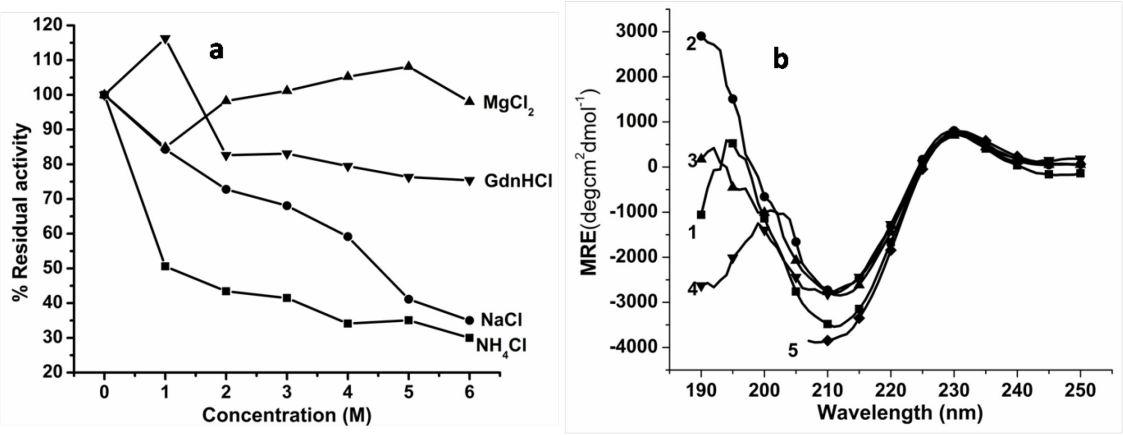


Figure 8: Effect of different cations on NprotI **a)** Activity profile **b)** CD profile of NprotI incubated with 1 M of different salts as: 1, native protein; 2, with NaCl; 3, with NH_4Cl; 4, with $MgCl_2$; 5 with GdnHCl.

The ions have ionic radii as NH_4: 175 pm, Na: 116 pm, Mg: 86 pm; hence the lowest activity in NH_4^+ could be due to the effect of larger sized ammonium ions disrupting the electrostatic interactions in the protein. Interestingly Gdn^+ (210 pm) showed stabilizing effect on the structure and activity of the enzyme. This effect could be due to ability of Gdn^+ to increase the structural content of PPII fold, while in the rest of the ions this effect might be lower.

4. References

1. E L Cunningham, S S Jaswal, J L Sohl, D A Agard. *Proc Natl Acad Sci.* **1999**, 96: 11008–11014.
2. S C Yadav, M V Jagannadham. *Eur Biophys J.* **2009,** 38: 981–990.
3. S B Rohamare, V S Dixit, P K Nareddy, D Sivaramakrishna, M J Swamy, S M Gaikwad. *Biochim Biophys Acta.* **2013**, 1834: 708-716.
4. B A Kelch, K P Eagen, F P Erciyas, E L Humphris, A R Thomason, S Mitsuiki, D A Agard. *J Mol Biol.* **2007,** 368: 870–883.
5. J B Matthew. *Annu Rev Biophys Chem.* **1985,** 14: 387–417.
6. V S Dixit, A Pant. *Biochim Biophys Acta.* **2000**, 1523: 261-268.
7. P X Qi, E D Wickham, H M Farrell Jr. *Prot J.* **2004,** 23: 389-402.
8. A L Rucker, T P Creamer. *Protein Sci.* **2002,** 11: 980–985.
9. A Shimizu, K Kawai, M Yanagino, T Wakiyama, M Machida, K Kameyama, Z Naito. *J Biochem.* **2007**, 142: 33–40.
10. O B Ptitsyn. *J Protein Chem.* **1987,** 6: 277–293.
11. J J Ewbank, T E Creighton. *Nature*. **1991,** 350: 518 - 520.
12. J Buchner, M Renner, H Lilie, H J Hinz, R Jaenicke, T Kiefhaber, R Rudolph. *Biochemistry.* **1991**, 30; 6922- 6929.
13. S K Haq, M F Ahmad, R H Khan. *Biochem Biophys Res Commun.* **2003,** 303: 685–692.
14. M Zafrullah, Z Khursheed, S Yadav, D Sahgal, S Jameel, F Ahmad. *Biochem Biophys Res Commun.* **2004,** 313: 67–73.
15. F Conejero-Lara, A I Azuaga, P L Mateo. *React Funct Polym.* **1997,** 34: 113-120.
16. A L Fink, L J Calciano, Y Goto, T Kurotsu, D R Palleros. *Biochemistry.* **1994**, 33: 12504-12511.
17. M Bostrom, F W Tavares, S Finet, F Skouri-Panet, A Tardieue, B W Ninham. *Biophys Chem.* **2005**,117:217 – 224.
18. S Finet, F Skouri-Panet, M Casselyn, F Bonnete´, A Tardieu. *Curr Opin Colloid Interface Sci.* **2004**, 9:112 – 116.
19. A F Drake, G Siligardi, W A Gibbons. *Biophys Chem.* **1988,** 31:143-146.
20. L A Rucker, T P Creamer. *Prot Sci.* **2002**, 11:980–985.

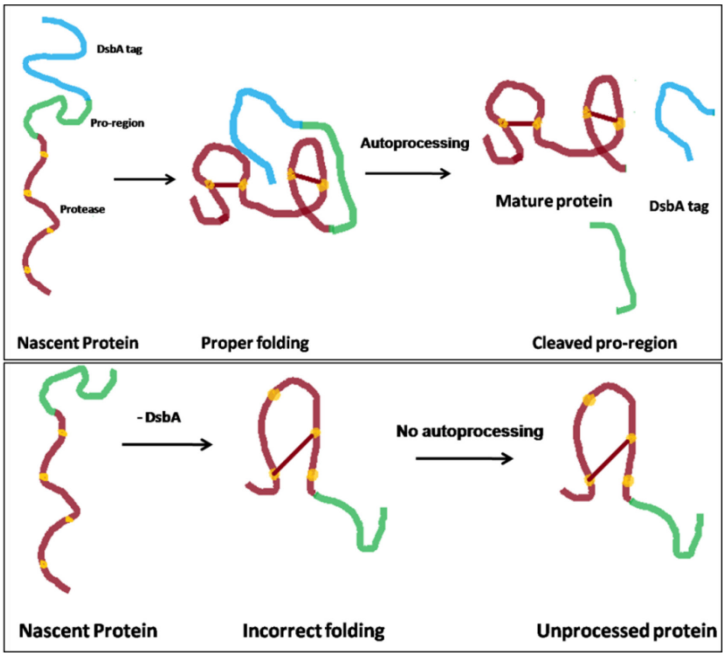
Chapter 5

Cloning and expression of a serine protease from Nocardiopsis sp. NCIM 5124

Cloning, expression and in silico studies of a serine protease from Nocardiopsis sp. NCIM 5124 (Manuscript communicated).

Summary

A serine protease (named as 'N.protease'), from a marine isolate of *Nocardiopsis sp.*, was cloned, expressed in *E.coli* and investigated for its potential kinetic stability. Expression was carried out using two vectors, pET-22b (+) and pET-39b (+) and comparisons were made based on proper folding and soluble expression of the protein. Studies into the role of incubation temperature for *E.coli* growth and inducer concentration on solubility of expressed protein were undertaken. pET-39b (+) was found to be a better vector for the soluble expression of this disulfide bonds containing protease. *In silico* studies were carried out for N.protease. Homology modelling of N.protease suggested it to be a part of PA clan of proteases. High temperature MD simulations were performed on N.protease to study its unfolding behavior and comparisons were made with alpha lytic protease (αLP), a model kinetically stable protein. A novel approach to study cooperativity of protein unfolding was undertaken, wherein 'P' value analysis based on φ and ψ values of the protein was performed. Gradual transition in P value analysis for N.protease suggested it to be comparatively less kinetically stable than αLP. Present study holds significance as the non-strepetomycete actinomycetes group is least explored and holds a promise for industrially important enzymes with exceptional stabilities.



Graphical abstract

1. Introduction

Proteases have been widely studied; sometimes as a model protein for enzyme catalysis and structure, as targets for drug development, or as reagents for green chemistry. Proteases are involved in virtually every physiological process. Serine proteases, with the classic Ser-His-Asp triad are the largest class of proteases. Controlling and regulating the protease activity has been of interest since long. Nature has its own ways of regulation; either there are inhibitors of proteases to control their activity or the proteases are synthesized as inactive precursors and are processed to mature proteases. Many of the bacterial proteases, which are needed to work in extracellular environment, are highly stable to various denaturing conditions and are called as kinetically stable proteins [1]. Most of these proteases are synthesized with a prodomain, which helps them to fold properly and is then removed autocatalytically to release the mature enzyme [2]. These prodomains are N-terminal extensions that are large enough to fold independently [3].

The unique feature of proteases is that although they catalyze the same reaction, their sequence, structure and stability differs with their source. Proteases from different sources are being studied widely to understand structural features responsible for their exceptional stabilities. E.g. alpha lytic protease (αLP) system and NAPase (*N. alba* protease) have been previously reported to possess the kinetic stability [4,5]. The kinetically stable proteases are difficult to unfold. If unfolded, their unfolding is irreversible. The pro-region (which is expressed along with these proteases) helps these proteases to attain their native conformation. Moreover unfolding transitions of such proteases are reported to be "cooperative" i.e. the unfolding event is an all or none process [6].

The biochemical heterogeneity, ecological diversity and the excellent capacity of secondary metabolite production of actinomycetes, makes them a potential source of enzymes with new activities and/or specificities; which has remained unexplored yet [7,8]. As, some of the actinomycetes grow in extreme environmental conditions, the enzymes produced by them are exceptionally stable [9,10]. Though, the enzymes from streptomycetes group of actinomycetes have been very well studied, the enzymes

from non-streptomycetes actinomycetes have not been studied thoroughly.

In the present study, an attempt has been made to clone and express a serine protease from *Nocardipsis sp.* NCIM 5124. There are many reports of production and purification of serine proteases from various actinomycetes [8,11]. But, these proteases have not been studied at molecular level with few exceptions such as the serine protease from *Nocardiopsis alba* (NAPase) which has been studied extensively for its high acid stability and NprotI from *Nocardipsis sp.* NCIM 5124 [12] have been studied for its kinetic stability.

Due to the inherent ability of proteases to cleave peptide bond, proteases represent one of the most challenging functional group of proteins for heterologous expression and purification. The proteases expression often results in formation of inclusion bodies, non-expression or cytotoxicity if they are produced unregulated in *E. coli* host cells as the cell suffers from critical stress. Hence, the molecular level study of proteases is challenging. For proper expression and purification of recombinant proteins, the fusion tags used in the expression vector are crucial components. Numerous tags are available which assists in enhancement of expression and solubility, immobilization, detection, quantification, and purification of proteins.

In this paper, we report the cloning and expression of a serine protease from *Nocardiopsis sp.*, containing the N-terminal prodomain and the mature protease domain, in *E. coli.* For expression studies we have used two different vectors pET-22b (+) and pET-39b (+) [13], both targeting proteins in the periplasmic space and both having a histidine tagging for easy purification. Additionally pET-39b (+) has a disulfide oxidoreductase A (DsbA) tag which helps in disulphide bond formation. A comparison between the two vectors has been done with respect to proper folding of the protease with disulphide bonds. The effect of growth temperature of *E. coli* on the maturation and solubility of the proteases was studied. Also*, in silico* studies have been carried out to study whether the protease in present study is a kinetically stable protein as some other proteases from microbial origin.

2. Materials and methods

2.1 Genomic DNA extraction of *Nocardiospsis sp.* NCIM 5124

The organism was isolated from an oil contaminated marine site near Mumbai harbor (India) [14]. The genomic DNA from this organism was isolated using CTAB extraction protocol to avoid any polysaccharide contamination [15].

2.2 Cloning and sequencing

Primers were designed based on the gene sequences of serine proteases from *Nocardiopsis dassonivilie,* as the whole genome sequence of the organism is known [16]. Following sets were designed:

No.	Forward Primer (5'-3')	Reverse Primer (5'-3')
1	AGCGTTACCCGGTCACCA	AGGGATCTATGCGACCCTCC
2	GAGCGTTACCCGGTCACCA	ATCTATGCGACCCTCCCCC
3	CGGTCACCAGGGACAGC	TAGGGATCTATGCGACCCTCC
4	CGGAGCGTTACCCGGTC	GATCTATGCGACCCTCCCCC
5	CCCGGTCACCAGGGACA	GGATCTATGCGACCCTCCCC

PCR amplification was done using these primers and genomic DNA from *Nocardiospsis sp.* NCIM 5124 as a template. The amplified fragments were sequenced and were then assembled using CAP3 program [17]. As the sequence was highly GC rich, it was codon optimized for expression in *E. coli* and then the assembled gene was synthesized from Life Technologies^{TM}, UK, into cloning vector pMA between restriction sites BamHI and XhoI.

2.3 Subcloning in expression vector

The cloning vector pMA with the gene sequence was digested with BamHI and XhoI endonucleases (NEB, USA). These fragments were purified using the Qiagen PCR Purification Kit (Qiagen, Valencia, CA,USA) and then ligated into the BamHI and XhoI sites of the pET-22b (+) and pET-39b (+) vectors which were digested with the same enzymes, generating two fusion expression vectors pET22b-pro-prot and

pET39b-pro-prot, where "pro" stands for pro-region. The resultant plasmids were transformed into *E. coli* strain DH-5α. The correct insertion was confirmed by DNA sequencing, restriction digestion and colony PCR.

2.4 Heterologous expression and purification of pro-prot in *E. coli* pET22b-pro-prot and DsbA-pro-prot in pET39b-pro-prot

The screened plasmids pET22b-pro-prot and pET39b-pro-prot were transformed into the *E. coli* strain BL-21(DE3) for expression. Expression of the recombinant protein was performed according to pET system manual (Novagen), induced by adding isopropyl β-D-thiogalactopyranoside (IPTG). The expression studies were carried out with inducer concentration of 1 mM to 0.1 mM and growth temperature of 37 °C to 16 °C to optimize the conditions for getting soluble expression. Cells were grown overnight after induction and were harvested by centrifugation at 4500×g for 15 min at 20 °C. The cell pellet obtained from 500 ml bacterial culture was resuspended in 15 ml of 20 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 mM phenyl methane sulphonyl fluoride (PMSF), and subjected to sonication at 4 °C for 20 cycles (each cycle consisting of 2 s on and 3 s off times). The total cellular proteins were then partitioned into soluble and insoluble fractions by centrifugation at 12,000×g for 45 min at 4 °C and the protein expression was analyzed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). It was speculated that during maturation of the proteases the tags will get cleaved off along with prodomain, so this processing was monitored with SDS-PAGE.

The purification by Nickel nitrilotriacetic acid (Ni-NTA) column was carried out at room temperature. The soluble fraction was batch absorbed for 90 min with Ni-NTA agarose (Qiagen, Germany) pre-equilibrated with Tris-HCl at room temperature. The resin was loaded on a glass column and washed with 5 column volume of TrisHCl with 20 mM imidazole to remove non-specifically bound contaminants. Finally the protein was eluted by passing the same buffer with 300 mM imidazole. Purification and elution were monitored by SDS-PAGE.

2.5 Western blot analysis

Expression of recombinant pro-prot and DsbA-pro-prot *E. coli* BL21 (DE3) were confirmed by Western blot analysis with rabbit antiserum raised against Histidine tag (Bangalore Genei, India) as primary antibody. For Western blot analysis, a standard protocol where the SDS-PAGE (12 %) separated proteins were transferred to a nitrocellulose membrane (Hybond-ECL, Amersham Biosciences, Germany), blocked with 3 % bovine serum albumin, incubated with primary antibody and followed by secondary antibody (conjugated with Horseradish Peroxidase) were performed. Finally, the blot was detected using chemiluminiscent substrate. The processing of the DsbA-pro-prot to get mature protein was also monitored with western blot analysis.

2.6 Molecular modeling studies

A similarity search carried out using the Basic Local Alignment Search Tool (BLAST) server [18], identified in Protein Data Bank (PDB) (http://www.rcsb.org), the NAPase; a serine protease from *Nocardiopsis alba*, (PDB ID: 2OUA, Resolution: 1.85Å) which had 85 % identity with the translated gene sequence of mature N.protease. The 3D model of prot was built with the MODELLER 9.10 program [19] using above homologus PDB structure as a template. Separate models were built for prodomain and mature protease. The N-terminal prodomain was submitted to the SWISS Model server (http://expasy.org/spdbv/) for comparative modelling employing the prodomain of alpha lytic protease (PDB ID: 2PRO, 3.00 Å) as template which had 25 % identity to the prodomain of N.protease. The molecular models were evaluated using Ramchandran plot obtained from PROCHECK (http://services.mbi.ucla.edu/SAVES/), Errat (version 2.0) [20], PDBsum [21] and ProSA-web Protein structure analysis [22].

2.7 Phylogenetic analysis

MEGA 6.0 software (http://www.megasoftware.net) was used for the phylogenetic analysis [23] of protein sequence of N.protease. Various sequences of proteases were obtained from National Centre for Biotechnology Information (NCBI). These protein sequences were initially aligned using **MU**ltiple **S**equence **C**omparison

by **L**og- **E**xpectation (MUSCLE). This alignment was used to build the phylogenetic tree using the Neighbor Joining (NJ) algorithm and the bootstrap was calculated using 1000 replications.

2.8 Molecular Dynamics Simulations

The thermal MD simulations were carried out for N.protease as well as for αLP, for comparison. The molecular model generated using MODELLER program was used for N.protease and 1SSX PDB was used for αLP. The structures were subjected to molecular dynamics simulations using the GROningen Machine for Chemical Simulations V4.5.4 (GROMACS) [24]. GROMOS96 43a1 force field was applied on the protein molecule placed in the centre of the dodecahedron box solvated with SPCE explicit water model with distance between the solute and the box set to 10 Å. The system was initially energy minimized by steepest descent minimization for 50,000 steps until a tolerance of 10 KJ/mol reach. In order to neutralize the whole system, total charge on the system was balances by adding 3 chloride counter-ions using genion program of GROMACS. After adding ions, the system was again energy minimized by steepest descent minimization retaining the same parameters. A cut off distance of 9 Å and 14 Å was set for Coulombic and van der Waals interactions. Simulations were carried out with periodic boundary conditions in all the three directions. A timestep of 2 fs was used and snapshots were saved every 2 ps. The solvent was equilibrated for 100 ps under NPT conditions using Berendsen coupling for both pressure (100 fs relaxation time) and temperature (2.0 ps coupling constant). For N.protease, 3 final runs of 10 ns at 298 K, 500 K and for αLP, 3 runs at 500 K and 1 at 298 K were performed. The trajectories were visually inspected using Visual Molecular Dynamics program (VMD) [25].

2.9 Structural persistence

The structural persistence parameter was used to monitor the changes in secondary structure with respect to the reference structure [26, 27]. To summarize briefly,

$$P = \frac{1}{N_{res}} \sum_{j=1}^{N_{res}} e^{-(\Delta\varphi_j / \Delta\varphi_{max})} . e^{-(\Delta\psi_j / \Delta\psi_{max})}$$

Here, $\Delta\varphi_j$ and $\Delta\psi_j$ respectively, are the changes in the φ and ψ torsional angles of residue j over the reference at a given point in time, and $\Delta\varphi_{max}$ and $\Delta\psi_{max}$ represent the maximal changes that can occur in the torsional angles. N_{res} represents the total number of amino acid residues in the protein. A value of P = 1 is observed for a conformation whose secondary structure is exactly same to the reference. This parameter was calculated for both N.protease and αLP.

3. Results and discussion

3.1 Cloning and sequencing

PCR amplification using the set of primers described earlier resulted in various sized fragments of gene sequence. The sequence obtained in fragments was assembled using CAP3 program to yield a full length sequence of 1155 bp, coding for signal peptide, prodomain (pro) and mature protease (prot) (Fig. 1). The gene sequence was sunbmitted to genbank under accession no. of KM268817. The 384 amino acid protein, codified from the ORF, contained three parts: a signal peptide of 29 amino acids, an N-terminal propeptide of 167 amino acids, mature enzyme of 188 amino acids. The estimated molecular mass of the mature protein is around 20 kDa with theoretical pI of 8.86 (http://web.expasy.org/cgi-bin/compute_pi/pi_tool). The protein was found to have 3 disulfide bonds as studied by PDBsum program. We got this gene synthesized from Life Technologies™ (UK), codons for the first 29 amino acids coding for signal peptide and the stop codon were excluded from the synthesis. The signal peptide from *E. coli* was used for effective expression. The gene was synthesized in pMA vector in between BamHI and XhoI restriction sites.

3.2 Subcloning in expression vector

The gene was successfully subcloned in expression vectors pET-22b (+) and pET-39b (+) (Fig.2). These created a fusion with six histidines and a pelB tag for pET22b-pro-prot and the transcript was named as pro-prot; for pET39b-pro-prot, a DsbA tag, six histidines and S tag were attached at the N-terminus of pro-prot and was named as DsbA-pro-prot.

```
atgagaccctccagcatcatctccgccgtgggcacaggagccctggccttcggtatggca
 M  R  P  S  S  I  I  S  A  V  G  T  G  A  L  A  F  G  M  A
ctggccatggccccccggagccctcgcggccccggcccccgtccccccagacccccgtcgcc
 L  A  M  A  P  G  A  L  A  A  P  A  P  V  P  Q  T  P  V  A
gacgacagcgccgccagcatgacggaggcgctcaagcgcgacctcgacctcacctcggcc
 D  D  S  A  A  S  M  T  E  A  L  K  R  D  L  D  L  T  S  A
gaggccgaggagctgctctcggcgcaggaagccgccatcgagaccgacgctgaggccacc
 E  A  E  E  L  L  S  A  Q  E  A  A  I  E  T  D  A  E  A  T
gaggccgcgggcgaggcctacggcggctccctgttcgacaccgagaccctcgaactcacc
 E  A  A  G  E  A  Y  G  G  S  L  F  D  T  E  T  L  E  L  T
gtgctggtgaccgactcctcggccgtcgaggccgtcgaggccaccggcgccgaggccacc
 V  L  V  T  D  S  S  A  V  E  A  V  E  A  T  G  A  E  A  T
gtggtctcacacggcaccgagggcctggccgaggtcgtcgaggacctcaacggcgccgag
 V  V  S  H  G  T  E  G  L  A  E  V  V  E  D  L  N  G  A  E
gctcccgcgagcgtcctcggctggtacccggacgtggagagcgacaccgtcgtggtcgag
 A  P  A  S  V  L  G  W  Y  P  D  V  E  S  D  T  V  V  V  E
gtgctggagggctccgacgccgacgtcgccgccctgctcgccgacgccggtgtggacgcc
 V  L  E  G  S  D  A  D  V  A  A  L  L  A  D  A  G  V  D  A
tcctcggtccgggtggagaaggccgaggaggccccgcaggtctacgccgacatcatcggc
 S  S  V  R  V  E  K  A  E  E  A  P  Q  V  Y  A  D  I  I  G
ggcctggcctactacatgggcggccgctgctccgtcggcttcgccgcgaccaacagcgcc
 G  L  A  Y  Y  M  G  G  R  C  S  V  G  F  A  A  T  N  S  A
ggccagcccggtttcgtcaccgccggccactgcggcaccgtcggcaccagcgtgaccatc
 G  Q  P  G  F  V  T  A  G  H  C  G  T  V  G  T  S  V  T  I
ggcaacggccgcggcaccttccagaactcggtcttccccggcaacgacgccgccttcgtc
 G  N  G  R  G  T  F  Q  N  S  V  F  P  G  N  D  A  A  F  V
cgcggcacctccaacttcaccctgaccaacctggtctcgcgctacaactccggcggctac
 R  G  T  S  N  F  T  L  T  N  L  V  S  R  Y  N  S  G  G  Y
cagtcggtgtccggcgccagccaggccccggccggctcggccgtgtgccgctccggctcc
 Q  S  V  S  G  A  S  Q  A  P  A  G  S  A  V  C  R  S  G  S
accaccggctggcactgcggcaccatccaggcccgcaaccagaccgtgcgctacccgcag
 T  T  G  W  H  C  G  T  I  Q  A  R  N  Q  T  V  R  Y  P  Q
ggcaccgtctactcgctcacccgcaccaacgtgtgcgccgagcccggtgactccggcggc
 G  T  V  Y  S  L  T  R  T  N  V  C  A  E  P  G  D  S  G  G
tcgttcatctccggctcgcaggcccagggcgtcacctccggcggctccggcaactgctcc
 S  F  I  S  G  S  Q  A  Q  G  V  T  S  G  G  S  G  N  C  S
gtcggcggcacgacctactaccaggaggtcacgccgatgatcaactcctggggcgtcatg
 V  G  G  T  T  Y  Y  Q  E  V  T  P  M  I  N  S  W  G  V  M
aaccgcaccagctga
 N  R  T  S  –
```

Figure 1: The complete gene and amino acid sequence. The signal peptide is shown in red, prodomain in blue and mature protease region in black.

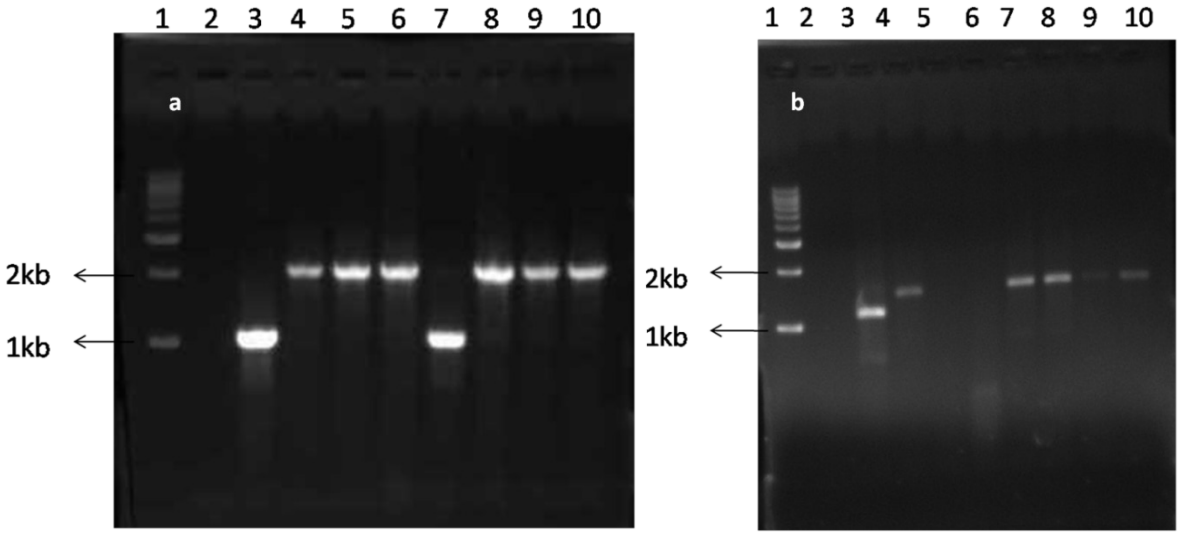


***Figure 2**: **Colony PCR to confirm correct gene insertion in the vector* a) 39b**, lane no.

4-6 & 8-10 shows colonies with correctly inserted gene **b) 22b**, lane no. 4 and 7-10 shows colonies with correctly inserted gene.

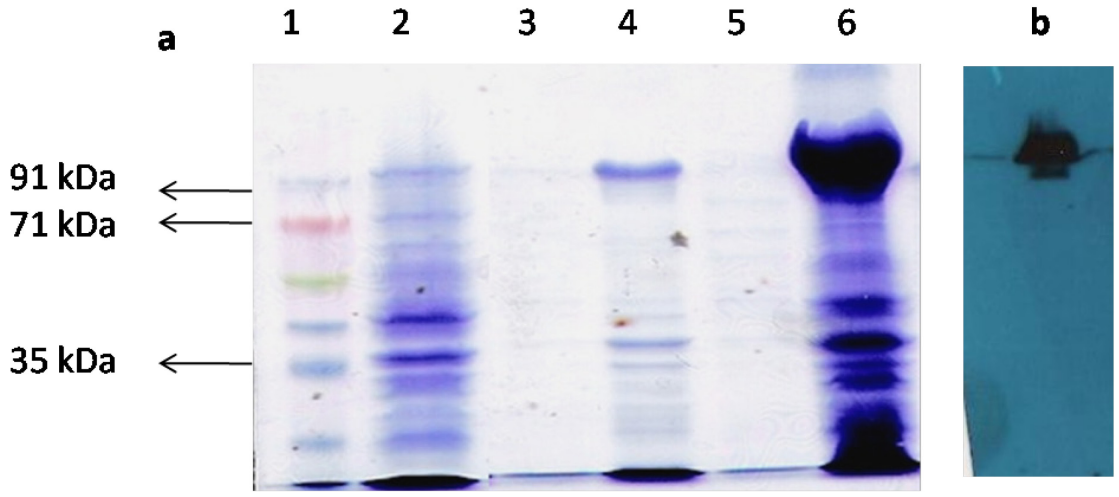


Figure 3: Protein expression using pET-39b(+) at 37°C, with 1 mM IPTG induction **a)10 % SDS PAGE.** Two clones (39b1 and 39b2) were checked for expression. Samples were loaded as followed; 1: prestained protein ladder, 2: uninduced whole cell fraction, 3: 39b1 soluble fraction, 4: 39b1 insoluble fraction, 5: 39b2 soluble fraction, 6: 39b2 insoluble fraction. **b) Western blot analysis** with anti His tag antibody showing unprocessed protein at 91 kDa in insoluble fraction.

3.3 Heterologous expression and purification of pro-prot in *E. coli* pET22b-pro-prot and DsbA-pro-prot in pET39b-pro-prot

As described earlier the expression was optimized, w.r.t. growth temperature after induction and the inducer concentration, to get the protein in soluble fraction. Initially with 1 mM concentration of IPTG and growth of *E. coli* at 37 °C (Fig. 3), led to protein getting expressed in insoluble fractions in both the vectors. Also maturation of the protein was not seen, as the over expressed protein was of higher molecular weight as seen from SDS PAGE. The mature protein was estimated to be of 20 kDa. The same findings were observed even at growth temperature of 28 °C and 0.5 mM concentration of inducer. But when the protein expression was checked at 16 °C and 0.1 mM concentration, the protein was found to express in soluble form for pET39b-pro-prot.

Also autoprocessing of the protease was seen very clearly in SDS PAGE as well as in western blot analysis. The soluble mature protein was partially purified using Ni-NTA matrix (Fig. 4).

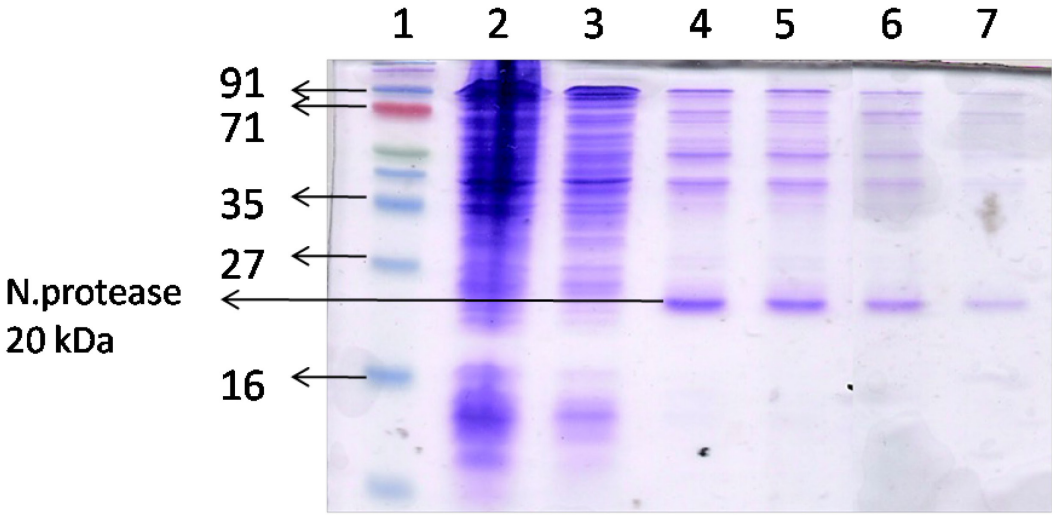


Fig. 4: Partial protein purification using IMAC (Ni NTA) column after expression of the protein in pET-39b, 16 °C, 0.1 mM IPTG. Soluble fraction was loaded on the column. Samples were loaded as followed; 1: prestained protein ladder, 2: Flow through from the column, 3: Column wash, 4-7: Column elutions in 300 mM imidazole.

Three stages of protease autoprocessing were seen (Fig.5). DsbA-pro-prot, first one with highest molecular weight around 91 kDa, then, pro-prot, was seen at 35 kDa, from which the DsbA tag was cleaved off but prodomain was still present in the expressed protein. Then prot was seen around roughly around 21 kDa from which both the prodomain and DsbA tag were cleaved off.

The protein expressed in pET-22b was in soluble form (Fig. 6) at the temperature of 16 °C and at 0.1 mM IPTG but autoprocessing of the enzyme was not seen as the overexpressed protein was seen at 35 kDa, i.e. with the prodomain. These observations suggested that protein was not properly folded when expressed in pET22b and hence, no autoprocessing was seen. The DsbA tag in pET-39b was important in catalyzing disulphide bond formation in the protease. And hence the protease was properly folded and autoprocessed.

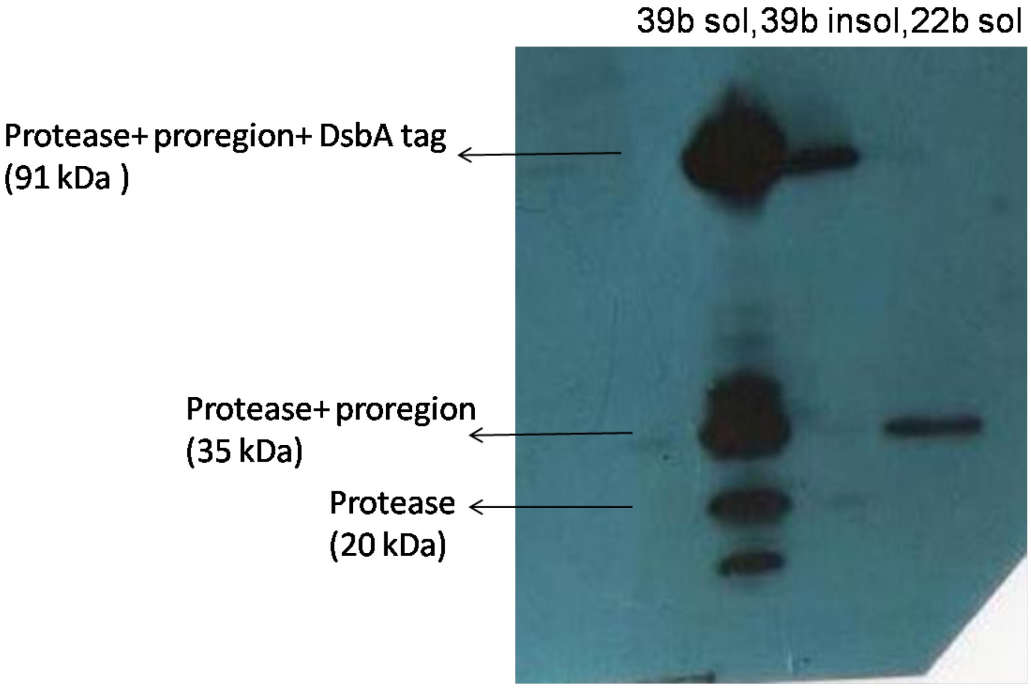


Figure 5: Western blot analysis showing different stages in protease processing. The soluble fraction after expression in pET-39b showed different stages in protease maturation, while the soluble fraction of protein expressed in pET-22b showed unprocessed protein at 35 kDa.

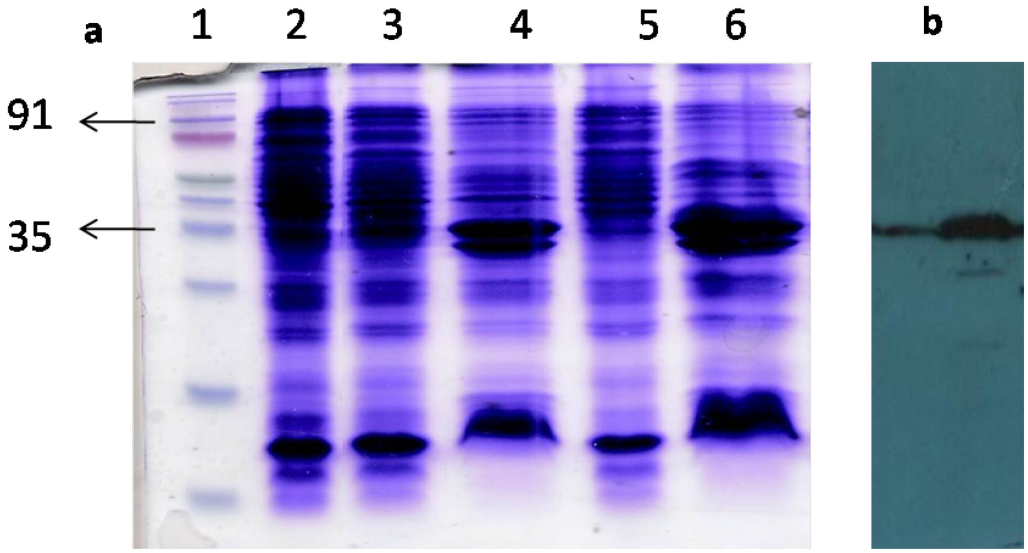


Figure 6: Protein expression using pET-22b at 16 °C, with 0.1 mM IPTG induction.

a) 15 % SDS PAGE. Two clones (22b1 and 22b2) were checked for expression. Samples were loaded as followed 1: prestained protein ladder, 2: uninduced whole cell

protein, 3: 22b1 soluble fraction, 4: 22b1 insoluble fraction, 5: 22b2 soluble fraction, 6: 22b2 insoluble fraction. **b) Western blot analysis** with anti His tag antibody showing the expression using pET-22b as unprocessed form at 35 kDa in both soluble and insoluble fractions.

3.4 *In silico* studies

Initially, the sequence of N.protease was analysed using Clustal X [28] (Fig. 7). It can be observed that the sequence of active protease region showed more homology than the pro-region to other proteases. Interestingly, it has been studied that the pro-regions of different extracellular proteases share structural similarity [29] but their active protease domains might not be similar; e.g. for alpha lytic protease, subtilisins and carboxypeptidase A, the secondary structure in the C-terminal portion of pro-region is helix–b-hairpin–helix–strand, with an extended region following the last strand.

3.4.1 Homology modeling

A. Construction of model

A model for N.protease was generated taking crystal structure of NAPase (PDB ID: 2OUA) as a template. NAPase is a kinetically stable protein from *Nocardiopsis alba* studied for its high acid stability. The major structural elements of the protein structure have been labeled in the Fig. 8a. The model evaluation parameters indicated that the model is of good quality (Fig.9). The protein structure from the model can be visualized by presence of two barrels which are arranged perpendicular to each other and the residues of catalytic triad are split between the two barrels. These are the typical structural properties of PA clan proteases [30]. The PA clan (Proteases of mixed nucleophile, superfamily A) is the largest group of proteases with common ancestry as identified by structural homology. Members of this clan have a chymotrypsin-like fold and similar proteolysis mechanisms but sequence identity of <10 %. N.protease demonstrated sequence similarity with various extracellular proteases. The gene was also found to have signal sequence for extracellular secretion; therefore it was envisaged to be a member of S1A family of PA clan proteases. The peptidases from S1A

family are involved in extracellular protein turnover and are mainly present in eukaryotes with a limited distribution in plants, prokaryotes and archaea [31].

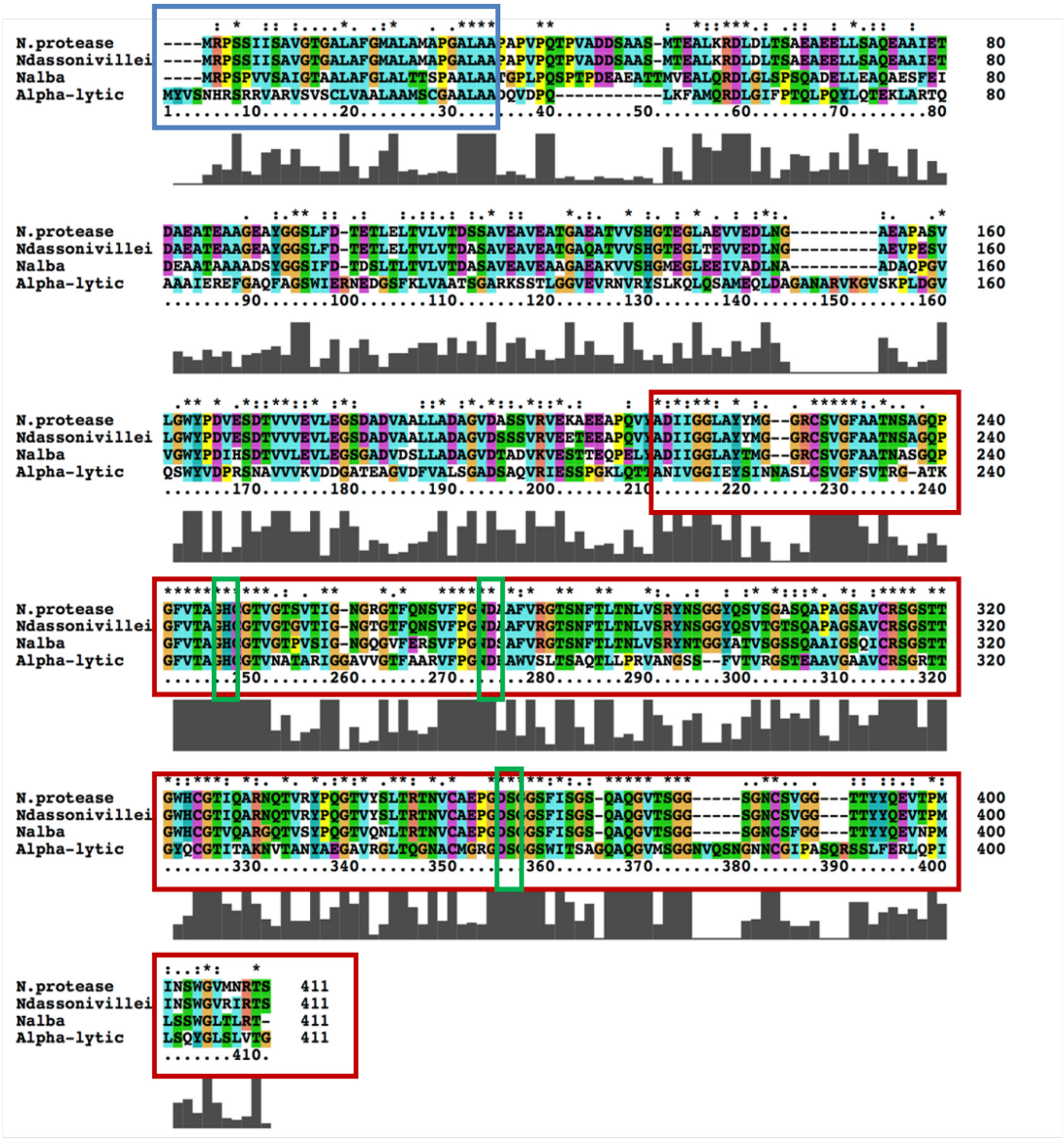


Figure 7: Protein sequence alignment in using program Clustal X. Protease in present study has been named N.protease in the above diagram. The blue box indicates signal peptide, the red box indicates active protease and remaining portion is pro-domain. The conserved catalytic triad has been showed in green boxes.

A model for prodomain was generated (Fig. 8b) in which the prodomain can be seen as a C shaped molecule with two domains; C terminal and N terminal domains, attached

with a hinge.

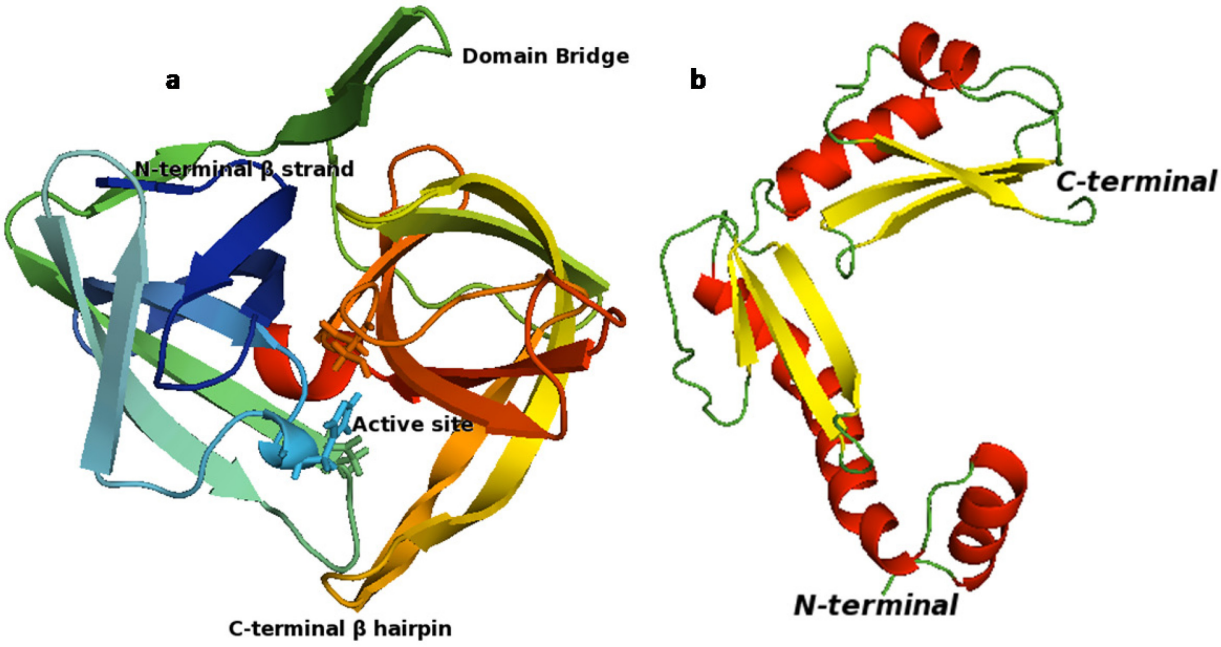


Figure 8: Homology modeling **a)** Model structure of N.protease generated using 2OUA as a template. The molecule is colored dark blue at the N-terminus progressing to red at the C-terminus. Important structural elements are labeled **b)** Model structure of pro-domain generated using 2PRO as a template.

B. Validation of the protein structure model

Stereochemical quality of the polypeptide backbone and side chains of the N.protease model was evaluated using the **Ramachandran plot** obtained from PROCHECK. It showed that 92.5 % of the residues are located in the most favoured region, 6.8 % in additionally allowed region and 0.0 % in generously allowed region of the Ramachandran plot (Fig. 9a). 0.7 % residues were found in the disallowed regions.

The **ProSA tool** was used to check the overall model protein structure for potential errors. The ProSAII program (Protein Structure Analysis) is an established tool, which is frequently employed in structure prediction, refinement and validation of experimental protein structures. It generates Z score of model, which is a measure of compatibility between its sequence and structure. Z score value obtained for the N.protease model (-6.1) indicated its location within the space of protein related to

X-ray (dark black dot in Fig. 9b).

The **ERRAT program** verifies the quality of the model. This program plots error values as a function of position in the sequence by sliding a nine residue window. The error function was based on the statistics of non-bonded atom-atom interactions in the template structure. The overall quality factor of modelled N.protease in ERRAT analysis was 94.94, expressed as the percentage of the protein for which the calculated error value falls below the 95 % rejection limit (Fig. 9c). On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value. Regions of the structure that can be rejected at the 95% confidence level are gray and regions that can be rejected at the 99% level are shown in dark black. Thus, validation results suggested that the predicted model was a reliable 3D structure of N.protease. Similary the model of pro-domain was validated (Fig. 10).

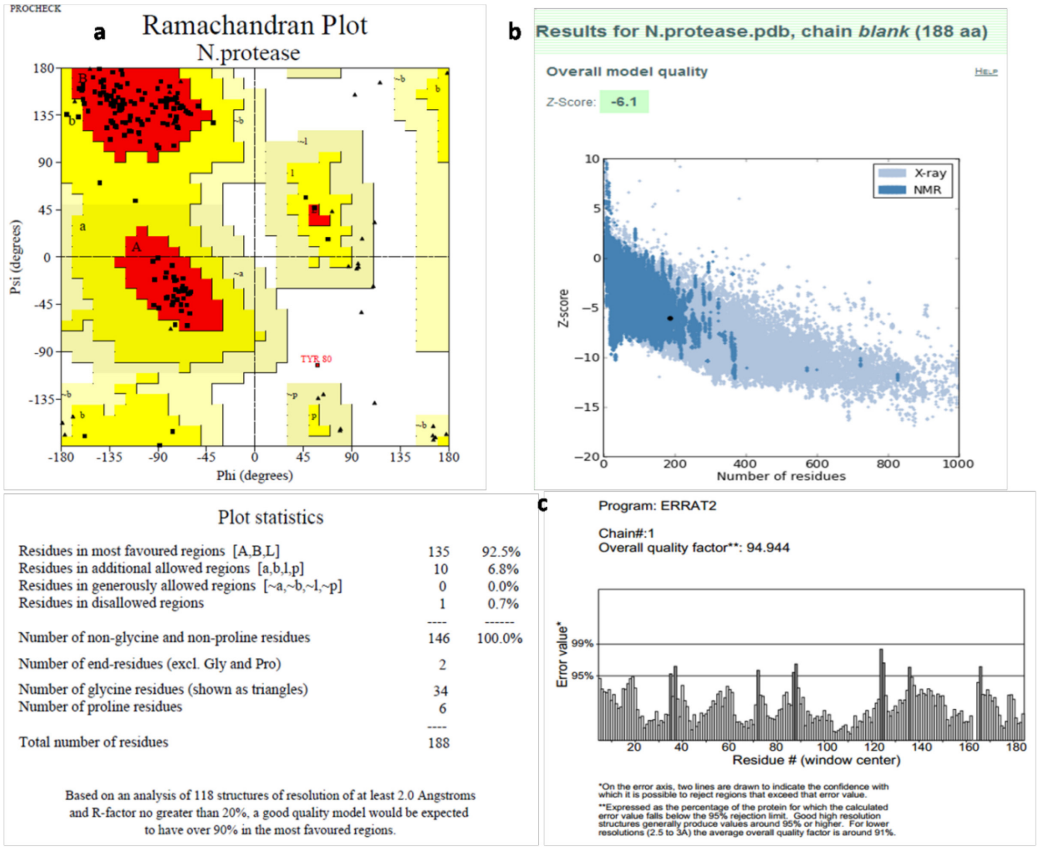


Figure 9: Evaluation parameters for the model generated of N.protease **a)** Ramachandran plot and its statistics **b)** ProSAII analysis **c)** ERRAT score.

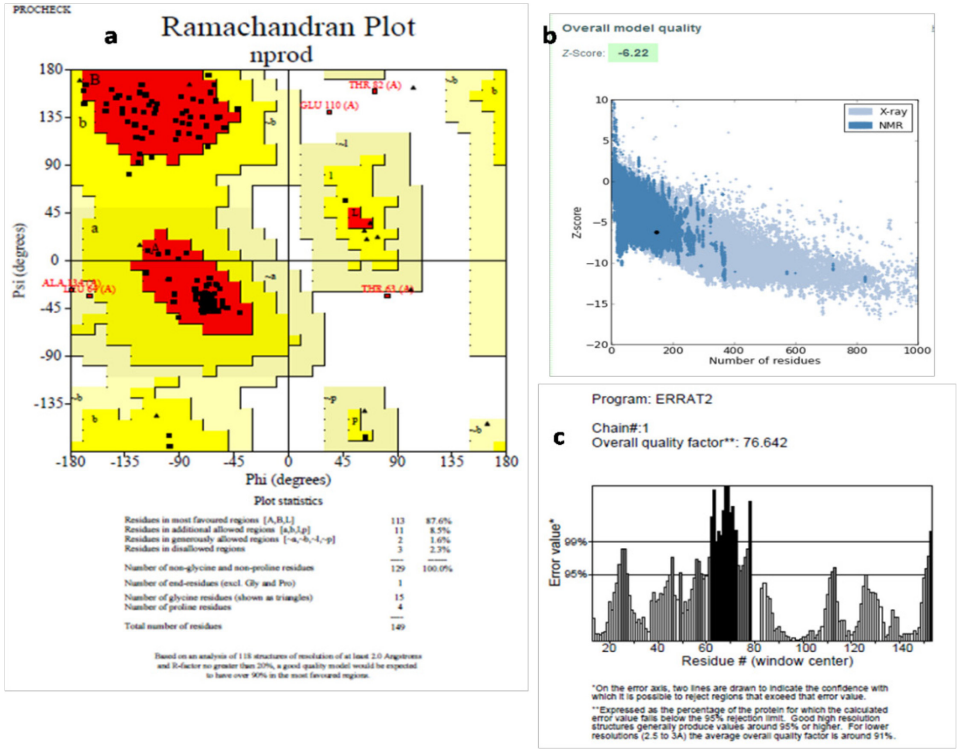


Figure 10**: **Evaluation parameters for the model generated of pro-region of N.protease **a)** Ramachandran plot and its statistics **b)** ProSAII analysis **c)** ERRAT score.

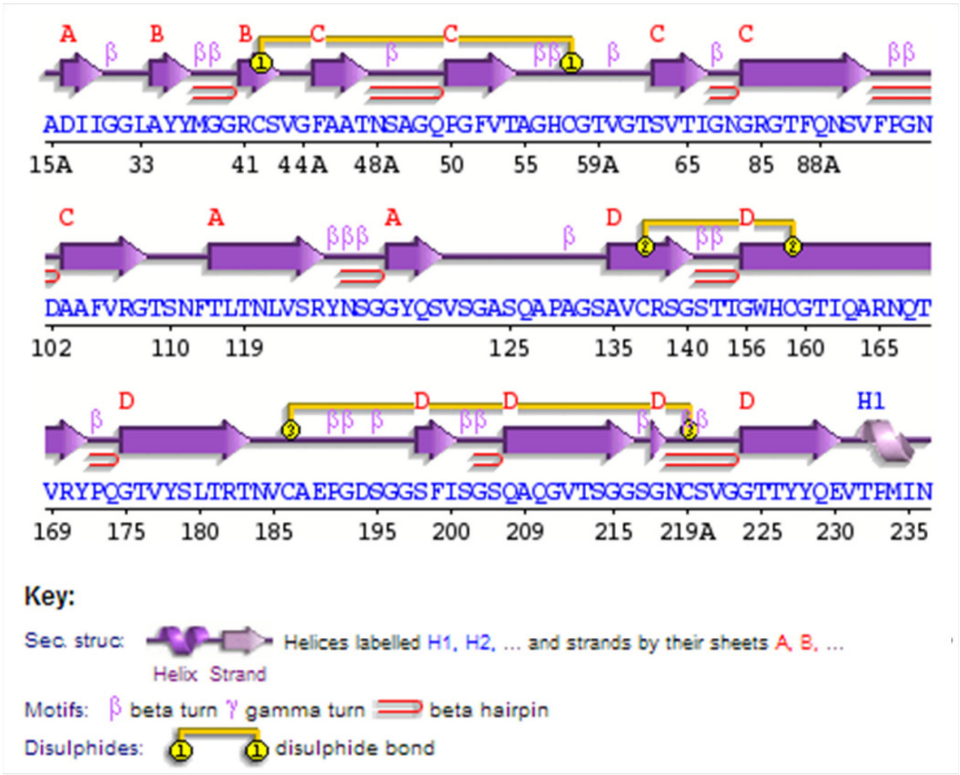


Figure 10**: **Diagrammatic representation of the secondary structure of N.protease as seen in PDBSum. Three disulphide bonds present in N.protease are seen in this

representation.

3.4.2 Phylogenetic analysis

In the phylogenetic tree (Fig. 11) N.protease was seen clustered with NAPase from *N.alba* with significantly high bootstrap value. αLP was seen in another cluster in same clade. Both of these proteases are well studied as kinetically stable proteases. The thermodynamically stable protein, chymotrypsin was clustered in totally different clade.

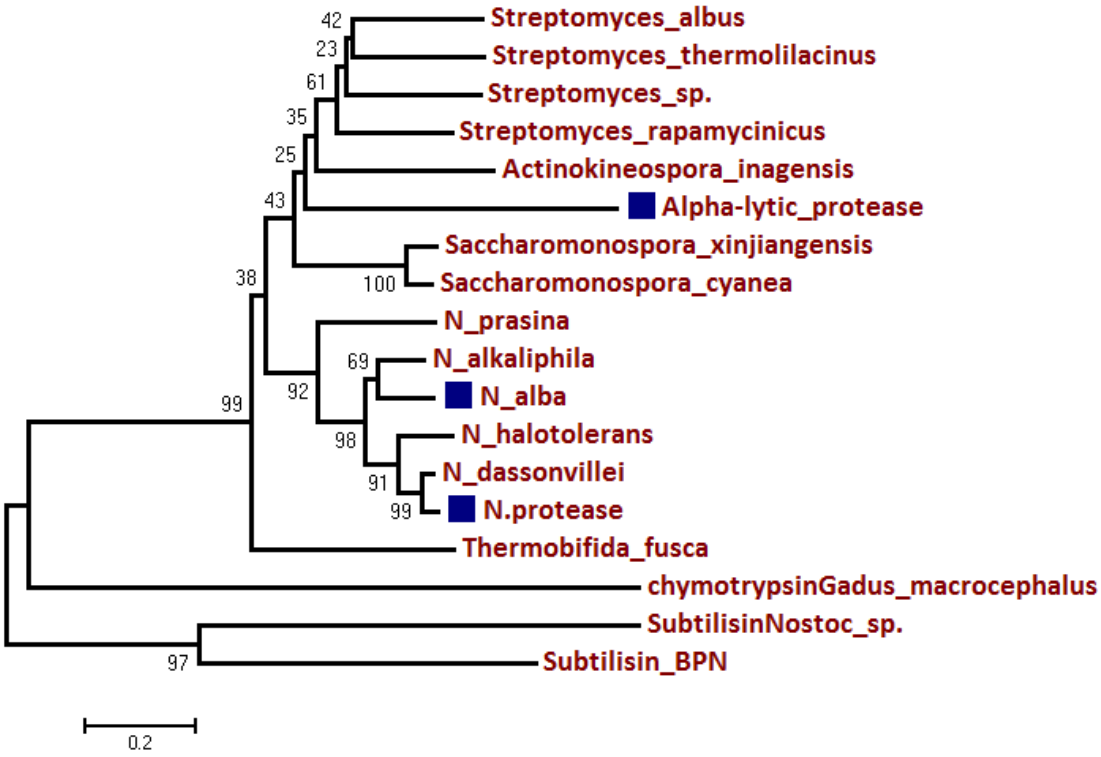


Figure 11: Phylogeny tree generated using MEGA 6.0 program. The kinetically stable proteases have been highlighted with squared boxes.

3.4.3 Thermal Simulations

High-temperature molecular dynamics (MD) unfolding simulations can be pursued for studying protein unfolding. As, the unfolding rates for proteins can be very slow under physiological conditions (for instance it can be years for proteins like αLP), the unfolding can be accelerated into the ns range required for computational analysis using very high temperatures (450 – 500 K) for simulations [6]. The thermal simulations for

N.protease were performed at 298 K and 500 K.

The N.protease was found to be stable during 10 ns of MD simulations at 298 K. The Cα RMSD for N.protease was plotted for each of the 3 simulations performed at 500 K (Fig. 12). The protein was found to be unfolded in each simulation based on visual inspection of trajectories (Fig. 13) and the high Cα RMSD attained. It was speculated that the major unfolding occurs in first 2 ns as high Cα RMSDs were attained in this time period. The unfolding events were studied with focus mainly on the N- terminal β strand, domain bridge and C-terminal β hairpin.

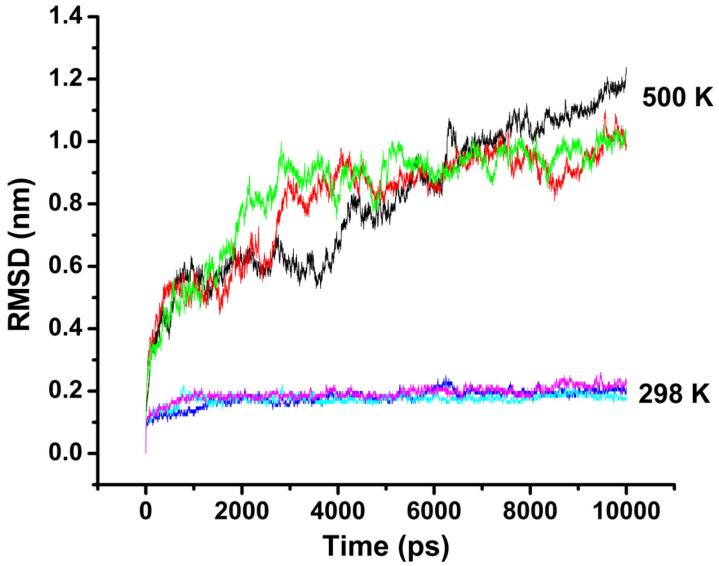


***Figure 12**: **Unfolding of N.protease monitored with RMSD values at 500 K and 298 K**.* Quick unfolding of N.protease could be seen below 2 ns at 500 K.

The domain bridge, which is the only covalent linkage in the two domains, was found to be intact even after 4 ns of simulation. The area covered by the domain bridge has been found to be an important factor in deciding the cooperativity of unfolding of kinetically stable proteins [5].

The C-terminal β hairpin becomes more mobile from 0.7 ns but remains intact till 1.8 ns, while the N- terminal β strand seems to be more flexible and pulled away from the protein body. The C-terminal β-barrel seems to unfold earlier than the N-terminal β barrel at 4 ns. At 7 ns very little residual structure remained.

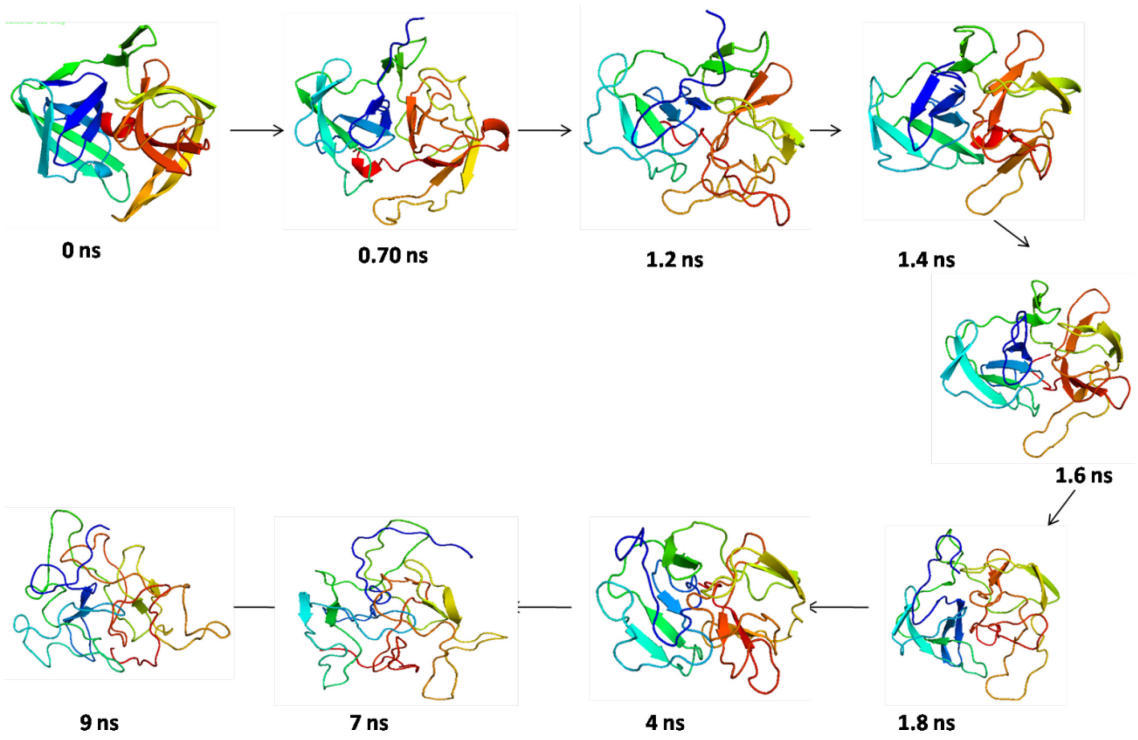


Figure 13: Unfolding simulations. The unfolding pathway of N.protease at 500 K illustrated by selected structures from the high temperature MD simulations.

3.4.4 P value analysis

To examine whether the high temperature unfolding behavior of N.protease is cooperative, we have performed P value analysis (Fig. 14). The high temperature unfolding cooperativity has been very well studied for αLP. Here, we have quantified the structural persistence during thermal simulation with respect to the reference structure i.e. the structure at 0 ns.

The P value analysis was also performed at 298 K for comparison. At 500 K, there was decrease in P value for both the proteins. But, this decrease was sharper for αLP compared to N.protease, indicating that N.protease followed comparatively gradual temperature unfolding. This gradual unfolding could also be seen in structural details (Fig. 13) where the domain bridge of αLP was seen to be unfolded by 2.4 ns while for N.protease it gradually unfolded after 4 ns. This comparative gradual unfolding of N.protease suggested its lesser kinetic stability compared to αLP.

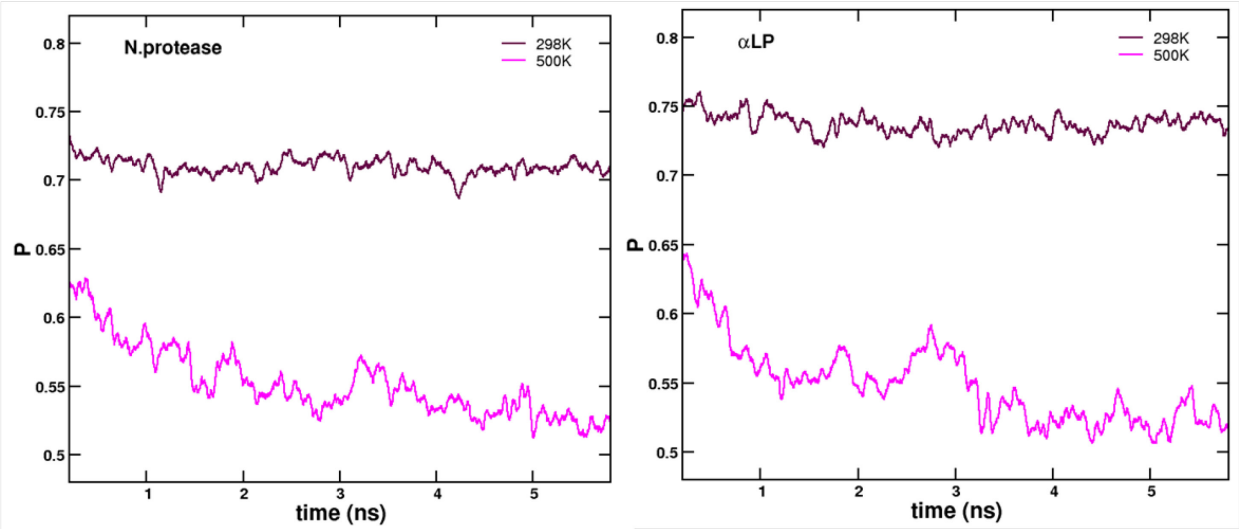


Figure 14: P value analysis. Plots of P value for N.protease and alpha lytic protease, at 300 K and 500 K, plotted as a function of simulation time.

We wished to study the sequence of NprotI and hence the cloning and expression studies were performed. A serine protease gene from *Nocardiopsis sp.* was successfully cloned and expressed in *E.coli*. Unfortunately, the cloned protein was different from NprotI as seen from CD spectrum and presence of disulfide bonds. However, the expression of the protease was studied in detail with respect to expression vectors, inducer concentration and *E.coli* growth temperature. The expressed protein was found to be active inside the *E.coli* cells (based on its autoprocessing to cut off pro-domain) but it was unstable once the cells were lysed. This could be the reason why caseinolytic activity was not detected recombinant protein.

In summary, highly GC rich gene sequence of a serine protease from *Nocardiopsis sp.* was codon optimized, cloned and expressed in *E.coli.* A comparison between pet-22b (+) and pet-39b (+) vectors for better expression and proper folding of N.protease suggested pet-39b (+) to be a better vector for disulfide bonds containing proteins. N.protease was found to be the member of S1A family of PA clan, whose members have limited distribution in prokaryotes. The phylogenetic relatedness of N.protease to kinetically stable proteins; NAPase and αLP, supported the previous finding that the less studied non-streptomycetes actinomycetes sp. can be potential sources of enzymes with exceptional stabilities. The *in silico* studies performed to test the kinetic stability of

N.protease indicated lesser stability compared to αLP, a model kinetically stable protein. Experimental evidences for these results are currently being pursued.

4. References

1. J M Sanchez-Ruiz. *Biophys Chem*. **2010**, 148:1–15.
2. E L Cunningham, D A Agard. *Biochemistry*. **2003**, 42(45):13212–13219.
3. N P Bryan. *Chem Rev*. **2002**, 102(12):4805–4816.
4. S S Jaswal, J L Sohl, J H Davis, D A Agard. *Nature*. **2002**, 415:343-346.
5. B A Kelch, K P Eagen, F P Erciyas, E L Humphris, A R Thomason, S Mitsuiki, D A Agard. *J Mol Biol*. **2007**, 368:870–883.
6. N L Salimi, B Ho, D A Agard. *PLoS Comput Biol*. **2010**, 6: e1000689.
7. W Peczynska-Czoch, M Mordarski. *Actino in Biotech*.**1988**, 220-283.
8. V J Mehta, J T Thumar, S P Singh. *Biores Tech.* **2006**, 97:1650–1654.
9. S P Singh, J T Thumar, S D Gohel, M K Purohit. *In Curr Res Technol Education Topics in Appl Microbial Biotechnol* Edited by Mendez A. **2010**.
10. S D Gohel, S P Singh. *J Chrom B.* **2012**, 889–890:61–68.
11. H N Wael, L Wen-Jun, A I A Mohammed, O Hammouda, M S Ahmed, X Li-Hua, X Cheng-Lin. *Int J Syst Evol Microbiol*. **2004**, 54:247–252.
12. S B Rohamare, V Dixit, P K Nareddy, D Sivaramakrishna, M J Swamy, S M Gaikwad. *Biochim Biophys Acta*. **2013**, 1834:708-716.
13. M Sone, Y Akiyama, K Ito. *J Biol Chem*. **1997**, 272:10349–10352.
14. V S Dixit, A Pant. *Biochim Biophys Acta.* **2000**,1523:261-268.
15. M S Shivji, S O Rogers, M J Stanhope. *Mar Ecol Progr Ser.* **1992**, 84:197–203.
16. H Sun, A Lapidus et al. *Stand Genomic Sci.* **2010**, 3(3):325–336.
17. X Huang, A Madan. *Genome Res.***1999,** 9:868-877.
18. S F Altschul, W Gish, W Miller, E W Myers, D J Lipman. *J Mol Biol.* **1990**, 215: 403-410.
19. N Eswar, M A Marti-Renom, B Webb, M S Madhusudhan, D Eramian, M Shen, U Pieper, A Sali. *Curr Prot in Bioinfo*. **2006**, 15:5.6.1-5.6.30.
20. C Colovos, T O Yeates. *Protein Sci.***1993**, 2:1511–1519.

21. T A de Beer, K Berka, J M Thornton, R A Laskowski. *Nucleic Acids Res.* **2014,** 42: D292-D296.

22. M Wiederstein, M J Sippl. *Nucleic Acids Res.* **2007**, 35: W407–W410.

23. K Tamura, G Stecher, D Peterson, A Filipski, S Kumar. *Mol Biol & Evol*. **2013**, 30: 2725-2729.

24. D van der Spoel, E Lindahl, B Hess, G Groenhof, A E Mark, H J Berendsen. *J Comp Chem.* **2005**, 26: 1701–1719.

25. W Humphrey, A Dalke, K Schulten. *J Molec Graphics*. **1996**, 14: 33-38.

26. P Chatterjee, N Sengupta. *Eur Biophys J.* **2002**, 41(5):483-489.

27. J C Jose, N Sengupta. *Eur Biophys J*. **2013**, 42:487-494.

28. M A Larkin, G Blackshields, N P Brown, R Chenna, P A McGettigan, H McWilliam, F Valentin, I M Wallace, A Wilm, R Lopez, J D Thompson, T J Gibson, D G Higgins. *Bioinformatics*. **2007**, 23:2947-2948.

29. D Baker. *Nature Struc Biol.* **1998**, 5 (12):1021-1024.

30. N D Rawlings, A J Barrett, A Bateman. *Nucleic Acids Res*. **2010**, 38:D227-233.

31. M J Page, E Di Cera. *Cell Mol Life Sci*. **2008**, 65:1220 – 1236.

Chapter 6

Discussion

1. Discussion

An extracellular alkaline serine protease from *Nocardiopsis sp.* NCIM 5124 was investigated for its biochemical and biophysical attributes. Previous studies by Dixit et al, had shown the enzyme to be of 21 kDa but with anomalous shape as it could pass through 3 kDa cut off filter membrane [1]. Also, it was already reported that the amino acid composition of the protein was unusual with very high content of proline residues (around 10 %).

It is known that, many of the extracellular proteases of microbial origin are highly stable to various denaturing conditions, as they are evolved to survive in the environment, where they need to function [2-5]. The enzymes from non-streptomycete actinomycetes have not been studied very well for their structure and function with few exceptions like NAPase from *N. alba* [6]. NAPase has been studied for its high acid stability.

With this background, we undertook the biochemical and biophysical characterization of NprotI, from *Nocardiopsis sp.* NCIM 5124.

The enzyme was characterized for its tryptophan environment, active site residues and inhibition by various protease inhibitors. The single tryptophan residue in NprotI was found to be surrounded by merely positively charged residues at native pH (pH 5.0) as well as at pH 1.0 and 10.0, with more positive charge at pH 1.0. Charge reorientation was seen in denatured protein as indicated by quenching with Cs^{+}. An Arg residue might be responsible for holding active catalytic conformation in NprotI.

The enzyme was found to possess PPII fold as its global conformation, which is very unusual as the PPII is found as local fold in many proteins. We speculated that, presence of this fold made the enzyme resistant to GdnHCl denaturation, proteolysis, and various solvents. Hence, this fold was found to be responsible for kinetic stability of NprotI. The structure-function relationship of NprotI is influenced by PPII fold. The PPII fold of the enzyme is more stable towards chemical denaturants and proteolytic enzymes than to the physical denaturant i.e. heat. This might be because PPII helix does not have internal hydrogen bonds for stabilization. The chemical reagents might be fulfilling the need of PPII helix for hydrogen bonding while heat

could be disrupting the other stabilizing non-covalent interactions. Melting of PPII helix could be seen, when the temperature unfolding monitored with far UV CD spectrum.

NprotI was also studied for its atypical stability and activity profile at acidic pH and higher temperature. Earlier, the PPII fold was shown to be responsible for the kinetic stability of NprotI. After that, the structural and functional stability of NprotI over a wide pH range at different temperatures for extended time periods was studied. Although, the PPII fold is maintained throughout the broad pH range studied, the enzyme was not functionally stable when incubated alone in alkaline environment. Also, the CD and DSC profiles indicated lesser thermal stability of the enzyme at alkaline pH. The functional rigidity is supported by structural rigidity as indicated by thermostability of the acid induced state.

The unusual stability of NprotI towards high concentration of denaturing agent, organic solvents, proteolytic enzymes and in wide pH range makes this enzyme an interesting candidate for structural investigations. The stability was attributed to the presence of PPII conformation observed for the first time as a global conformation in a non structural protein of microbial origin. The exact implication of presence of PPII helix in NprotI, apart from providing kinetic stability, needs to be studied further.

Further, to study the sequence of NprotI, cloning and expression studies were performed. A serine protease gene from *Nocardiopsis sp.* was successfully cloned and expressed in *E.coli*. Unfortunately, the cloned protein was different from NprotI based on CD spectrum and presence of disulfide bonds. But, the expression of the protease studied in detail with respect to expression vectors, inducer concentration and *E.coli* growth temperature. Two expression vectors, pET-22b (+) and pET-39b (+), were compared for proper folding and soluble expression of the protease (N.protease), which contained 3 disulfide bonds. In that pET-39b (+), with DsbA tag which helps in disulfide bond formation, was found to be a better vector. Homology models were built for N.protease and it's prodomain. The *in silico* studies performed to test for kinetic stability of N.protease indicated lesser stability compared to αLP, a model kinetically stable protein.

2. Conclusions

The highlights of the thesis are:

1. The single Trp residue present at the surface of NprotI is in positively charged environment.
2. The far UV CD profile and denaturation profile in presence of GdnHCl confirmed **presence of a unique PPII fold** in NprotI.
3. NprotI was found to be resistant towards proteolytic enzymes as well as organic solvents.
4. More pronounced secondary structure (far UV CD spectrum) was observed in presence of different denaturants indicating **PPII fold to be responsible for the kinetic stability** of the enzyme
5. PPII fold in NprotI was more resistant to chemical denaturation than thermal denaturation, as monitored in CD and DSC studies, while at pH 1.0, reversal of this behavior was observed.
6. GuSCN was more effective in denaturing NprotI. Enhancement in enzyme activity was observed at lower concentration but complete inhibition and unfolding was seen at higher concentration.
7. Enzyme seemed to be structurally stable in the wide pH range. Although, an alkaline serine protease, NprotI was more stable at highly acidic pH (pH 1.0-3.0), at 25 °C for 24 h. The activity of NprotI was found to be enhanced at pH 1.0-3.0, 50-60 °C.
8. At pH 1.0, 5.0 and 10.0, different modes of thermal denaturation were observed.
9. Uncompetitive inhibition of NprotI was observed with CanPI 7, an inhibitor of plant origin, indicating that the inhibitor binds to the site other than active site.
10. The enzyme was found to follow reverse Hofmeister series when activity and CD profile were studied in presence of different metal ions of the series.
11. A serine proteases from *Nocardiopsis sp.* NCIM 5124 were cloned and expressed in *E. coli*, structurally it resembled to PA clan superfamily.
12. pET-39b was found to be a good vector for the proper folding and soluble expression of proteins with disulphide bonds.

13. The cloned protein was checked for probable kinetic stability by performing high temperature MD simulations and P value analysis.

3. References

1. V S Dixit, A Pant. *Biochim Biophys Acta.* **2000,** 1523**:**261-268.
2. M Manning, W Colon. *Biochemistry*. **2004**, 43:11248-11254.
3. J M Sanchez-Ruiz. *Biophys Chem.* **2010**, 148: 1–15.
4. J L Sohl, S S Jaswal, and D A Agard. *Nature.* **1998,** 395: 817–819.
5. E L Cunningham, D A Agard. *Protein Sci.* **2004,** 13: 325–331.
6. B A Kelch, K P Eagen, F P Erciyas, E L Humphris, A R Thomason, S Mitsuiki, D A Agard. *J Mol Biol.* **2007,** 368: 870–883.