STUDIES ON PROTEASE INHIBITOR, ITS COMPLEX WITH TRYPSIN AND CHILECTIN FROM *Tamarindus indica*

A THESIS

Submitted in partial fulfilment of the requirements for the award of the degree of

DOCTOR OF PHILOSOPHY

in

BIOTECHNOLOGY

by

PATIL DIPAK NARHARI

DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247 667 (INDIA)

MAY, 2012

INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **"STUDIES ON PROTEASE INHIBITOR, ITS COMPLEX WITH TRYPSIN AND CHILECTIN FROM *Tamarindus indica*"** in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy and submitted in the Department of Biotechnology of the Indian Institute of Technology Roorkee, Roorkee is an authentic record of my own work carried out during the period from July 2007 to May 2012 under the supervision of Dr. Pravindra Kumar, Assistant Professor, Department of Biotechnology, Indian Institute of Technology Roorkee, Roorkee.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Dpatil

(PATIL DIPAK NARHARI)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Pkumar

(Pravindra Kumar)
Supervisor

Date: May 4, 2012

The PhD Viva-Voce Examination of **Patil Dipak Narhari,** Research Scholar has been held on...25/06/12.....

Pkumar
Supervisor

Chairman SRC 25.06.2012

External Examiner

Head of the Department

ABSTRACT

Seeds of leguminous plants are known to contain proteins that are important for a number of biological processes. Proteinase inhibitors, chitinases, lectins, globulins etc. from these seeds are the most studied ones, since Leguminosae family members are recognized as an excellent source of these biologically significant proteins. The Kunitz type inhibitors of serine proteases form an important field of scientific investigation due to their multifunctional abilities. The Kunitz STI inhibitors are included as Kunitz-P inhibitors in the I3A family of the MEROPS database (http://merops.sanger.ac.uk). Specific inhibitors are reported as insecticidal, anti-inflammatory, anti-cancer, anticoagulant, antiviral, antiparasitic etc. These multifunctional plant Kunitz type inhibitors are crucial tools enabling us to gain knowledge of the basic principles of protein interactions and could be important for human beings in control of many diseases by using them to discover highly potential drugs. So, it is necessary to search for novel Kunitz inhibitors with important functions. Applications of these types of multifunctional inhibitors are very extensive which give an inspiration to perform their structure based study to correlate their structure with their activities. The structural properties of inhibitors that control the activity of the proteinases could be targeted as potential pharmaceutical products.

Plant lectins are heterogeneous and highly diverse class of non-immune origin (glyco) proteins because of their carbohydrate-binding specificity, differences in molecular structure and biochemical properties. They are classified into twelve diverse families of evolutionary and structurally related lectin domains. There is no precise description of the biological functions of plant lectins because of their diverse classes and carbohydrate specificities. Several functions of specific lectins are reported which include antifungal, insecticidal, antiviral, antiproliferative, apoptosis-inducing and symbiosis mediating between nitrogen fixing microorganisms and legume plants. Lectins show extensive structural diversity and considering the structural folds, plant lectins are grouped into seven folds. Previous reports showed that plant lectins are also members of chitinase family in fold while they might possess or lack chitinase activity. To gain further insight into structure-function relationship of lectins in general, it is necessary to carry out structural study of these types of proteins which will also lead to understand their evolution and how they acquire new functions using the same parental fold.

My thesis encompasses the cDNA cloning, isolation, purification, characterization, crystallization and structural studies of a Kunitz type inhibitor, its complex with trypsin and

purification, biochemical and structural studies of chitinase-like lectin (chilectin) from Tamarindus indica.

The first chapter gives a general introduction and a brief review of the work done nationally and internationally for Kunitz type inhibitors and lectins.

The second chapter describes purification and biochemical characterization of a Kuntiz type inhibitor from seeds of a leguminous plant, Tamarindus indica named as TKI (Tamarind Kunitz type inhibitor). TKI was isolated using a series of columns, initially Affi-blue column followed by DEAE an anion exchange column. The approximate molecular mass of the TKI was determined to be 21 kDa from SDS-PAGE analysis. The protein yield was 0.6 mg per gram of seeds and consisted of a homogenous mixture. The trypsin inhibitory activity and N-terminal amino acid sequence analysis revealed the isolated protein to be a protease inhibitor belonging to Kunitz type inhibitor family. The N-terminal amino acid sequence of TKI showed 79% homology with factor Xa inhibitor (BuXI) from Bauhinia ungulata and 64% with Trypsin inhibitor (BvTI) from Bauhinia variegata. TKI was tested for its anticoagulant, factor Xa inhibitory and antiproliferative activities and was found to possess anticoagulant and FXa inhibitory activities. Furthermore it also showed dose dependent suppression of the cancer cell growth of MCF-7, HepG2 and PC-3 cells.

The third chapter contains cDNA synthesis, cloning, expression and purification of TKI. Briefly, the total RNA was isolated using three month old seeds after flowering and cDNA was synthesized. It was further amplified and cloned into pGEM-T easy vector and sequenced. The cDNA fragment amplified with PCR was ~700 bp in length consisting of 3' UTR region and poly (A+) tail. A 555 bp ORF obtained after sequencing of the TKI gene coded for a polypeptide of 185 amino acid residues with a calculated molecular mass of 20575 Da. Sequence analysis showed that TKI belongs to Kunitz type STI family, however, has distorted Kunitz signature signature sequence due to insertion of Asn15 in the signature motif. TKI has maintained two characteristics of Kunitz STI family such as having molecular weight ~20-22 kDa and conserved two disulphide bonds.

The full length TKI encoding gene was cloned into pET-28cTEV vector and expressed into BL21- DE3 competent Escherichia coli cells which showed expression and solubility of TKI but with negligible trypsin inhibitory activity which might be due to improper folding and disulfide bonds formed. Formation of correct disulfide bonds and proper folding might be

required for TKI to maintain its inhibitory activity, so we expressed TKI in chemically competent E. coli K12 cells shuffle T7 which is engineered to form disulfide bonded proteins in the cytoplasm. SDS-PAGE analysis showed that rTKI was expressed as a soluble protein which has potent trypsin inhibitory activity like native TKI. Recombinant TKI was purified using Ni–NTA column and pure rTKI exhibited trypsin inhibitory activity. This result shows that rTKI is a stable protein and able to preserve its functional property and can be used further for biochemical studies.

The fourth chapter encloses the structure of TKI, its complex with trypsin and molecular docking studies with factor Xa. Crystal structure of free TKI was solved at 1.9Å showing β-trefoil fold consisting of 12 anti-parallel β-strands with two disulphide bridges and an exposed reactive site loop having Arg at P1. To understand the mode of interaction of TKI with trypsin, crystal structure of TKI:pancreatic porcine trypsin (PPT) was determined at 2.2Å. The stable complex formed by TKI blocked the S1 pocket of PPT by its exposed reactive site and stabilized the complex by making interaction with the subsites of PPT. The electron density for Asn15 was clearly visible in both free as well as TKI:PPT complex structure.

In order to analyze the possible binding of TKI with FXa, molecular docking was performed. TKI P1 and P3 residues showed the interactions with S1 and S4 pocket of FXa, respectively in a classical L-shaped substrate-like conformation. Arg64, which interacts with cation hole with its side chain occupying the aryl binding site of S4, is the most important residue of TKI which may provide specificity towards FXa. The Asn15 is important residue insertion in TKI which is involved in recognition and stabilization of complexes. Asn15 interacts with four residues of PPT and two residues of FXa and participates in stabilization of complex. Furthermore, it interacts with those residues which are important for specificity of FXa which point towards its role in recognition of specific proteases. Comparison of the reactive site of TKI with Kunitz STI inhibitors and factor Xa inhibitors show that it has distinct reactive site residues which possibly inhibit other trypsin-like serine proteases and could play multifunctional roles.

The fifth chapter encompasses purification, characterization, crystallization and structure determination of the chitinase-like lectin (TCLL) and its complex with GlcNAc. The TCLL was purified using Affi-blue and DEAE column chromatography was found to be monomeric 34 kDa protein which shows negligible chitinase activity. The structure was determined at 1.49Å and shows (βα)8 barrel topology. The sequence of TCLL was proposed from crystallographic data

and by mass spectroscopy. TCLL showed significant sequence similarity to the reported class III chitinase of GH18 family members. Evaluation of active site and substrate binding subsites revealed that it has mutations of key residues which are required for chitinase activity.

Remarkably, the complex structure of TCLL with GlcNAc reveals that the latter binds in a pocket formed by two loops β4α4 and β5α5 which have different conformation in structures of GH18 members. From biochemical and structural studies, we concluded that TCLL is a chilectin which explored novel carbohydrate binding site other than chitin-binding groove in GH18 family members. However, TCLL does not fit anywhere into this classification of plant lectins by Van Damme et al., (2008). Consequently, TCLL structure signifies a new class of plant lectins and confers evidence for evolutionary link of lectins with chitinases. The structure of TCLL also depicts how plants utilize the existing structural scaffolds ingenuously to attain new functions. So we propose a new class of lectin which is evolutionarily related to class III chitinase. Moreover, the one of the existing 'Class V chitinase homologs' class of lectin and our proposed class 'Class III chitinase homologs' have same fold. To avoid number of classes in lectin classification, we propose the new class "TIM barrel domain" rather than existing one and these two classes can be grouped into TIM barrel domain as one subclass. Furthermore, one cannot exclude the possibility that the glycosyl hydrolase family members or other families having TIM barrel topology from plants will not emerge as lectin. So the TIM barrel domain will be the main class and the forthcoming classes will be subclasses.

ACKNOWLEDGEMENT

It gives me great pleasure in expressing my sincere gratitude and immense veneration towards my supervisor, Dr. Pravindra Kumar for his meticulous guidance, prudent support and keen interest throughout the course. It was a great joy and an invaluable experience to work with him. Under his esteem guidance, I have strengthened my skills in research and also developed a never ending quest for knowledge. I really thank him to provide me the opportunity to work under his guidance and am highly obliged to him for his expert guidance, perceptual encouragement, caring attitude and continuous support.

I endow special thanks to Dr. Ashwani Kumar Sharma and Dr. Shailly Tomar for helping in molecular and biochemical techniques. Without their encouragement and constant guidance, it was not possible for me to complete my work successfully. I would like to express sincere thanks to Dr. Partha Roy for helping me in performing anticancer activity experiments in his laboratory. I am thankful to Dr. Sumar Kundu, Department of Biochemistry, University of Delhi south campus, New Delhi for internal sequencing of the protein.

I gratefully acknowledge the help rendered from time to time by the members of my Student Research Committee (SRC): Prof. G. S. Randhawa (present DRC chairman), Dr. Ashwani Kumar Sharma and Dr. Ramesh Chandra for their scholarly suggestions; prudent admonitions and immense interest that have made this task successful. I am also grateful to the faculty members Prof. Ritu Bartwal, Prof. R. P. Singh, Prof. R. Prasad, Prof. H. S. Dhaliwal, Dr. Vikas Pruthi, Dr. Bijan Choudhury, Dr. Sanjay Ghosh, Dr. Naveen Kumar Navani, Dr. Ranjana Pathania and Dr. Maya Nair for their support and encouragement.

I am obliged to my labmates Shivendra, Aditya, Pramod, Sonali, Preeti, Tapas, Megha and Manju for their amiable attitude and coopearation while working in lab. I am thankful to Girijesh, P. Selva Kumar, Bibekanand and Prabhat Tomar for their helpful discussion and support. I am thankful to Nikheel for making effort to perform anticancer activity. I am grateful to all those who have helped me directly or indirectly in the successful completion of this thesis.

This work would not have been possible without the help and support of all the members of the office staff Mr. Lokesh Kumar, Mr. Ved Pal Singh Saini, Mrs. Surita Sharma, Mrs. Shashi Prabha, and Mr. Sunil Kumar Sharma.

I owe my loving thanks to my wife, Mrs. Supriya Patil, for all the support she gave me during the years I have been working and while writing the thesis and manuscripts. It was not possible for me to finish my thesis without her help, unconditional support, encouragement and her unwavering confidence in me. I am really thankful to her for correcting my manuscripts and thesis critically and patiently

Last, but not the least I thank my parents, Narhari Patil and Pushpa Patil, for giving me life, educating me and encouraging me to pursue my interest. I thank my sisters Mrs. Sangita Patil, Mrs. Sunita Patil and Mrs. Prita Patil for their love and moral support. I thank my parents in law Mr. O P Deepak and Mrs. Raj Deepak and my brother in law Sumeet Deepak for their love and constant inspiration.

Finally, I am thankful to Ministry of Human Resource Development, Govt. of India for providing the financial support.

(PATIL DIPAK NARHARI)

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PUBLICATIONS

1. **Patil DN**, Datta M, Chaudhary A, Tomar S, Sharma AK, Kumar P. (2009). Isolation, purification, crystallization and preliminary crystallographic studies of chitinase from tamarind (*Tamarindus indica*) seeds. **Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.** 65(Pt 4): 343-345

2. **Patil DN**; Preeti, Chaudhry A, Sharma AK, Tomar S, Kumar P. (2009). Purification, crystallization and preliminary crystallographic studies of a Kunitz-type proteinase inhibitor from tamarind (*Tamarindus indica*) seeds. **Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.** 65(Pt 7): 736-738

3. Tomar S, **Patil DN**, Datta M, Tapas S, Preeti, Chaudhary A, Sharma AK, Tomar S, Kumar P. (2010). Crystallization and preliminary X-ray diffraction analysis of the complex of Kunitz-type tamarind trypsin inhibitor and porcine pancreatic trypsin. **Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.** 65(Pt 11): 1179-1181

4. **Patil DN**, Chaudhary A, Kumar N, Roy P, Sharma AK, Tomar S, Kumar P. Structural insights into a multifunctional tamarind Kunitz type inhibitor (TKI) showing factor Xa inhibitory and anticancer activity and its complex with trypsin. (communicated)

5. **Patil DN**, Singh N, Dasauni P, Kundu S, Sharma AK, Tomar S, Kumar P. Structural investigation of tamarind chitinase-like lectin (TCLL), a novel N-acetyl glucosamine binding lectin, signifies emergence of a new class of plant lectins and reveals evolutionary relationship with class III chitinases (communicating soon)

6. **Patil DN,** Kumar P. Structural studies of tamarind chitinase-like lectin (TCLL) complex with xanthine derivatives. (Manuscript under preparation)

7. **Patil DN,** Toma S, Kumar P. Cloning, expression, purification and biochemical characterization of tamarind Kunitz inhibitor (TKI). (Manuscript under preparation)

8. Sakshi, **Patil DN**, Tomar S, Sylvestre M, Kumar P. (2010). Expression, purification, crystallization and preliminary crystallographic studies of cis-biphenyl-2,3-dihydrodiol-2,3-dehydrogenase from *Pandoraea pnomenusa* B-356. **Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.** 66(Pt 11):1517-1520

9. Dhindwal S, **Patil DN**, Mohammadi M, Sylvestre M, Tomar S, Kumar P. (2011). Biochemical studies and ligand-bound structures of biphenyl dehydrogenase from *Pandoraea pnomenusa* strain B-356 reveal a basis for broad specificity of the enzyme. **J. Biol. Chem.** 286(42):37011-37022

10. Narayanan A, Paul LN, Tomar S, **Patil DN**, Kumar P, Yernool DA. (2012). Structure-function studies of DNA binding domain of response regulator KdpE reveals equal affinity interactions at DNA half-sites. **PLoS One.** 7(1):e30102

11. Kumar P, **Patil DN**, Chaudhary A, Tomar S, Yernool D, Singh N, Dasauni P, Kundu S, Kumar P. Purification and biophysical characterization of a novel 11S lectin-like globulin from *Wrightia tinctoria* (submitted to Plant Physiol. Biochem.)

Gene submitted to NCBI database

TKI gene

- *Tamarindus indica* trypsin inhibitor mRNA, partial cds with accession number HQ385502

Atomic Coordinates and Structure Factors submitted to EMBL-EBI's Protein Data Bank in Europe (PDBe) with accession numbers-

4AN6: Kuntiz type trypsin inhibitor with factor Xa inhibitory activity

4AN7: Kunitz type trypsin inhibitor complex with porcine trypsin

3ZQ7: Structural studies of DNA binding domain of response regulator KdpE

2Y93: Crystal structure of cis-biphenyl-2,3-dihydrodiol-2,3-dehydrogenase (BphB) from *Pandoraea pnomenusa* strain B-356

2Y99: Crystal structure of cis-biphenyl-2,3-dihydrodiol-2,3-dehydrogenase (BphB) from *Pandoraea pnomenusa* strain B-356 complex with co-enzyme NAD+

3ZV3: Crystal structure of cis-biphenyl-2,3-dihydrodiol-2,3-dehydrogenase (BphB) from *Pandoraea pnomenusa* strain B-356 in intermediate state of substrate binding loop

3ZV4: Crystal structure of cis-biphenyl-2,3-dihydrodiol-2,3-dehydrogenase (BphB) from *Pandoraea pnomenusa* strain B-356 in apo form at 1.8 angstrom

3ZV5: Crystal structure of cis-biphenyl-2,3-dihydrodiol-2,3-dehydrogenase (BphB) from *Pandoraea pnomenusa* strain B-356 complex with co-enzyme NAD+ and product 2,3-dihydroxybiphenyl

3ZV6: Crystal structure of cis-biphenyl-2,3-dihydrodiol-2,3-dehydrogenase (BphB) from *Pandoraea pnomenusa* strain B-356 complex with co-enzyme NAD+ and product analog 4,4'-dihydroxybiphenyl

And the structure coordinates for tamarind chitinase like lectin (TCLL) has to be submitted

CHAPTER 1

LITERATURE REVIEW

1.1. Introduction

Seeds of leguminous plant are important in human and animal nutrition because of their rich protein and carbohydrate content. Proteins like lectins and proteinase inhibitors are commonly found [183]. These proteins can interfere with physiological processes such as germination and seed maturation, or can be defense proteins against the attack of seed predators [172, 184]. Apart from these, they are also reported to have pharmaceutical importance and can be utilized for human welfare. Other proteins which are also present in seeds are chitinases, glucanases, thaumatin-like proteins, thionins, cyclophilin-like proteins, ribosome inactivating proteins, ribonucleases, deoxyribonucleases, peroxidases, allergens, albumins and globulins. These proteins, including protease inhibitors and lectins play important biochemical and diverse cellular functions due to their ability to bind other molecules specifically and tightly. This binding ability is mediated by the structure of the protein, chemical properties of the surrounding amino acids side chains and physiological conditions.

1.2. Protease inhibitors

Proteases perform an indefinite number of hydrolytic reactions both intra- and extracellularly [97, 138]. They are found in all living organisms from bacteria to mammals. Proteinases are hydrolytic enzymes which in vivo catalyse cleavage of peptide bonds in protein and peptide substrates. Proteolytic enzymes comprise very broad range of substrate specificities applying several distinctly different chemical mechanisms to carry out peptide bond hydrolysis. There is a continuously increasing recognition of the function of the proteinases in a broad range of physiological processes of vital importance [38, 137, 145]. Proteinases are important for extracellular metabolism playing a vital role in defence mechanisms that protect an organism from tissue damage and infection (proteolytic cascades of blood coagulation, fibinolysis and complement systems). Proteinases generally act as regulatory elements through proteolytic activation of prohormones and zymogens, release of physiologically active peptides and are also active in macromolecular assembly of viruses and fibrin. Proteinases are mostly classified according to the main catalytic amino acid residue in their active site: serine proteinases, with a serine and a histidine; (2) cysteine proteinases, with a cysteine; (3) aspartic proteinases, with an aspartate group and metalloproteinases, with a metallic ion (Zn2+, Ca2+, or Mn2+) [137] in their active site.

However, beside their physiological necessity, proteases are potentially harmful to their proteinaceous environment and their activity must be correctly controlled by the respective cell or organism. When not properly controlled, proteases can be responsible for serious diseases. The basic level of regulation is by controlled expression/secretion, by activation of proproteases [90], and by degradation of the mature enzymes. A second level of regulation is by inhibition of their proteolytic activity. Almost all known naturally occurring inhibitors directed toward endogenous cognate proteases are proteins; only some microorganisms secrete small non-proteinaceous compounds which block the host protease activity. A huge number of inhibitors have been described; they were isolated from various cells, tissues and organisms. Often they accumulate in high quantities in plant seeds, bird eggs and various body fluids. Inhibition of proteinases by proteins itself appears to be a paradox. In fact, nature developed many different structural adaptations in protein structures to overcome the potential risk of proteolysis and develop specificity of recognition. Proteinase inhibitors adopt many different structures, ranging in size from mini-proteins to large macromolecular structures, much larger than the target enzyme

Nevertheless, there are very common examples of inhibition of proteases by structurally unrelated proteins [20, 97, 145]. In fact, inhibitor structures, modes of inhibition and the nature of enzyme-inhibitor complexes are surprisingly different. Previously, inhibitors were believed to be specific for one of the four mechanistic classes of proteases (serine, cysteine, aspartic, or metalloproteases). While this is probably true in a prevailing number of cases, there are also known examples of proteins that are able to inhibit cysteine and aspartic protease, a serine and metalloprotease, a serine and aspartic protease, or a serine protease and amylase which utilize different, non-overlapping binding sites. From the structural point of view, blocking of the enzyme active site is attained by docking of exposed structural components, like loops or protein termini, either independently or in combination of two or more such components. Canonical inhibitors of serine proteinases typically bind to target enzymes through an exposed proteinase binding loop [145].

1.2.1. Protease inhibitors of serine proteases

The serine proteinases are a large family of enzymes involved in a wide variety of vital biological processes. The crucial physiological functions of these enzymes in metabolism, blood coagulation, fibrinolysis, complement pathways, viral maturation, apoptosis and cancer make

them important targets for designing potent and specific inhibitors. Macromolecular proteinase inhibitors bind to the target proteinase through a single loop that indicates the critical P1 residue. This residue fits into the binding pocket of the target proteinase in a substrate-like conformation to lock the enzyme in a complex formed between the proteinase and inhibitor. Serine proteinases and their protein inhibitors have been the most extensively studied group of protein-protein complexes. Protein inhibitors of serine enzymes do not use a single mechanism to inhibit the cognate proteinase. They are categorized into three groups on the basis of different inhibition mechanisms. The largest group is formed by canonical inhibitors which follow the standard mechanism [105].

1.2.2. Canonical inhibitors

Canonical inhibitors form the largest group of protein inhibitors that act according to the standard mechanism of inhibition. These inhibitors are usually distributed in all groups of organisms and constitute 27 to about 200 amino acid residues. The standard mechanism implies that inhibitors are peculiar protein substrates containing the reactive site P1-P1' peptide bond located in the most exposed region of the proteinase binding loop (P1, P2 and P1', P2' designate inhibitor residues amino- and carboxy-terminal to the scissile peptide bond; S1, S2 and S1', S2' denote the corresponding subsites on the proteinase (Fig. 1.1) [173]. The reactive site can be selectively hydrolyzed by the enzyme. The binding loop is similar, so called canonical, conformation in inhibitor structures representing different families [19, 145]. It is usually assumed and in most cases has been verified experimentally that the standard mechanism inhibitors show canonical conformation of the binding loop. In principle, very different design concepts can be imagined, such as blocking the enzyme's active site in a substrate-like manner, docking adjacent to the active/substrate binding site, or allosterically impairing the proteolytic activity/substrate interaction of the enzyme via binding to quite distantly located enzyme exosites (Fig. 1.2)

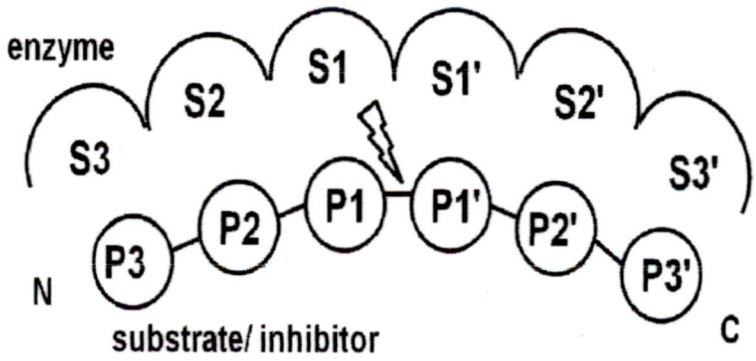


Figure 1.1. Schematic representation of the active site of a typical serine protease and reactive site of inhibitor. The substrate/inhibitor reactive site represented as P3-P3' and enzyme active site as S3- S3'. The scissile peptide bond is located between P1 and P1'.

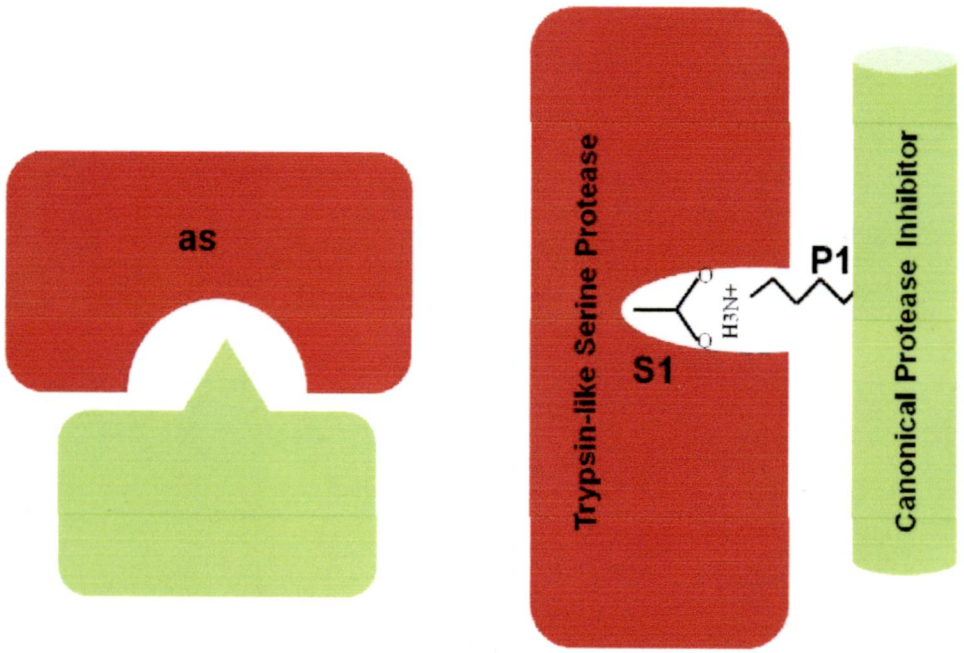


Figure 1.2. A. Schematic representations of substrate like inhibition of trypsin like serine proteases (red) by canonical inhibitors (green). B. Schematic representations of typical interactions between proteinases (red) and their protein inhibitors (green). P1 side chain (here that of a Lys residue) of reactive site loop forming hydrogen bonds with the S1 pocket residue (here Asp residue). Modified from Bode and Huber, 2000 [20].

1.2.2.1. The standard mechanism

The canonical inhibitor-cognate protease interaction is preserved in all cases tested and called the standard mechanism [105]. The interaction between enzyme and inhibitor can be presented in a simplified form as a hydrolysis/resynthesis reaction of the P1-P1' reactive-site peptide bond:

$$E + I \underset{k_{off}}{\overset{k_{on}}{\longleftrightarrow}} EI \underset{k^*_{on}}{\overset{k^*_{off}}{\longleftrightarrow}} E + I^*$$

Where E is the protease, I is the inhibitor, I* is the reactive- site-cleaved inhibitor, EI is the stable complex, k_{on} and k^*_{on} are respective second-order association rate constants, and k_{off} and k^*_{off} are respective first-order dissociation rate constants of the complex.

1.2.2.2. Classification of Canonical inhibitors

A classification of canonical inhibitors was originally proposed by Laskowski and Kato in 1980 [105]. They distinguish it into eight families on the basis of the disulfide bond topography, location of the reactive site, and sequence homology. Now, 18 inhibitor families are recognized [97] (Table 1.1). Crystal and/or solution structures are known for representatives of almost all families (Table 1.1). Since the inhibitors are small, rigid, and stable, these structures have often been determined with high resolution and accuracy. Extensive structural information is also available for protease-inhibitor complexes. Most often they comprise either purely β sheet or mixed α/β proteins; they can also be α-helical or irregular proteins rich in disulfide cross-links.

Family	Representative members	Enzyme-inhibitor complex structure
BPTI	BPTI	BPTI: rat trypsin
Kazal	OMSVP3	OMTKY: chymotrypsin
Potato 1	CI-2	eglin c: subtilisin
Squash	CMTI I	CPTI II: trypsin
Ecotin 1	ecotin	ecotin: crab collagenase
STI	STI	STI: porcine trypsin
BBI	BBBI	MbBBI: Ns3-protease
BBI (SFTI)	SFTI-1	SFTI-1: trypsin
Antistasin	hirustasin	bdellastasin: porcine trypsin
Ascaris	AMCI	C/E-1 inhibitor: porcine elastase
Grasshopper	PMP-C	PMP-C: chymotrypsin
SSI	SSI	SSI: subtilisin
Potato 2	T1	PCI 1: SGPB
Cereal	CHFI	-
Chelonianin	R-elafin	Elafin: porcine elastase
Rapeseed	ATTp	-
Arrowhead	API-A	API-A:trypsin

Table 1.1. Classification of canonical inhibitors and their representative members whose X-ray/NMR three-dimensional structures have been solved along with the enzyme complexes. Modified from Krowarsch *et al*, 2003 [97].

1.2.3. Non-canonical inhibitors

Non-canonical inhibitors bind to the enzyme active site through their N-terminal segment but also contact the proteinase at more distant site(s). Due to the extensive contact area, these inhibitors form very tight and specific complexes with serine proteinases. Interestingly, such an interaction features also proteins possessing folds of canonical inhibitors like BPTI or Kazal-type inhibitors but with distorted conformation of the binding loop. There are also extensive secondary interactions which provide an additional buried area and contribute significantly to the strength and specificity of interaction. In the studied cases there is two-step kinetics of association — the initial slow binding step occurs at the secondary binding site, then the N-terminus locks in the active site of proteinase. The first recognized inhibitor of this class was hirudin — a 66 amino-acid residues protein from the saliva of medical leech. Both for hirudin in solution and for its complexes the structure is known [145, 164].

1.2.4. Serpins (serine proteinase inhibitors)

There are several notable features in which serpins differ from the standard mechanism inhibitors. Serpins are single domain proteins of about 400 amino-acid residues which are abundant in plasma often in variably glycosylated forms [152, 202]. Like the canonical inhibitors, they interact with their target enzyme in a substrate-like manner through the exposed loop of poorly defined structure. In the case of serpins, however, cleavage of the P1-P1' peptide bond leads to dramatic structural and stability changes. In contrast to canonical inhibitors, the reactive site loop of serpins is lexible and can assume a number of different conformations. Serpins are the only family of serine proteinase inhibitors for which complex formation with non-serine enzymes cysteine proteinases and aspartyl proteinases has been demonstrated [93, 118, 125, 145].

1.2.5. The Kunitz-type inhibitor family

The Kunitz type inhibitors of serine proteases form an important field of scientific investigation due to their diverse role. It began when Kunitz isolated and characterized a trypsin inhibitor from *Glycine max* seeds [99, 100]. Since then, vast number of inhibitors has been isolated and their structures solved. Plant Kunitz inhibitors are extensively distributed in plants and are largely concentrated in leguminous seeds of taxonomic subfamilies Mimosoideae, Caesalpinioideae and Papilionoideae. They mainly inhibit serine, cysteine and aspartic

proteinases and no inhibition of metalloproteinase was reported by these inhibitors. The Kunitz type inhibitors of serine proteinases are further subdivided into two families namely bovine pancreatic trypsin inhibitor (BPTI) and soybean trypsin inhibitor (STI). The inhibitors from BPTI family have molecular weight of about 6.5 kDa and three disulfide bridges. STI family inhibitors have molecular mass of about 20 kDa with one or two polypeptide chains and less cysteines [96, 110, 140]. Most of them are single polypeptide chain of about 20 kDa with two disulfide bridges and single reactive site (see Fig. 1.3) [91, 107]. The Kunitz STI inhibitors are included as Kunitz-P inhibitors in the I3A family of the MEROPS database (http://merops.sanger.ac.uk) [159].

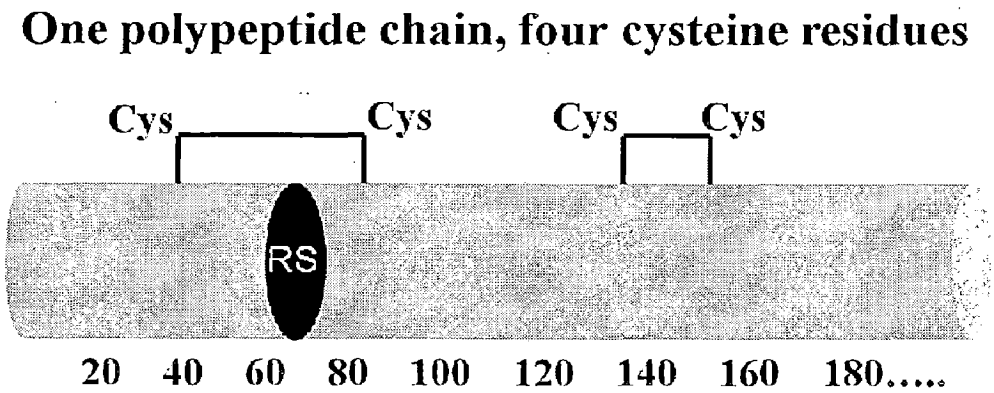


Figure 1.3. The primary structure of classical plant Kunitz STI type inhibitors having one polypeptide chain with two disulfide bonds and single reactive site (RS). Adapted from Oliva *et al*, 2010 [142]

1.2.6. Biochemical and Biophysical studies of Kunitz inhibitors

1.2.6.1. Isolation, purification and general properties of Kunitz STI type inhibitors

The first plant proteinase inhibitor to be isolated and characterized was soybean trypsin inhibitor (STI) by Kunitz. STI also inhibits chymotrypsin weakly. It is inactivated by heat and by gastric juice. The purification, crystallization, kinetics of the interaction and complex formation of STI with trypsin comprise a major landmark in the study of protein proteinase inhibitors [99, 100]. The numerous studies on STI concerning specificity, stability, physical, kinetic and other properties have been compiled and summarized by Kassell [86] and Birk [17]. STI consists of

181 amino acid residues and includes two disulfide bridges. The precursor for STI has a molecular mass of ~24,000 Da and the mature STI has a molecular mass of ~20,000 Da.

The Kunitz type inhibitors from other than soybean, the *Erythrina* seeds received the greatest amount of attention. The inhibitors have similar molecular mass (~20,000 Da) and share many other chemical characteristics. They are divided into three groups on the basis of their relative abilities to inhibit chymotrypsin, trypsin and tissue plasminogen activator (tPA). Group (a) inhibitors were relatively specific for chymotrypsin; they were poor inhibitors of trypsin and had no apparent effect on tPA. Group (b) proteins inhibited trypsin strongly and chymotrypsin slightly less effectively. They had no effect upon tPA. Group (c) inhibitors inhibited trypsin, chymotrypsin and tPA. The sequence of these inhibitors shows a high degree of homology to those of Kunitz-type trypsin inhibitors from soybean and winged bean seeds. Onesti *et al*, also investigated the effects of pH and temperature on kinetic and thermodynamic parameters for the binding of ETI to different serine proteinases and defined their perspective affinities [144, 165].

Studies on *Erythrina variegata* trypsin and chymotrypsin inhibitors were reported by Kouzuma *et al,* [94]. The stoichiometry of trypsin inhibitors with trypsin was 1:1, while that of the chymotrypsin inhibitor with chymotrypsin was 1:2 molar ratio. According to Kouzuma *et al* the inhibitors show structural features characteristic of the Kunitz-type soybean trypsin inhibitor and exhibit a significant homology to the storage proteins, sporamin in sweet potato and the taste-modifying protein, miraculin, in miracle fruit. A proteinase inhibitor from *Enterolobium contortisiliquum* seeds was isolated, characterized and sequenced by Batista *et al*, [12]. The inhibitor contains 174 amino acid residues in two polypeptide chains, an alpha-chain consisting of 134 residues and a beta-chain made up of 40 residues, linked by a disulfide bridge. The inhibitor displays a high degree of sequence identity with other Kunitz-type proteinase inhibitors isolated from the Mimosoideae subfamily. Several Kunitz type inhibitors have been reported since the discovery of soybean trypsin inhibitor. Some of them are studied well to know the mechanism of inhibition and their detailed study of protease-inhibitor interaction.

1.2.6.2. Solubility properties

To be biologically active, proteins must adopt specific folded three-dimensional tertiary structures in soluble condition. Protein solubility is a complex function of the physiochemical nature of the proteins, pH, temperature, organic solvents, detergents and the concentration of the salts used. Protein solubility is highly dependent upon the ionic strength. The spectroscopic and

biochemical studies have shown that the Kunitz type inhibitors are highly stable over a broad range of temperature and pH and are remarkably resistant to proteolytic degradation [6, 144, 191].

1.2.6.3. Circular dichroism analysis and fluorescence studies

Circular dichroism (CD) spectroscopy is a well-established method to understand the structure- function relationship of proteins. The sensitivity of far-UV protein CD spectra to protein secondary structure is used in one of the most successful applications of CD to determine the secondary structure composition of a protein. Based on circular dichroism spectra, soybean Kunitz trypsin inhibitor (SKTI) has been classified in the β-II or the 'disordered' class of all antiparallel β-sheet proteins [11, 124, 166]. CD spectra of SKTI during denaturation in the far-UV region show decreased ellipticity, indicating loss of secondary structure [16, 126, 167]. Fluorescence spectroscopy is a very powerful technique to monitor conformational changes in proteins. The fluorophores can be either intrinsic (tryptophan) or extrinsic (ANS: 8-anilino-1-naphthalene sulfonate) probes. The fluorescence property of tryptophan has long been utilized to understand the folding/unfolding, substrate binding and conformational heterogeneity in different physicochemical conditions [127, 128].

1.2.6.4. Structural studies

Protein crystallography is very important and essential technique to determine the three dimensional structure at molecular level. The crystallographic studies help in understanding the structure-function relationship and mechanism of action of protein. The interaction between the macromolecules which is important for molecular recognition can be studied by determining the complex structures. Many proteins including antigen, antibody, hormone, receptors and various enzymes like proteinases and substrate/inhibitors have been characterized by X-ray crystallography.

The crystallographic studies have shown that Kunitz soybean trypsin inhibitor (STI) family members are predominantly β-sheet proteins with little or no α-helical structure. They consists of 12 antiparallel β-strands forming β-trefoil fold. The three-dimensional structure of STI in complex with porcine trypsin was also determined [18, 191]. The complex structure shows that P1 Arg residue enters into S1 pocket of trypsin and forms salt bridge with Asp189. The complex is stabilized by network of hydrogen bonds between reactive site residues of STI

and active site pocket of trypsin [187]. The predominant conformation found in STI has been demonstrated to be approximate beta-sheet structures, with a small amount of regular sheet. Onesti *et al*, studied the structure and properties of a Kunitz-type trypsin inhibitor from *Erythrina caffra* seeds [192]. This inhibitor consists of 172 amino acid residues with two disulfide bridges. The three-dimensional structure of ETI consists of 12 antiparallel beta-strands connected by long loops. Six strands form beta-barrel and the other six form a "lid" like structure. The scissile bond (Arg63-Ser64) of ETI is located on an external loop that protrudes from the surface of the molecule. The overall structure of ETI is similar to the partial structure of STI.

Several structures of Kunitz STI type inhibitors have been reported showing that they have β-trefoil fold consisting of 12 anti-parallel β-strands connected by long loops. Their exposed reactive site loop has a characteristic canonical conformation with Arg/Lys at P1 position [25, 41, 96, 144, 187]. Conserved Asn (Asn13 in STI) residue is observed in structures of Kunitz STI inhibitors which maintains the canonical conformation of the reactive site loop [40, 187]. The structure of STI complex with porcine trypsin reveals that P1 residue; Arg63 enters in S1 pocket of trypsin and makes a salt bridge with Asp189 and inhibit it in a substrate-like manner [187]. Moreover, some structural reports show variation in structures compared to the structures of classical Kunitz STI inhibitors. For example, CTI has β-trefoil fold composed of two non-covalently bound polypeptide chains with only single disulfide bridge but its reactive site loop is still able to maintain canonical conformation [95]. DrTI has one amino acid insertion between P1 and P2 of reactive site distorting its conformation [96]. BbCI, a Kunitz type inhibitor has conservative β-trefoil fold but lacks disulfide bonds [65]. Crystal structure of MKMLP from seeds of *Murraya koenigii* having trypsin inhibitory activity also shows conservative β-trefoil fold but has seven cysteines forming three disulfide bridges, Asn65 at P1 instead of Arg/Lys, Asn13 replaced by Ala13 and carbohydrate moieties linked to Asn64 [56, 185]. API-A, a double-headed arrowhead protease inhibitor similar to Kunitz STI inhibitor have β-trefoil fold but possess two P1 residues Leu87 and Lys145 [10].

1.2.7. Functions of Kunitz type inhibitors

The high content of protease inhibitors in seeds of various plant species motivated the interest in their physiological function. It is suggested that they contribute in storage of nutrients, defense against predators, besides regulation of proteolysis by controlling proteolytic enzymes activity during quiescence and germination. The Kunitz protease inhibitors possess inhibitory activity on various proteolytic enzymes such as trypsin, chymotrypsin, and similar enzymes involved in digestive processes, blood coagulation enzymes such as plasma kallikrein, factors XIIa, XIa, and Xa enzymes such as plasmin, fibrinolytic system enzymes like tissue plasminogen activator and enzymes involved in inflammatory processes such as elastase and cathepsin G. Some inhibitors of this family shows multifunctional role (Table 1.2), as is the case of BbCI that inhibits the cysteine proteinases cathepsin L and cruzipain and the serine proteinases HNE (human neutrophil elastase) and cathepsin G [30, 114, 142, 155].

Inhibitor	Inhibitory Activity Kiapp (nM)	Studies on Biological Models
EcTI	Human plasma kallikrein = 1.0 Human plasmin = 10.0 Human FXIIa = 100.0 Bovine chymotrypsin = 100.0 Bovine trypsin = 1.0	Prolongs the coagulation time
LITI	Human plasma kallikrein = 6.3 Human plasmin = 3.2 Bovine trypsin = 0.25 Bovine chymotrypsin = 0.14	Prolongs the coagulation time; inhibits kinin release from HMWK by human plasma kallikrein; decreases paw edema induced by carrageenin in male Wistar rats
ApTIA	Human plasma kallikrein = 550.0 Bovine trypsin = 2.31 Alpha-chymotrypsin = 10.0	Prolongs the coagulation time; antifungi properties on *Aspergillus niger*, *Thielaviopsis paradoxa* and *Colletotrichum* sp. P10
PdKI	Bovine trypsin = 35.6 Papain = 761.0	Inhibits digestive proteinases from Coleopteran, Lepidopteran and Dipteran pests.
BuXI	Human plasma kallikrein = 6.9 Human plasmin = 76.0 Human FXIIa = 74.0 Human FXa = 18.4 Bovine trypsin = 28.0 Bovine chymotrypsin = 2.7	Prolongs the coagulation time; inseticidal activity on *Abracris flavolineata*, *Musca domestica* and *Spodoptera frugiperda* midgut; inhibits a prothrombin activator proteinase isolated from *Lonomia obliqua* venom
BvTI	Human plasma kallikrein = 23.0 Human FXIIa = 110.0 Bovine trypsin = 2.1 Bovine chymotrypsin = 12.0 Murine plasma kallikrein = 2.2	Inseticidal activity on *Abracris flavolineata*, *Musca domestica* and *Tenebrio molitor* midgut
BrTI	Human plasma kallikrein = 14.0 Bovine trypsin = 2.9 Murine plasma kallikrein = 13.0	Inseticidal activity on *C. maculatus* larvae
BbKI	Human plasma kallikrein = 2.4	Prolongs the coagulation time

	Human plasmin = 33.0 Bovine trypsin = 2.0 Bovine chymotrypsin = 2,600.0 Porcine pancreatic kalikrein = 200.0 Murine plasma kallikrein = 5.2	
BbCI	Human neutrophil elastase = 5.3 Cathepsin L = 0.22 Cathepsin G = 160.0 Bovine pancreatic elastase = 40.0 Cruzipain = 1.3 Cruzain = 0.3	Decreases pulmonary edema in isolated perfused rabbit lungs
CeKI	Human plasma kallikrein = 3.1 Human plasmin = 0.18 Human FXIIa = 0.18 Human FXa = 490,000.0 Bovine trypsin = 21.5	Prolongs the coagulation time
DrTI	Human plasma kallikrein = 5.25 Bovine trypsin = 21.9	Inseticidal activity on *C. cephalonica*
DtTCI	Bovine trypsin = 0.17 Bovine Chymotrypsin = 0.125	Antiparasitic agent

Table 1.2. Inhibitory Properties of Kunitz protease Inhibitors from Leguminosae. EcTI, *Enterolobium contortisiliquum* trypsin inhibitor, LlTI, *Leucaena leucocephala* trypsin inhibitor, ApTIA, *Acacia plumosa* trypsin inhibitor A, PdKI, *Pithecelobium dumosum* trypsin–papain inhibitor, BuXI, *Bauhinia ungulata* factor X inhibitor, BvTI, *Bauhinia variegata* trypsin inhibitor, BrTI, *Bauhinia rufa* trypsin inhibitor, BbKI, *Bauhinia bauhinioides* Kallikrein inhibitor, BbCI, *Bauhinia bauhinioides* cruzipain inhibitor, CeKI, *Caesalpinia echinata* kallikrein inhibitor, DrTI, *Delonix regia* trypsin inhibitor, DtTCI, *Derristrifoliata* trypsin chymotrypsin inhibitor. Adapted from Oliva *et al*, 2011 [114].

1.2.7.1. Role in plant defense

Kunitz inhibitors are considered important for plant defense against insects. The inhibition of digestive proteinases, leading to inadequate digestion and absorption of essential amino acids, interferes with growth and may lead to starvation. This effect can be observed when insects are fed with artificial diets containing specific inhibitors of the major class of their gut proteinases. In addition, a synergistic action of the inhibitors may occur, developing insecticides more effective by mixing different classes of inhibitors. With the progress of the studies on molecular structures, cloning and gene expression, new inhibitors against predatory insects, fungi with potential use in plant genetic engineering to develop transgenic resistant plants were characterized. Serine proteinase inhibitors from plants have been shown to have potential usefulness as defense tools to protect the plants from invading pests. A few plant proteinase inhibitors have been shown to possess fungicidal activity

1.2.7.2. Pharmacological role

The pharmacological action of proteinase inhibitors has been studied. The antitumor action of proteinase inhibitors was mainly reported in Bowman-Birk and Kunitz families. Trypsin inhibitors from *Peltophorum dubium* (PDTI) and soybean (STI), both belonging to the Kunitz family, induce cell death in human leukemic Jurkat cell line of human leukemia, via activation of caspase-3 and -8. In addition, significant release of cytochrome c by mitochondria was not observed, suggesting that the intrinsic mitochondrial pathway is not predominant in this apoptotic process. The inhibitor from seeds of Chinese black soybean *Glycine max* suppressed proliferation of breast MCF-7 and hepatoma HepG2 cancer cells.

Characterization of Kunitz inhibitors and their involvement with proteolytic enzymes have been investigated in the control of other pathophysiological processes such as inflammation, thrombosis, AIDS, parasitic diseases and fungal infection. The action of plant peptidase inhibitors as antibiotic and antiparasitic agents is attributed to their interference with host protein catabolism causing lowered amino acid availability. This interference would delay the synthesis of proteins necessary for the development, growth, and invasion.

The serine protease inhbibitors play an important role in regulating many physiological processes, such as inflammation, coagulation, fibrinolysis, complement activation intracellular protein breakdown, cell cycle, transcription and apoptosis [58, 77, 188, 197]. Because of all these

pharmaceutical advantages, the inhibitors can have many potential applications in different fields of biotechnology. Apart from this, the complex with their cognate proteases is an important model for studying protein-protein interaction.

1.3. Lectins

The specific recognition of carbohydrates by lectins is so remarkable that they are commonly defined as carbohydrate binding proteins (other than enzymes or antibodies) capable of specific and reversible interaction with the carbohydrates [92, 182]. Lectins, a group of highly diverse non-immune origin proteins ubiquitously distributed in plants, animals and fungi which have at least one non-catalytic domain that facilitates them to selectively recognize and reversibly bind to specific free sugars or glycans present on glycoproteins and glycolipids without altering the structure of carbohydrate. Several reports have shown the importance of carbohydrates and lectins in immunology, oncology or medicine and discussed various methods of lectin classification or gave detailed accounts of specific lectins and their properties [60, 92]. Studies on lectins are performed to characterize the structure of lectin molecules, their carbohydrate-binding specificity, conformational/functional properties of their carbohydrate-binding 'pocket', and their biological significance. The application of lectins with known specificity as powerful tools to study the function of glycoconjugates, both in solution and on cell surfaces [219]. Lectins have been extensively used for preparative and analytical purposes in biochemistry, cell biology, immunology and related areas.

1.3.1. Lectin research: History and Occurrence

1.3.1.1. Historical aspects

The search for the toxic principle in castor beans (*Ricinus communis*, Euphorbiaceae) prompted Rudolf Kobert, one of the most prominent pharmacologists, to ask his medical student Hermann Stillmark to study this plant. Stillmark described that extracts from castor beans and four other Euphorbiaceae plants are able to agglutinate blood cells from different animals, i.e. rabbits, horses, dogs and cats. He thought that toxicity and agglutinating capability to originate from the same substance, ricin, which he regarded to be an enzyme. As we now know, castor beans contain a tetrameric protein called *Ricinus communis* agglutinin that is able to agglutinate cells but is hardly toxic. A second, dimeric protein, now called ricin or *Ricinus* toxin, is closely

related to the agglutinin, and its enzymatic subunit acts as a highly specific RNA N-glycosidase on 28S rRNA but is only a weak agglutinin. Soon after Stillmark's discovery, ricin and the related toxin from jequirity beans (*Abrus precatorius*, Leguminosae) played a fundamental role as model antigens in the pioneering studies of Paul Ehrlich. Looking back to the starting of lectin research, progress in the field was hindered in the first decades due to crudity of the fractionation techniques. The first agglutinin to be isolated was concanavalin A from Jack bean (*Canavalia ensiformis*) seeds. The eminent American biochemist James B. Sumner succeeded in purifying also the agglutinating principle from these seeds by crystallization and called it concanavalin A (ConA). He discovered that ConA is capable to interact with RBCs, starch, glycogen and mucins, and that this interaction can be prevented by low-molecular weight carbohydrates such as sucrose [62, 169, 189]. This result was the first clear experimental sign that an agglutinin binds carbohydrates and had significant implications for the study of cell membrane constituents. Systematic screening of plant extracts for agglutinating activities led to a breakthrough in haematology and paved the way for coining the term lectin. In fact, Landsteiner's discoveries of human isoagglutinins in 1900 in Vienna and the species specificity of plant agglutinins and his comparison of haemagglutinating/ haemolyzing activities with natural antibodies lead W. C. Boyd "to test seeds for blood group specificity". he found A-type specificity with extracts of the lima bean (*Phaseolus lunatus limensis*, Leguminosae) and proposed "the term 'lectin' (from the Latin legere, to choose or to pick out) for these and other antibody like substances" [23, 63, 133]. From this starting point, the chemical nature of the blood group substances as oligosaccharides could be explained. With this focus on haemagglutination it is no surprise that the initial definition of the term lectin placed special emphasis on just this aspect [169]. Lectin study had progressed beyond the serological level and it was a matter of time before molecular level analysis of lectin activity would start in earnest [92, 153].

1.3.1.2. Occurrence of lectins

Since the discovery of ricin (*Ricinus* agglutinin) and ConA, it is known that the richest source for most lectins is seeds or, more generally, the storage organs of plants. Lectins are studied extensively not only from seeds bur also from roots (*Urtica, Phytolacca, Sambucus, Trichosanthes, Calystegia*), tubers or bulbs (*Solanum, Galanthus, Scilla, Allium, Crocus, Tulipa, Iris*), bark (*Sambucus, Sophora, Robinia, Maackia, Laburnum, Cytisus, Cladrastis, Hevea, Abies*) or leaves (*Aloe, Lactuca, Vicia unijuga, Viscum album*) [169, 200]. Inside the cells, lectins are primarily found in protein bodies. They are synthesized in the endoplasmic reticulum (ER)

and transported via the Golgi apparatus and originate by subdividing the vacuole. Analysed from their origin and their role for protein turnover, protein bodies are related to lysosomes. The main content of protein bodies are storage proteins (vicilin, legumin and convicilin in Leguminosae, related proteins in other plants), lectins, hydrolases (glycosidases, phosphatases) and phytin to store phosphate.

1.3.2. Definition of lectins and their classification: the ongoing debate

Lectins display considerable specificity in binding oligosaccharides and still possess huge structural diversity. Such a group which is diverse both in structure and function, arriving at a common definition that suitably describes all their characteristics is obviously not simple. The initial definition of lectins as synonymous with agglutinins made way for a newer and more general definition of lectins: proteins that possess at least one non-catalytic domain, which binds reversibly to specific carbohydrates. By this definition, agglutination no longer remained the pivotal property by which a lectin is defined, although in practice it continued to be a useful marker. Carbohydrate binding and specificity became the new criteria for the definition of a lectin. To qualify as lectin today, a (glyco)protein must meet three distinct requirements: A lectin is a (glyco)protein that binds carbohydrate. (2) Lectins are separated from immunoglobulins. (3) Lectins do not biochemically modify the carbohydrates which they bind [169].

A useful and popular method of classifying lectins, particularly from plant sources, is based on the mono- or disaccharide specificities of the lectins, even though as mentioned earlier, amongst lectins where subsite binding is important, the oligosaccharide specificities may be very dissimilar despite similar monosaccharide specificities. Yet perhaps due to both historical and practical reasons, this method of classification has continued to be popular. However, it is difficult to apply such a classification universally. For example, applying such a framework it would simply not be promising to classify a lectin like galectin-10, whose soluble form apparently binds galactose but the crystalline state recognizes mannose instead. Similarly, jacalin, the T-antigen binding lectin from jackfruit seeds, is capable of binding both galactose and mannose sugars. Previous studies also show that plant and animal lectins could have one or more hydrophobic binding sites different from their carbohydrate binding site(s) and these could play a role in protein–protein interactions. A number of new questions arise from these studies. Are lectins multi-functional proteins invested with the power to trigger biological activity by using additional binding sites to their carbohydrate binding ones? What functions do these

recognitions confer upon lectins and how important were they to the role of these proteins in biological systems? Moreover, there are also proteins that have single or weak carbohydrate recognition domains (CRDs) [64, 92].

Yet another group of proteins which could be classified as lectins based on extensive primary sequence homology with certain lectins without possessing any marked saccharide binding ability. Lis and Sharon proposed a method of classification based on structural features of the proteins of interest. Thus they classified lectins into three groups: simple, mosaic and macromolecular assemblies [111]. Van Damme *et al,* classified plant lectins are into 12 different families of evolutionary and structurally related lectin domains according to their different carbohydrate-binding specificities, such as Agaricus bisporus agglutinin homologs, (2) Amaranthins, (3) Class V chitinase homologs with lectin activity, Cyanovirin family, (5) EEA family, (6) GNA family, (7) proteins with hevein domains, (8) Jacalins, (9) proteins with legume lectin domains, (10) LysM domain, (11) Nictaba family (formerly Cucurbitaceae phloem lectins), (12) Ricin- B family [204] (Table 1.3).

As new terrains are explored, new questions come up as well. Should lectins be classified based on their carbohydrate specificities, or by comparing with other lectin-like proteins which share sequence homology with them, or with those with whom they share a common fold? Or should it be a judicious combination of the three? Or should the most significant biological function of the protein be the ultimate consideration? These questions need to be addressed in order to arrive at a classification that can be followed by all workers in this field [67, 92, 134].

Representative lectin	Abbreviation	Family	Sugar-binding specificity
Agaricus bisporus agglutinin	ABA	Agaricus bisporus agglutinin homologs	Galactose
		Amaranthins	
Chitinase-related agglutinin	CRA	Class V chitinase homologs with lectin activity	High-mannose N-glycans
Cyanovirin-N	CV-N	Cyanovirin family Mannose	
Euonymus europaeus agglutinin	EEA	EEA family	Mannose/galactose
Polygonatum cyrtonema lectin	PCL	GNA family	Mannose/sialic acid
Wheat germ agglutinin	WGA	Proteins with hevein domains	N-acetyl-d-glucosamine
Jacalin	JAC	Jacalins	Mannose
Concanavalin A		ConA Proteins with legume lectin domains	D-mannose
		LysM domain	
Cucurbitaceae phloem lectin	CPL	Nictaba family	
European mistletoe lectin	ML-I	Ricin-B family	Beta-galactose

Table 1.3. The classification of plant lectins according to Van Damme *et al.*, 2008 [204].

1.3.3. Chitinase-like lectins (chi-lectins)

A lectin from black locust (*Robinia pseudoacacia*) bark shares approximately 50% sequence identity with plant class V chitinases but is essentially devoid of chitinase activity. Specificity studies indicated that the black locust chitinase-related agglutinin (RobpsCRA) preferentially binds to high-mannose N-glycans comprising the proximal pentasaccharide core structure. So this lectin has been placed in new family of plant lectins that is structurally and evolutionary closely related to class V chitinases of the glycoside hydrolase 18 (GH18) family which is an ancient chitinase family found in all kingdoms from bacteria to fungi, animals, and plants. Moreover, sequence data clearly point out that all GH18 chitinases, including those from plants, have a common ancestor. This, taken together with the fact that GH18 chitinases or corresponding genes have been found in numerous plant species, implies that the GH18 structural scaffold is quite common in higher plants. Considering the apparent confinement of RobpsCRA orthologs to the legume family, it is intersting to hypothesize that a catalytically active chitinase from an ancient legume species or possibly an earlier ancestor served as a structural scaffold for the development of a small family of carbohydrate-binding proteins. This evolutionary process most likely involved gene duplication followed by neofunctionalization [203].

Interestingly, a similar conversion of a chitinase into a lectin, but catalytically inactive, homolog also occurred in higher animals. YKL-40 (Human cartilage glycoprotein-39, chitinase 3-like protein 1) is a mammalian protein which in humans is expressed by a variety of cell types including synovial cells, chondrocytes and smooth muscle cells, and neutrophils and macrophages. Expression appears to be induced by a change in the extracellular matrix environment, is thought to have a role in development, tissue remodelling and inflammation, and has been associated with diseases including arthritis, cancer and liver fibrosis. YKL-40 binds specifically to collagen triple helices and regulates cleavage and fibril formation. YKL-40 binds chitin and chito-oligosaccharides using 9 GlcNAc-binding subsites. Heparan sulphate, found on proteoglycans, is also a potential ligand and cell surface receptor for YKL-40 [70, 76]. Yml is a murine protein secreted in large quantities by peritoneal macrophages upon nematode infection and is a marker for alternatively activated macrophages, which exert anti-inflammatory effects and promote wound healing, as well as combating parasitic infections. The function of Yml is not known, and there is conflicting evidence regarding both chitinase activity (weak or absent) and sugar-binding activity. SPR studies suggested that Yml exhibits calcium-independent

binding at low pH to glucosamine (GlcN) and galactosamine (GalN), and especially to oligomers of these sugars, including heparin, but not to the unsubstituted or N-acetylated sugar [83, 129, 190]. However, both plants and mammals used a homologous chitinase as a structural scaffold, there are important differences in the conversion of a carbohydrate modifying into a carbohydrate-binding protein [130].

1.3.4. Isolation of lectins

In the beginning, purification of lectins followed the system used for proteins in general without exploiting their special characteristics. The methods that were used included precipitation by salts, acids and organic solvents. Naturally, the preparations thus obtained were far from pure, and results acquired with them have to be interpreted very carefully. Progress in this area was achieved by the introduction of preparative chromatographic methods with ion exchangers, gel filtration chromatography, affinity adsorbents and cross-linked dextrans found versatile application in one-step purification schemes for Glc-binding plant lectins. Agarose, a commercially available natural polysaccharide containing Gal, was utilized as an affinity adsorbent for Gal-binding lectins like those from *Ricinus communis, Bauhinia purpurea, Glycine max* and *Wisteria floribunda* [85, 115, 169, 220].

1.3.5. Structural studies and folds of lectins

In the fields of lectin research, structural investigations began with the lectin from the Jack bean *Canavalia ensiformis* using x-ray crystallography [48]. ConA mainly built up of β-sheets that are connected by loops whereas α-helical elements are almost absent. In the subsequent years, this folding pattern was likewise seen for various other legume lectins by x-ray crystallography, circular dichroism and calculation of hydropathic profiles from amino acid sequences. Over the years, however, x-ray crystallography established to be the method of choice to map plant lectin structure. Despite their diversity in carbohydrate-binding specificity, the folding patterns of subunits of legume lectins are superimposable. This jelly-roll motif comprises a nearly flat six-stranded 'back' β-sheet, a curved seven-membered 'front' β-sheet, a short five-membered 'top' β-sheet, which has an important role in holding the two larger sheets together, and a number of loops interconnecting the sheets, as well as the strands that they contain.

The structural information available on plant lectins exhibiting other folds is much less extensive, although the second lectin to be examined using X-ray crystallography was the cereal

lectin from wheat germ, involving hevein domains [122, 218]. The threefold symmetric β-trefoil fold has been taken up by lectins and lectin-like carbohydrate-binding proteins. Two other novel threefold symmetric lectin folds have been revealed. One, which of jacalin, occupies three Greek-key β sheets, with the strands oriented nearly parallel to the threefold axis (β-prism I). The other, first observed in snowdrop lectin, encloses three four-stranded β-sheets that are nearly perpendicular to the threefold axis (β-prism II). The structure of another lectin, that from *Maclura pomifera*, with a β-prism I fold and those of garlic, daffodil and bluebell, with β-prism II fold have recently solved [211]. Following to the legume lectin fold, the C-lectin fold is the most broadly observed structural motif in lectins. The structure containing this domain includes endostatin, intimin, CD94, tetranectin, an antifreeze protein and human lung surfactant protein [211]. Three new lectin folds were fully characterized recently are the I-lectin fold in a sialoadhesin domain, the P-lectin fold in a cation dependent mannose 6-phosphate receptor and a fivefold symmetric β-propeller fold in tachylectin. So considering the structural folds, plant lectins are grouped into seven folds, which are con-A like β-sandwich, β-prism-I, β-prism-II, β-trefoil, knottins, double-psi-β-barrel, and P-domain of calnexin/calreticulin folds. Additionally, some plant lectins adopting a 6/7-bladed β-propeller fold and a double-stranded β-helix fold [29, 132].

1.3.6. Carbohydrate-binding site.

There is significant progress in illuminating the features of lectins involved in carbohydrate binding in few past years. X-ray crystallography of the proteins complexed with their ligands, site-directed mutagenesis experiments and molecular modelling have permitted the recognition of the chemical groups belonging to both interacting species involved in the binding and of the types of bond formed. Studies of lectin–oligosaccharide complexes provide the basis for the understanding of the proteins interaction with natural ligands. In general, lectins show specificity for di-, tri- and tetrasaccharides. Carbohydrate-binding sites are frequently shallow depressions that are found on the surface of the protein. In all legume lectins, despite of their specificity, four invariant amino acid residues participate in the ligand binding: an aspartic acid, an asparagine, a glycine (conserved in all the lectins of the family apart from Con A) and an aromatic amino acid or leucine (Asp83, Gly104, Asn127 and Tyr125 for PNA) [3, 9, 209, 210]. Replacement of the aspartic acid or asparagine by site-directed mutagenesis results, in several cases, in the loss of the lectin sugar-binding ability [1, 212]. However, having conserved key amino acids in the carbohydrate binding, different legume lectins can show different specificity.

Therefore, while the group of highly conserved amino acids provides the framework required for binding, specificity apparently arises from the variability of amino acid residues in other regions of the combining pocket. Lectins bind carbohydrates through a network of hydrogen bonds and hydrophobic interactions. Contacts between the ligand and the protein are often mediated by water molecules. Water acts as a molecular "mortar" its small size and its ability to behave as both hydrogen donor and acceptor make it near-ideal for this function. Thus, water plays a major role in carbohydrate recognition, providing in some cases exquisite specificity. Generally, water molecules in the carbohydrate-binding region mimic the ligand to a significant extent not only at the primary site, but also in the regions adjacent to it [3].

1.3.7. Physiological functions of lectins

Despite their long history, the true physiological function of plant lectins is still not well understood. Many hypotheses have been formulated in the course of the years but, at present, no physiological function for any lectin has been established with certainty. The difficulty in assigning a precise role arises from several of their features. The defining characteristic of all lectins is their carbohydrate-binding ability. This activity has been preserved during evolution, suggesting that it is essential for the exploitation of their function. Even for lectins with homologous sequences a common function cannot be readily approved to them because individual parameters such as carbohydrate specificities, location and time of appearance differ. Therefore, every description of a function requires specific evidence which cannot be substituted by analogy considerations. Rigorous investigations need to be carried out for an understanding of the biological roles of individual lectins.

1.3.7.1. Protection from insects

Previous reports show that some lectins have broad insecticidal activity. Feeding trials to assess the insecticidal action of lectins as potential biological pesticide have focused on the Man-binding lectin GNA, as already mentioned. After initial experiments in which various insects were fed with artificial diets supplemented with GNA, transgenic plants containing the GNA gene were increasingly used. In a study in which ladybirds fed on aphids that had been sucking on transgenic GNA-containing potato plants, adverse effects on the development of the coleopterous predators were observed [169, 179]. Numerous reports have already shown that expression of plant lectins in genetically engineered (crop) plants and/or incorporation of the

purified lectins in artificial diets negatively affect the performance of many pest insects belonging to different orders such as Lepidoptera, Coleoptera, Diptera and Hemiptera [207].

1.3.7.2. Protection from fungi

Along with the insecticidal action of plant lectins the spectrum of external functions is certainly not yet completely covered. Binding to cell wall constituents in fungi can impede their growth. Fungal cell walls have chitin, the β1 → 4 linked polymer of GlcNAc. The GlcNAc-binding lectins action as fungicidal might be due to targeting the chitin. In fact, this is the case for several lectins in vitro. In order to unequivocally demonstrate the fungicidal action of a lectin, however, it is important to avoid the use of lectins that are contaminated by other fungicidal proteins [169].

1.3.7.3. Symbiosis with bacteria

Plant lectins are not only restricted as a part of defense system of plant. However, their interaction with cell surface compounds is also considered to initiate new and desired contacts. Several plants, in particular Leguminosae family members are known for their capacity to establish a symbiosis with soil bacteria of the genus *Rhizobium* and related genera which are capable of fixing atmospheric nitrogen without supply of external nitrogen fertilizer. This symbiosis is species specific: a given *Rhizobium* strain will nodulate only a single legume species or at best a limited number. It is thus suggestive, together with the widespread and abundant occurrence of lectins in the seeds of Leguminosae members that lectins might participate in establishing the symbiosis between plant roots and rhizobial bacteria. Initial studies appeared to support this hypothesis but were questioned by further scrutiny [168]. A new characteristic emerged with the detection of the nodulation (Nod) factors, i.e. lipochitooligosaccharides that are produced by the bacteria as a response to plant-derived stimuli (flavonoids). That a lectin is part of the communication network involved in nodulation is substantiated by the transformation of clover roots with the pea lectin gene.

1.3.7.4. Potent inhibitors of coronaviruses

The antiviral activity of plant lectins with specificity for different glycan structures against the severe acute respiratory syndrome coronavirus (SARS-CoV) and the feline infectious peritonitis virus (FIPV) in vitro was evaluated. The plant lectins possessed marked antiviral properties against both coronaviruses with EC50 values in the lower microgram/ml range. The

strongest anti-coronavirus activity was found mostly among the mannose-binding lectins. In addition, a number of galactose-, N-acetylgalactosamine-, glucose-, and N-acetylglucosamine-specific plant agglutinines displayed anti-coronaviral activity. Two possible targets for the lectin as a antiviral were identified. The first target most possibly viral attachment and the second target is at the end of the infectious virus cycle [89, 222].

1.3.7.5. As antitumor agents

Plant lectins including Ricin-B family, proteins with legume lectin domains and GNA family can induce cancer cell death through targeting programmed cell death pathway [54]. *Polygonatum cyrtonema* lectin (PCL), a mannose/sialic acid lectin belonging to GNA family has drawn many interests for its anti-tumor activities toward HeLa, MCF-7, A375 and L929 cells. Importantly, accumulating evidence has exposed that PCL can induce cancer cell death targeting PCD pathways, such as the caspase-dependent pathways, mitochondrial ROS-p38-p53 pathway, Ras-Raf and PI3K-Akt pathways [181, 213]. ConA, a legume lectin was reported to induce apoptosis in balb/c 3T3 (3T3) and diploid human gingival fibroblasts (HGF). Recent studies have demonstrated that ConA have apoptosis-inducing activities, and initiation of this apoptotic cell death can be mediated by mitochondria [112, 199].

1.3.8. Applications of plant lectins

Plant lectins have been studied well with structural characteristics. Numerous questions in basic and medical sciences have been and are being addressed by exploiting plant lectins (Table 1.4) [169]. Laboratory manuals illustrate that lectin-involving methods have well matured and attained the category of routine handling. Plant lectins have made themselves known to assess a defined aspect of activity of the complex machinery for glycan assembly and modification, starting with synthesis of nucleotide sugars and their transport into the Golgi lumen [160, 169]. The development of sequential lectin affinity chromatography had a major impact on glycan isolation and achieved access to prominent sequence information. The topological route how these lectin-reactive epitopes are transported after their synthesis and then presented on the cell surface plus their lateral movements in the membrane are readily chased by lectin cytochemistry using fluorescent or biotinylated lectins or lectin-coated colloidal gold granules [169, 180].

Biochemistry • Detection of defined carbohydrate epitopes of glycoconjugates in blots or on thin-layer chromatography plates • Purification of lectin-reactive glycoconjugates by affinity chromatography • Glycan characterization by serial lectin affinity chromatography • Glycome analysis (glycomics) • Quantification of lectin-reactive glycoconjugates in enzyme-linked lectin-binding assays • Quantification of activities of glycosyltransferases/glycosidases by lectin-based detection of products of enzymatic reaction
Cell biology • Characterization of cell surface presentation of glycoconjugates and their preceding intracellular assembly and routing in normal and genetically engineered cells • Analysis of mechanisms involved in correct glycosylation by lectin-resistant cell variants • Fractionation of cell populations • Modulation of proliferation and activation status of cells • Model substratum for study of cell aggregation and adhesion
Medicine • Detection of disease-related alterations of glycan synthesis • Blood group typing and definition of secretor status • Quantification of aberrations of cell surface glycan presentation, e.g. in malignancy • Cell marker for diagnostic purposes incl. infectious agents (viruses, bacteria, fungi, parasites)

Table 1.4. Common applications of plant lectins as tools in basic and medical sciences. Adapted from Rüdiger and Gabius, 2001 [169].

CHAPTER 2

ISOLATION, PURIFICATION AND BIOCHEMICAL CHARACTERIZATION OF TAMARIND KUNITZ TYPE INHIBITOR (TKI)

2.1. Introduction

Leguminous seeds contain several proteins that are important for a numerous of biological processes. Proteinase inhibitors from the leguminous family are the most studied, since this family is known as an exceptional source of the protease inhibitors. Among them, Kunitz-type inhibitors are present in comparatively large quantities in the seeds of the Leguminosae subfamilies, Mimosoideae, Caesalpinoideae and Papilionoideae [142, 143]. The Kunitz type inhibitors of serine proteinases are further subdivided into two families namely bovine pancreatic trypsin inhibitor (BPTI) and soybean trypsin inhibitor (STI). The inhibitors from BPTI family have molecular weight of about 6.5 kDa and three disulfide bridges. STI family inhibitors have molecular mass of about 20 kDa with one or two polypeptide chains and less cysteines [96, 110, 140]. Most of them are single polypeptide chain of about 20 kDa with two disulfide bridges and a single reactive site [91, 107, 142]. The Kunitz STI inhibitors are included as Kunitz-P inhibitors in the I3A family of the MEROPS database (http://merops.sanger.ac.uk) [159].

The plant Kunitz type inhibitors are not restricted to trypsin inhibition only, but also show inhibition of several serine and other proteases. It exhibit inhibitory activity on various proteolytic enzymes such as trypsin, chymotrypsin, and other similar enzymes involved in digestive processes, blood coagulation enzymes for example plasma kallikrein, factors XIIa, XIa, and Xa, enzymes of the fibrinolytic system such as plasmin and tissue plasminogen activator and enzymes involved in inflammatory processes like elastase and catepsin G [114]. Some of the pharmacological action of these proteinase inhibitors has been studied as anticoagulant, anti-inflammatory, antiproliferative, antimetastatic agents, anti-invasive property on cancer cells and Anti-HIV-1-RT activity [43, 52, 81, 163]. Due to their multifunctional properties, they have attracted the attention of researchers. These multifunctional plant Kunitz type inhibitors are crucial tools to gain knowledge of the basic principles of protein interactions and could be important for human being for control of many diseases by using them to discover highly potential drugs. So, it is necessary to search for novel Kunitz inhibitors and explore it important functions such as anticoagulant, anticancer, antiviral and anti-inflammatory properties.

In the search of such multifunctional inhibitors which are abundant in leguminous seeds, we have isolated a Kunitz type inhibitor from the seeds a medicinal plant, *Tamarindus indica* of Leguminosae family. *Tamarindus indica* L., usually known as Tamarind tree is one of the most

important multipurpose tropical fruit tree species in the Indian subcontinent. It is traditionally used in abdominal pain, diarrhea and dysentery, helminthes infections, wound healing, malaria and fever, constipation, inflammation, cell cytotoxicity, gonorrhea, and eye diseases. It also reported to possess antidiabetic, antimicrobial, antivenomic, antioxidant, antimalarial, hepatoprotective, antiasthmatic, laxative and anti-hyperlipidemic activity [14].

A protease inhibitor belonging to Kunitz type serine protease inhibitors family isolated here from tamarind seeds has been named "TKI" and this TKI has been shown previously to have trypsin inhibitory and weak elastase inhibitory activity but is inactive towards chymotrypsin and cysteine proteinases (papain and bromelain) [5, 157]. This Kunitz type inhibitor is also shown to possesses *in vitro* and *in vivo* potential bio-insecticidal activity against different insect pests from Lepidoptera, Coleoptera, and Diptera orders [5]. In the present chapter, we have isolated, purified and biochemically characterized TKI for its anticoagulant and anticancer activity. Furthermore, which specific protease is targeted by TKI through which it shows anticoagulant activity has been carried out. In this study, we have explored the pharmaceutical role of TKI and also suggested that TKI act as a multifunctional inhibitor.

2.2. Materials and Methods

2.2.1. Materials

Tamarind seeds were obtained from a tamarind tree growing locally. Affi-Gel Blue matrix, MacroPrep DEAE and MacroPrep CM (carboxymethyl) support, Molecular weight standards, protein assay kit (concentration estimation), were obtained from Biorad (Bio-Rad Laboratories, Hercules, CA, USA). Gel filteration column Superdex-200 was obtained from GE Healthcare. Centricon and Centriprep were purchased from Amicon, Millipore (Beverly, MA). Dialysis membrane was obtained from Pierce. Factor Xa from Thermo Scientific, USA. Other chemicals were purchased from Himedia chemicals, India. N-benzoyl-arginine-p-nitroanilide–HCl, porcine trypsin and factor Xa chromogenic substrate were obtained from Sigma-Aldrich Pvt. Ltd. Human breast, prostate and liver cancer cell lines, MCF-7, PC-3 and HepG2 were obtained from National Center for Cell Science (NCCS), Pune, India. The cell culture reagents were from GIBCO (Invitrogen, USA).

2.2.2. Methods

2.2.2.1. Purification of TKI

Tamarind seeds were collected from a tamarind tree growing locally and seed coat was removed. The seed kernels were soaked in 100 mM Tris buffer pH 7.4 (buffer A) overnight. Then seed kernels were crushed in grinder and slurry were incubated with stirring at 4 °C. The crude extract was collected by centrifugation at 12000 rpm. The pellet was discarded and supernatant were again centrifuge to remove fats. The clear crude extract was obtained which is free from any particles. The crude extract then loaded onto the Affi-Gel Blue matrix column which had been pre-equilibrated with buffer A. After loading the sample, 10 column volumes (CV) of buffer A were used for washing and the bound proteins were eluted with a step gradient of NaCl from 0–1.5 M (0.1, 0.2, 0.3, 0.5, 1.0 and 1.5 M) in buffer A. The fractions eluted with 0.5 M NaCl containing trypsin inhibitory activity were pooled, concentrated using Amicon Ultra 15 (10 kDa cutoff, Millipore) and dialyzed against buffer A. The dialyzed sample was loaded onto a weak cation-exchange matrix [MacroPrep CM (carboxymethyl) support, Bio-Rad] pre-equilibrated with buffer A. The CM flow through containing trypsin inhibitory activity was pooled and loaded onto a weak anion-exchange matrix [MacroPrep DEAE (diethylaminoethyl), Bio-Rad] pre-equilibrated with buffer A. Trypsin inhibitory activity was detected in the DEAE flow through, which was collected and concentrated to about 10 mg/ml. The homogeneity and the molecular weight of the concentrated TKI was determined by 15% SDS–PAGE stained with Coomassie Brilliant Blue.

2.2.2.2. SDS-PAGE analysis

Sodium dodecyl sulfate-polyacrylamide gel (15%) electrophoresis (SDS-PAGE) under both reducing and non reducing conditions was done as described by Laemmli [103]. Relative molecular weight was determined by performing SDS-PAGE of protein with molecular weight standards under reducing condition calibrated on Tris-HCl gel. The molecular weight standards used were myosin (209 kDa), β-galactosidase (124 kDa), bovine serum albumin (80 kDa), ovalbumin (49 kDa), carbonic anhydrase (34.0 kDa), trypsin inhibitor (28 kDa), lysozyme (20.6 kDa) and aprotinin (7.1 kDa). The proteins were detected by staining the gel with 0.1% Coomassie brilliant blue R-250. Molecular weight of TKI was determined by comparing the distance travelled by protein compared to the molecular weight markers.

2.2.2.3. Protein estimation

Protein concentration in crude extract and fractionated protein samples were estimated by Biorad (Bradford) assay using bovine serum albumin (Sigma) as standard. TKI was concentrated and concentration was checked by Biorad (Bradford) assay.

2.2.2.4. The N-terminal sequencing

Purified TKI was subjected to SDS-PAGE (15%) and electroblotted onto a PVDF membrane (Immobilone-P SQ Millipore) in 100 mM Caps-buffer, pH 11. The N-terminal amino acid sequence analysis was performed by Edman degradation on a Shimadzu automated protein sequencer (PPSQ-20) at the Biophysics Department, All India Institute of Medical Sciences, New Delhi, India.

2.2.2.5. Biochemical characterization of TKI

2.2.2.5.1. Trypsin inhibitory activity

The trypsin-inhibitory activity was determined by measuring the residual enzymatic activity towards the substrate BAPNA–HCl (N-benzoyl-arginine-p-nitroanilide–HCl) predissolved in 20% dimethylsulfoxide (DMSO) [51]. The total reaction mixture volume was 1.5 ml. 20 µl trypsin at 1 mg/ml in 50 mM Tris–HCl pH 7.5 containing 20 mM $CaCl_2$ was incubated with 20 µl protein sample in 160 µl 50 mM Tris–HCl pH 7.5 for 10 min at 310 K. The reaction was initiated by the addition of 1 ml 1.5 mM BAPNA solution. The reaction mixture was incubated at 310 K and 0.3 ml 30% acetic acid solution was added after 10 min to terminate the reaction. The enzymatic hydrolysis of the BAPNA was determined spectrophotometrically by measuring the release of p-nitroaniline at 410 nm.

2.2.2.5.2. Anticoagulant activity

Activated partial thromboplastin time (APTT) [73] and prothrombin time (PT) [154] assays were performed using human normal plasma and commercially available kits (Liquiplastin for PT and Liquicelin - E for APTT, Tulip Diagnostics Pvt. Ltd., India) in presence and absence of TKI. For in vitro APTT assay, human normal plasma (100 µl), kit reagent (100 µl) and TKI (2.5–10 µM, previously diluted in 100 µl of 0.1 M Tris–HCl buffer pH 8.0) were mixed and incubated for 3 min at 37 °C. To initiate the coagulation process, 100 µl of 0.025 M $CaCl_2$ was added, and the time for clot formation was recorded. To measure in vitro PT assay,

human normal plasma (100 µl) was incubated with different concentrations of TKI (2.5– 10 µM previously diluted in 100 µl of 0.1 M Tris–HCl buffer at pH 8.0) for 2 min at 37 °C. Then, 100 µl of the kit reagent was added and the clotting time was measured. Both the experiment was done in triplicates and average value was taken.

2.2.2.5.3. Factor Xa inhibitory assay

Factor Xa inhibitory assay was performed as described previously with modification [13, 206]. 0.25 µM human factor Xa (Pierce, Thermo scientific, USA) was incubated with different concentration of TKI (1-60 µM) in 50 mM Tris buffer pH 8.0 and 300 mM NaCl at 37 °C upto 30 min on waterbath. After incubation, the chromogenic substrate CH3OCO-D-CHA-Gly-Arg-pNA-AcOH (Sigma-Aldrich, F3301) was added and incubated for 3 min at 37 °C on water bath. The reaction was stopped by adding 2% citric acid. The color formation by the liberated chromophoric group pNA was measured spectrophotometrically at 405 nm.

2.2.2.5.4. Anticancer activity

Cytotoxicity assays- MTT (3-(4, 5-dimethyl-2-thiazolyl)2,5diphenyl-2H-tetrazolium bromide) assay was carried out as described previously [131]. In brief, 5 X 103 cells in 200 µl of medium were seeded in 96-well plates (Griener, Germany). After cell attachment the medium was removed and the cells were treated with TKI (0, 25, 50, 75, 100 and 125 µM) in 10 mM HEPES buffer (pH-7.4) The cultures were assayed after 24 h by the addition of 20 µl of 5 mg/ml MTT and incubating for 4 h at 37 °C. The MTT containing medium was aspirated and 200 µl of DMSO (Himedia, Mumbai, India) and 25 µl of Sorensen glycine buffer (0.1 M glycine and 0.1 M NaCl, pH 10.5) were added to lyse the cells and solubilize the water insoluble formazone. Absorbance of the lysates was determined on a Fluostar optima (BMG Labtech, Germany) microplate reader at 570 nm. The percentage inhibition was calculated as:

$$\frac{\text{Mean OD of vehicle treated cells (negative control)} - \text{mean OD of treated cells X 100}}{\text{Mean OD of vehicle treated cells (negative control)}}$$

2.3. Results

A proteinase inhibitor was isolated and purified to homogeneity from *T. indica* seeds. Three steps of chromatography involving affinity chromatography and ion-exchange chromatography were used for purification of the tamarind seed proteinase inhibitor. A trypsin inhibitory activity assay was performed at each step of purification. The supernatant obtained after centrifugation of the crude extract was applied onto an Affi-Gel Blue column and the bound proteins were eluted with a step gradient of 0–1.5 M NaCl. Trypsin inhibitory activity was present in the fractions eluted with 0.5 M NaCl, but the fractions were heterogeneous when analyzed on SDS–PAGE. Therefore, fractions containing trypsin inhibitory activity were pooled, concentrated and dialyzed against buffer A. The protein sample was then loaded onto a CM column and the CM flow through containing trypsin inhibitory activity was collected (Fig. 2.1). It was then loaded onto a DEAE column and the flow through was collected. The collected DEAE flow through exhibiting trypsin inhibitory activity was concentrated and assessed for homogeneity by SDS–PAGE; a single protein band appeared. The estimated yield of purified protein was about 0.6 mg per gram of seeds. Molecular-weight estimation of tamarind seed proteinase inhibitor was performed using protein molecular-weight markers and reducing SDS–PAGE. The estimated molecular weight of the purified protein was about 21 kDa (Fig. 2.2; [5, 157]). The first 15 amino-acid residues of the purified protein were determined by automated N-terminal amino-acid sequencing and the obtained sequence was identical to that of the recently reported proteinase inhibitor from tamarind belonging to the Kunitz inhibitor family [157].

The N-terminal amino acids sequence of the first 15 of TKI obtained was Asp-Thr-Val-His-Asp-Thr-Asp-Gly-Lys-Pro-Val-Leu-Asn-Asn-Ala. The obtained amino acid sequence was searched against the NCBI BLAST short sequence search protein data base. The 15 amino acid residue did not show any conserved domain in this stretch of amino acid. The sequence gave 79% identity with Factor Xa inhibitor (BuXI) from *Bauhinia ungulata* followed by substantial identity with Trypsin inhibitor (BvTI) from *Bauhinia variegata* (Table 2.1).

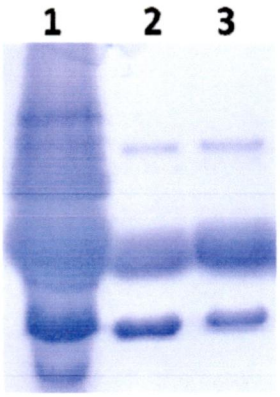


Figure 2.1. Elution profile of TKI on SDS-PAGE. Lane 1: crude extract; Lane 2: Affi-gel blue (0.5 M elute); Lane 3: CM flow through

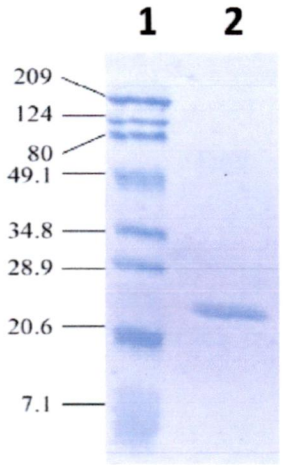


Figure 2.2. SDS-PAGE analysis of molecular weight of TKI from *Tamarindus indica.* Lane 1- molecular weight marker; Lane 2- pure TKI

TKI	DTVHDTDGKPVLNNA	
Name of protein	**Sequence**	**Identity**
Factor Xa inhibitor BuXI (*Bauhinia ungulata*)	1 DTVHDTDGKPVLNN 14 D V DTDGKPV NN 1 DIVLDTDGKPV-NN 13	79%
Trypsin inhibitor BvTI (*Bauhinia variegata*)	1 DTVHDTDGKPVLNN 14 DT DTDG V NN 1 DTLLDTDGEVVRNN 14	64%
Trypsin inhibitor CTI (*Copeifera langsdorffii*)	5 DTDGKPVLNN 14 DTDGKP+ N+ 4 DTDGKPIEND 13	70%

Table 2.1. Sequence homology study of the N-terminal amino acid sequence of TKI (residues 1-15) using NCBI BLAST showing sequence identity with Kunitz type inhibitors of different species.

As TKI has maximum sequence identity with factor Xa inhibitor BuXI, the other functions of TKI (anticoagulant and anticancer activity) along with trypsin inhibition, were tested. Incubation of increasing concentration of TKI with fresh human plasma shows significant prolongation of coagulation time of partial activated thromboplastin time (APTT) and prothrombin time (PT). The TKI (10 μM) extended the APTT of normal plasma 2.6 folds and PT 5 folds (Fig. 2.3). Prolongation of blood coagulation time by TKI and highest sequence identity with factor Xa inhibitor pointed out the role of TKI in inhibition of factor Xa. So to check the specific activity against it, we performed factor Xa chromogenic assay. Factor Xa acts on the chromogenic substrate, CH3OCO-D-CHA-Gly-Arg-pNA-AcOH and cleaves after Arg liberating pNA which forms yellow color. When the increasing amount of TKI was incubated with the factor Xa, decrease in color formation of liberated pNA was detected at 405 nM. This suggests that TKI was able to inhibit the activity of factor Xa. Residual enzyme activity is plotted against increasing concentration of TKI represented in Fig 2.4. The TKI was also tested for its cytotoxic activity in selected human cancer cell line of breast (MCF-7), prostate (PC-3) and liver (Hep G2)

by using MTT assay. The cells were incubated with various concentrations (0-125 μM) of TKI for 24 h on the exponentially growing cells. The result showed that TKI caused dose dependent inhibition of cell viability of cancer cells. Among these cells MCF-7 and HepG2 were more sensitive whereas PC-3 was least sensitive. After 24 h of treatment TKI caused about 14 % and 13 % reduction in cell viability of breast (MCF-7) and liver (HepG2) cancer cells respectively at 125 μM concentration whereas there was only 6% reduction in cell viability of prostate cancer (PC-3) cells at the same dose of 125 μM (Fig. 2.5).

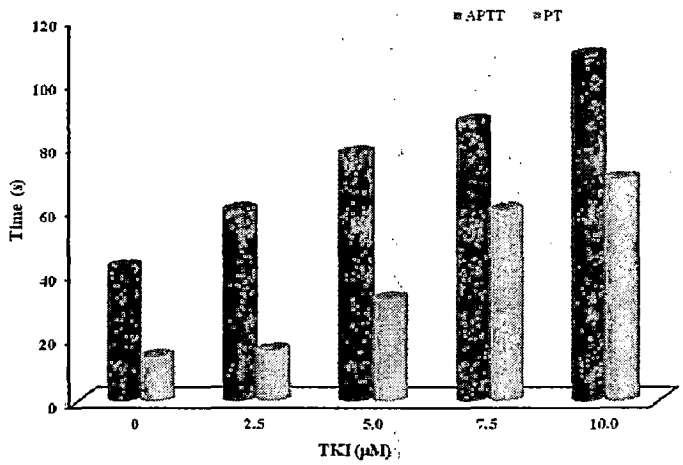


Figure 2.3. Anticoagulation activity of TKI. For blood coagulation time measurement APTT (black) and PT (grey) were determined by a standard procedure. All incubations were performed in the presence of TKI (2.5–10.0 μM) to study its inhibitory activity on blood clotting. 0.1 M Tris–HCl buffer pH 8.0 was used in the control experiment.

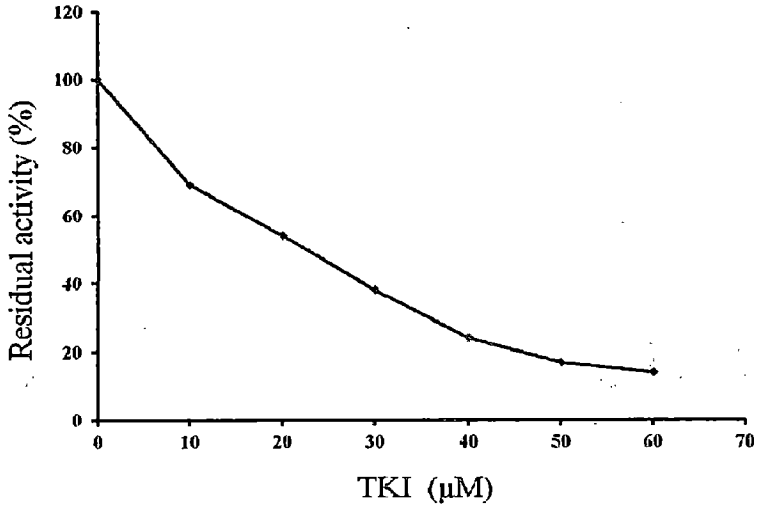


Figure 2.4. Factor Xa inhibition in the dose dependent manner of TKI. Residual activity of factor xa is shown in percentage.

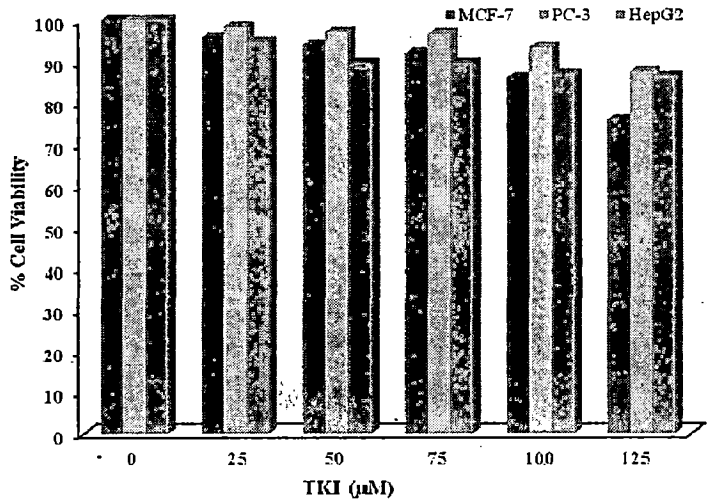


Figure 2.5. Effect of TKI on viability of cancer cells MCF-7 (breast), PC-3 (prostate) and Hep G2 (liver) measured by the MTT assay. Cells were treated with TKI (0-125 μM) for 24 h. The experiment was done in triplicates and average values are taken.

2.4. Discussion

The Leguminaceae family plant, *Tamarindus indica* is a well known for its many medicinal and industrial applications. In this chapter, isolation, purification and biochemical characterization of a Kuntiz type inhbitor (TKI) from the seed kernel of *Tamarindus indica* has been described. The TKI was purified by a three step procedure. The first step involved usage of an affinity matrix affi-gel blue gel followed by anion CM and cation DEAE column matrix at pH 7.4. The approximate molecular mass of the TKI was determined to be 21 kDa from SDS-PAGE analysis. The protein yield was 0.6 mg per gram of seeds and consisted of a homogenous mixture. The trypsin inhibitory activity and N-terminal amino acid sequence analysis revealed the isolated protein is a protease inhibitor belonging to Kunitz type inhibitor family. The N-terminal amino acid sequence of TKI showed 79% homology with factor Xa inhibitor (BuXI) from *Bauhinia ungulata* and 64% with Trypsin inhibitor (BvTI) from *Bauhinia variegata.*

The plant Kunitz type inhibitors are not restricted to trypsin inhibition only and have been observed in some members of Kunitz type inhibitors that it shows multifunctional activities [114]. To explore the fuctions of TKI, we performed anticoagulant, factor Xa inhibitory and anticancer activity. TKI shows increase in clotting time of human normal plasma in dose dependent manner. N-terminal sequence of TKI showed highest similarity with factor Xa inhibitor and prolongation of blood clotting time indicate the role of TKI in factor Xa inhibition. TKI also inhibit factor Xa in a dose dependent manner. To discover the anticancer activity of TKI, MTT assay was performed. 125 μM TKI showed 14 % and 13 % reduction in cell viability of breast (MCF-7) and liver (HepG2) cancer cells respectively.

To sum up, TKI is a single polypeptide 21 kDa protein belonging to Kunitz type inhibitor isolated from the kernel of tamarind seed. TKI shows factor Xa inhibition along with trypsin and act as anticoagulant. It further possesses anticancer activity. This Kunitz type inhibitor also shows in vitro and in vivo potential bioinsecticidal activity against different insect pests from Lepidoptera, Coleoptera, and Diptera orders [5]. Previous studies reported that TKI is also a weak inhibitor of porcine pancreatic elastase [5]. Taken together, the purified protein TKI has multifunctional role as it can act as bioinsecticidal by acting on gut proteases of insects, anticoagulant by inhibiting factor Xa, anti-inflammatory by blocking elastase activity and anticancer. So these properties can be exploited in designing a new drug against them and in anticancer therapy. Conclusively, multifunctional role of TKI has pharmaceutical significance

which can be utilized for human welfare such as treating various clotting disorders, certain inflammatory conditions, cardiovascular diseases, cancer and as an anticoagulant in several surgeries. Further, the bio-insecticidal role of TKI can be used to produce transgenic plant for insect control.

CHAPTER 3

cDNA SYNTHESIS, CLONING, EXPRESSION AND PURIFICATION OF TKI

3.1. Introduction

Kunitz type inhibitors which belong to serine protease inhibitors are abundant in plants. Most of these inhibitors are canonical inhibitors which bind tightly to the enzyme's active site in a substrate like manner through its convex reactive site loop resulting in a stable complex unlike that of the weak complexes between enzyme-substrate and enzyme-product which dissociate in a little period of time. Most plant Kunitz inhibitors has a single polypeptide chain of approximately 20 kDa with two disulfide bonds and a single reactive site. Many exceptions have been discovered recently which appropriately do not follow these characteristics of Kunitz inhibitor family [142]. Kunitz type inhibitors also play several other roles such as bioinsecticidal, anticoagulant, anticancer, anti-inflammatory, antiviral, and antifungal [5, 52, 81, 114].

We have isolated, purified and biochemically characterized a Kunitz type inhibitor, TKI. To attain information about its primary structure, cysteine residue content and its reactive site residues which can important for understanding protease-protease inhibitor interactions, we described here cDNA synthesis of TKI. The nucleotide sequence that was obtained from sequencing of cDNA was translated to obtain the amino acid sequence of TKI. Moreover, the sequence analysis was performed using bioinformatics tools such as BLAST, Clustal W, EsPript, ExPASy. Phylogenetic tree was constructed using MEGA version 5 program to evaluate evolutionary relationship among other inhibitors. Additionally, the TKI gene was cloned into bacterial expression system to further gain knowledge about the functional property of recombinant form as well as site directed mutagenesis studies in future. The recombinant protein was checked for its trypsin inhibitory activity. The complete amino acids sequence of TKI can be further utilized for its structure determination.

3.2. Materials and Methods

3.2.1. Materials

Plasmid isolation kit, PCR purification kit, gel extraction kit and Ni-NTA beads were obtained from Qiagen. M-MuLV reverse transcriptase and Taq polymerase were obtained from Bangalore GeNei (Bangalore). pGEM-T easy vector from Promega, Madison, WI, USA. All restriction enzymes and DNA ligase were from New England Biolabs Inc. All other chemicals were obtained from Himedia, India. Sequencing of TKI gene was done at Ocimum Biosolutions, Hyderabad and Vimta Labs, Hyderabad, India.

3.2.2. Methods

3.2.2.1 cDNA cloning

3.2.2.1.1. Isolation of total RNA from tamarind seeds

In order to obtain the complete amino acid sequence of TKI, three months old *Tamarindus indica* seeds were collected locally. Total RNA was isolated from the seeds as described previously [192] with slight modifications such as use of 4% polyvinylpyrrolidone (PVP) in the extraction buffer to remove the phenolic compounds and polysaccharide, which yield high quality RNA. Also to significantly inhibit RNase activity during the extraction and remove the other insoluble materials, extraction buffer with slight higher pH (pH 8.5) was used [121]. Quality of RNA sample was checked on 1.2% agarose gel in 0.5 X TBE buffer (50 mM TRIS, 50 mM boric acid, 0.2 mM EDTA, pH 8.5) at 35 V up to 60 min, in which little DNA contamination was observed. To remove the DNA contamination, 9 µl RNA sample was incubated with 0.5 µl DNase I in 0.5 µl 10X DNase I buffer (100 mM Tris, 25 mM $MgCl_2$, 5 mM $CaCl_2$ Water) at 37 °C for 30 min. DNase I was inactivated by heating the reaction sample at 70 °C for 5 min followed by rapid chilling on ice. The purity of RNA was checked on 1.2 % agarose gel.

3.2.2.1.2. RT-PCR and cDNA synthesis

Four micrograms of total RNA was hybridized with 50 pmol of oligo$(dT)_{18}$ adaptor primer (5'CCAGTGAGCAGAGTGACGACTCGAGCTCAAGCTTTTTTTTTTTTTTTT-3') and RNAs were reverse transcribed in 20 µl reaction volume containing 50 mM Tris-HCl pH 8.5, 8 mM $MgCl_2$, 30 mM KCl, 1 mM DTT, 1 mM of dNTPs, 20 U of RNasin and 100 U of M-MuLV reverse transcriptase. The RT cycle comprised of incubation of the reaction mix at 25 °C for 5 min followed by a cycle of 37 °C for 60 min in an Eppendorf thermal cycler. The RT reaction was terminated with a cycle of 5 min at 95 °C. The product of RT reaction (3 µl) was utilized as a template for the synthesis of cDNA by 30 cycles of polymerase chain reaction (PCR). Forward degenerate primer (F1: 5'-GAYACHGTNCAYGAYACHGAYGG-3') was designed on the basis of previously reported N-terminal sequence of TKI [157] and the oligo$(dT)_{18}$ adaptor primer was used as the reverse primer (R1). cDNA was prepared in 50 µl total volume containing 5 µl of 10X reaction buffer, 2 µl of 20 mM dNTP, 1 mM primers (F1 and R1) and 3U of Taq polymerase. PCR amplifications was performed in a thermocycler (Eppendorf AG Hamburg, Germany) with an initial denaturing step of 94 °C for 5 min, followed

by 30 amplification cycles of 94 °C for 60 s, 51 °C for 60 s, 72 °C for 60 s and a final extension cycle at 72 °C for 5 min.

The PCR product was subjected to electrophoresis on 1% agarose gel with DNA marker and gel purified using QIAquick gel extraction kit (Qiagen, Germany). Then amplified cDNA of TKI gene was cloned into pGEM-T easy vector according to the instructions provided (Promega, Madison, WI, USA). DH5α competent cells were transformed using ligated product by the heat shock method and plated on LB agar plates containing ampicillin, IPTG and X-gal. Recombinant bacterial colonies were identified by blue/white screening. Obtained white colonies were picked and 5 ml LB containing 80μg/ml ampicillin was inculcated. Bacterial cultures were grown overnight at 37 °C and plasmids were isolated using a Qiaprep miniprep kit (Qiagen, Germany). TKI gene insert was confirmed in the isolated plasmids by PCR and double restriction digestion using *Nde*I and *Nco*I. Obtained recombinant plasmid containing TKI gene insert (pGEMT-TKI) were sequenced in both the directions using M13 forward and reverse primers on an automated sequencer by DNA sequencing services of Ocimum Biosolutions, Hyderabad & Vimta Labs, Hyderabad, India. The obtained nucleotide sequence of TKI cDNA has been submitted to GenBank database under the accession number HQ385502.

3.2.2.1.3. Analysis of TKI gene

The ORF finder tool from NCBI was used to deduce the primary amino acid sequence of TKI from the TKI gene sequence. Analysis of TKI sequence was done using ProtParam in ExPASy server [57]. The similar sequences available in database were identified by the programme BLAST [2]. Multiple sequence alignment was performed using CLUSTAL W [196] with default parameters. The prediction of the Kunitz signature and probable motifs involved in post translational modifications were done by using Motif scan server [78]. Phylogenetic tree was constructed using MEGA version 5 program using the Neighbor-Joining method. The evolutionary distances were computed using the Poisson correction method [193]. A figure of amino acids sequence alignment was prepared by ESPRIPT programme [61].

3.2.2.2. Cloning of TKI

The TKI cDNA was used as the template for the polymerase chain reaction amplification. Two oligonucleotide primers containing *Nde*I and *Xho*I restriction endonuclease sites (Table 3.1) were designed to anneal to the 5' and 3' ends of the gene respectively for full length TKI gene

amplification. The *Nde*I and *Xho*I restriction sites introduced in the forward and reverse primers provided cloning of TKI gene into the corresponding sites of pET-28cTEV vector. The gene encoding for TKI was amplified by using these primers in a PCR reaction and amplified DNA fragment was purified using PCR purification kit according to manufacturer's instructions to remove unincorporated nucleotides and primers. The plasmid pET-28cTEV was isolated by using plasmid miniprep kit from the culture of DH5α cells containing the plasmid overnight grown at 37 °C. Purified PCR fragment and pET-28cTEV vector were restriction digested with *Nde*I and *Xho*I. The digested DNA fragments were run on 1% agarose gel and purified by using DNA gel extraction kit. Then both the products were ligated at 15 °C using T4 DNA ligase. E. coli DH5α cells were used for transforming the ligated product by heat shock method [82]. Competent cells were prepared by using chemical method ($CaCl_2$ treatment). The transformed cells were plated onto Luria-Bertani (LB) agar plates containing 50 µg of kanamycin/ml. The colonies obtained were picked, grown overnight and plasmids were isolated using MiniPrep plasmid isolation method. For screening purpose isolated plasmids were restriction enzyme digested for the presence of TKI gene with *Nde*I and *Xho*I. Polymerase chain reaction was also performed for further screening with same primers. The cloned plasmid was used for further experiment. This cloning was done in such a way that the cloned plasmid should contain the nucleotide sequence encoding 6x-His residues at the amino-terminus of the TKI with TEV (Tobacco etch virus) protease cleavable site to facilitate purification with Ni-NTA column.

Primers	Primer sequence	% GC	Length	Tm (°C)
FWD	5'-TCAGCTGCATATGGATTATACTGTGCATGA-3'	40	30	59.5
REV	5'-TAGTGCTCGAGTCAAACAATACTCCATTCC-3'	43.3	30	59.7

Table 3.1. Primer sequence for TKI gene amplification. The highlighted bases in the primer sequences indicate *Nde*I and *Xho*I digestion sites in FWD and REV primers, respectively.

3.2.2.3. Expression of rTKI in *E. coli* BL21 (DE3)

Plasmid pETHis6-TKI containing TKI gene was used for transforming *E. coli* expression strain BL21 (DE3). The expression of His-tagged TKI protein was tested by growing 5 ml bacterial cultures in Luria broth (LB) supplemented with appropriate antibiotics (Kanamycin) at 37°C for overnight. 100 µl of the overnight culture was further inoculated for fresh culture. Cells were grown at 37 °C to an optical density of 0.6 at 600 nm (OD_{600}) and protein expression was induced with 0.4 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). Cultures were induced for ~4 h at 37 °C and 18 °C for 16 h. Cells were harvested by centrifugation at 5,000 x g at 4 °C and cell pellets were re-suspended in 500 µl of 20 mM Tris pH 7.4. Cells were lysed by sonication on ice (30% amplitude and a pulse of 20 s). The cell lysates were centrifuged at 15,000 x g at 4 °C for 2 min to separate soluble and insoluble fractions. Samples were analyzed for expression and solubility on 15% SDS-PAGE gels and visualized by coomassie staining. Trypsin inhibitory activity was checked.

3.2.2.3.1. Trypsin inhibition assay

The trypsin inhibitory activity of recombinant TKI was determined by measuring the residual enzymatic activity towards the substrate BAPNA–HCl (N-benzoyl-arginine-p-nitroanilide–HCl) predissolved in 20% dimethylsulfoxide (DMSO; [51]). The total reaction mixture volume was 1.5 ml. 20 µl trypsin at 1 mg/ml in 50 mM Tris–HCl pH 7.5 containing 20 mM $CaCl_2$ was incubated with 20 µl protein sample in 160 µl 50 mM Tris–HCl pH 7.5 for 10 min at 310 K. The reaction was initiated by the addition of 1 ml 1.5 mM BAPNA solution. The reaction mixture was incubated at 310 K and 0.3 ml 30% acetic acid solution was added after 10 min to terminate the reaction. The enzymatic hydrolysis of the BAPNA was determined spectrophotometrically by measuring the release of p-nitroaniline at 410 nm.

3.2.2.3.2. Expression of rTKI in engineered *E. coli* K12 cells (shuffle T7)

Recombinant TKI expressed in BL21 cells didn't show any trypsin inhibitory activity. So we postulate that this might be due to improper folding and disulphide bond formation. To overcome this problem we used chemically competent *E. coli* K12 cells (shuffle T7 from New England biolabs inc.) which are engineered to form disulfide bonded proteins in the cytoplasm. These cells are suitable for T7 promoter driven protein expression. Protein expression was induced by the addition of 0.4 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) and cells were

cultured at the 37 °C for 4 h and 18 °C for an additional 16 h. Samples were analyzed for expression and solubility on 15% SDS-PAGE gels. Supernatant was checked for its trypsin inhibitory activity. For large scale production, we expressed TKI in 1 L Luria–Bertani broth (LB) culture of shuffle T7 cells which was induced with 0.4 mM isopropyl-β-thiogalactopyranoside (IPTG) at 18 °C for an additional 16 h and harvested the cells by centrifugation (8000g, 10 min, 279 K).

3.2.2.4. Purification of rTKI

For purification, the cell pellet was washed two times with purification buffer (PB buffer) comprising of 20 mM Tris buffer pH 7.4, 5% glycerol, 150 mM NaCl and resuspended in 40 ml PB buffer containing 35 µl DNase I (1 U/µl). Cells were disrupted at 138 MPa using a one-shot cell disrupter (Constant Systems Ltd, Daventry, England). The lysate was clarified by centrifugation at 14 000g for 45 min at 277 K and the supernatant was collected. The supernatant was added to 3 ml Ni–NTA beads (Qiagen, USA) pre-equilibrated with PB buffer containing 5 mM imidazole and gently agitated at 277 K for 1 h. The supernatant/bead mixture was loaded onto a gravity-flow column and the flow through was collected. Column washing was subsequently performed with five column volumes of washing buffers WB1, WB2 and WB3 (20 mM Tris pH 7.4, 5% glycerol, 300 mM NaCl and 20, 40 or 60 mM imidazole, respectively). The bound protein was eluted with 50 ml elution buffer (20 mM Tris buffer pH 7.4, 5% glycerol, 150 mM NaCl and 250 mM imidazole) and collected in 5 ml fractions. The collected fractions were analyzed by SDS–PAGE and those showing a single band corresponding to TKI were pooled and dialyzed against 2 L dialysis buffer (20 mM Tris buffer pH 7.4, 5% glycerol, 50 mM NaCl) at 277 K two times for 6 h each to remove imidazole. The trypsin inhibitory activity of purified recombinant TKI was measured.

3.3. Results

3.3.1. Total RNA isolation, cDNA cloning and sequence analysis

RNA contaminated with low amount of DNA was isolated with high purity after the DNase I treatment (Fig. 3.1). Amplification was obtained by using forward degenerate primer designed by N terminal sequencing of TKI. Approximately a 700 bp fragment was obtained, which includes the 3' UTR region with poly (A+) tail (Fig. 3.2). The complete nucleotide and derived amino acid sequences of TKI without signal sequence have been determined and deposited in the NCBI gene databank under accession no. HQ385502. A 555 bp ORF obtained after sequencing of the TKI gene coded for a polypeptide of 185 amino acid residues (Fig. 3.3). The mature TKI consisted of 185 amino acid residues polypeptide with a calculated molecular mass of 20575 Da. TKI showed theoretical pI: 5.53 and ext. coefficient:19940. The amino acid composition of TKI is summarized in Table 3.2.

TKI showed significant similarity to the reported Kunitz-type inhibitors in homology search of database using NCBI BLAST. These results demonstrated that TKI belongs to soybean Kunitz family inhibitors. The primary amino acid sequence of TKI does not show any conserved motif related to Kunitz-type (STI) trypsin inhibitor despite of having striking structure similarity with later. This is due to insertion of Asn15 in between the Kunitz signature sequence. This has also been observed in *Bauhinia variegata* trypsin inhibitor (BvTI). A protein sequence BLAST with non-redundant database showed that TKI shares highest sequence identity 45%, 42%, 39% and 33% with factor Xa inhibitor from *Bauhinia ungulata* (BuXI), *Bauhinia variegata* trypsin inhibitor (BvTI), trypsin inhibitor from *Erythriana caffra* and soybean trypsin inhitor (STI) respectively. The chain length of these proteins ranges from 172 to 185 residues. Four cysteins are linked by two disulphide bonds in TKI as in ETI and STI (Cys 42-86, Cys 134-144). Phylogenetic analysis and multiple sequence alignment of the representative Kunitz-type inhibitors from different plants were performed to understand any evolutionary relationship among these inhibitors. In both the analysis, TKI sequence formed close relationship with classical Kunitz family members with closest relation with factor Xa inhibitor from *Bauhinia ungulata* (BuXI) (Fig. 3.4 and 3.5).

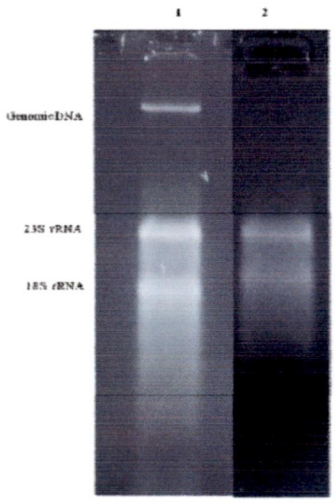


Figure 3.1. Isolation of total RNA from tamarind seed. Lane 1 – 23S and 18S rRNA bands are indicated with genomic DNA contamination, Lane 2 – treatment of total RNA with DNase I.

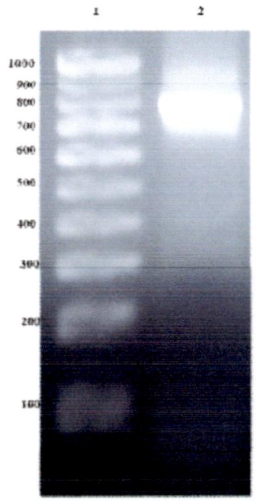


Figure 3.2. Amplification of TKI cDNA. Lane 1 – 100bp DNA ladder, Lane 2 – TKI gene PCR amplification product.

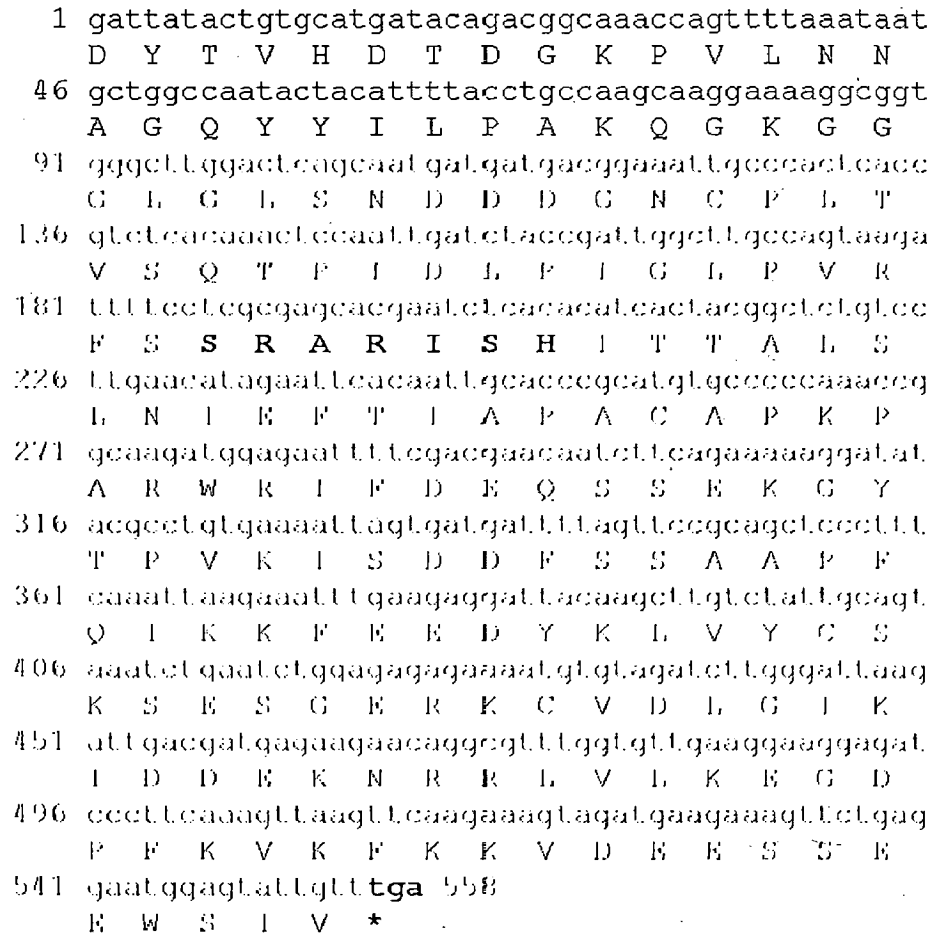


Figure 3.3. Nucleotide and deduced amino acid sequence of TKI. The translated sequence of mature TKI consists of 185 amino acids. The active site residues are highlighted (grey background). The genebank accession numbers of the nucleotide sequences of TKI cDNA is HQ385502.

Amino acids	Total no.	Content (%)
Ala (A)	10	5.4%
Arg (R)	8	4.3%
Asn (N)	6	3.2%
Asp (D)	16	8.6%
Cys (C)	4	2.2%
Gln (Q)	5	2.7%
Glu (E)	13	7.0%
Gly (G)	13	7.0%
His (H)	2	1.1%
Ile (I)	13	7.0%
Leu (L)	13	7.0%
Lys (K)	18	9.7%
Met (M)	0	0.0%
Phe (F)	8	4.3%
Pro (P)	12	6.5%
Ser (S)	17	9.2%
Thr (T)	8	4.3%
Trp (W)	2	1.1%
Tyr (Y)	6	3.2%
Val (V)	11	5.9%

Table 3.2. Amino acid composition of TKI using ExPASy ProtParam server.

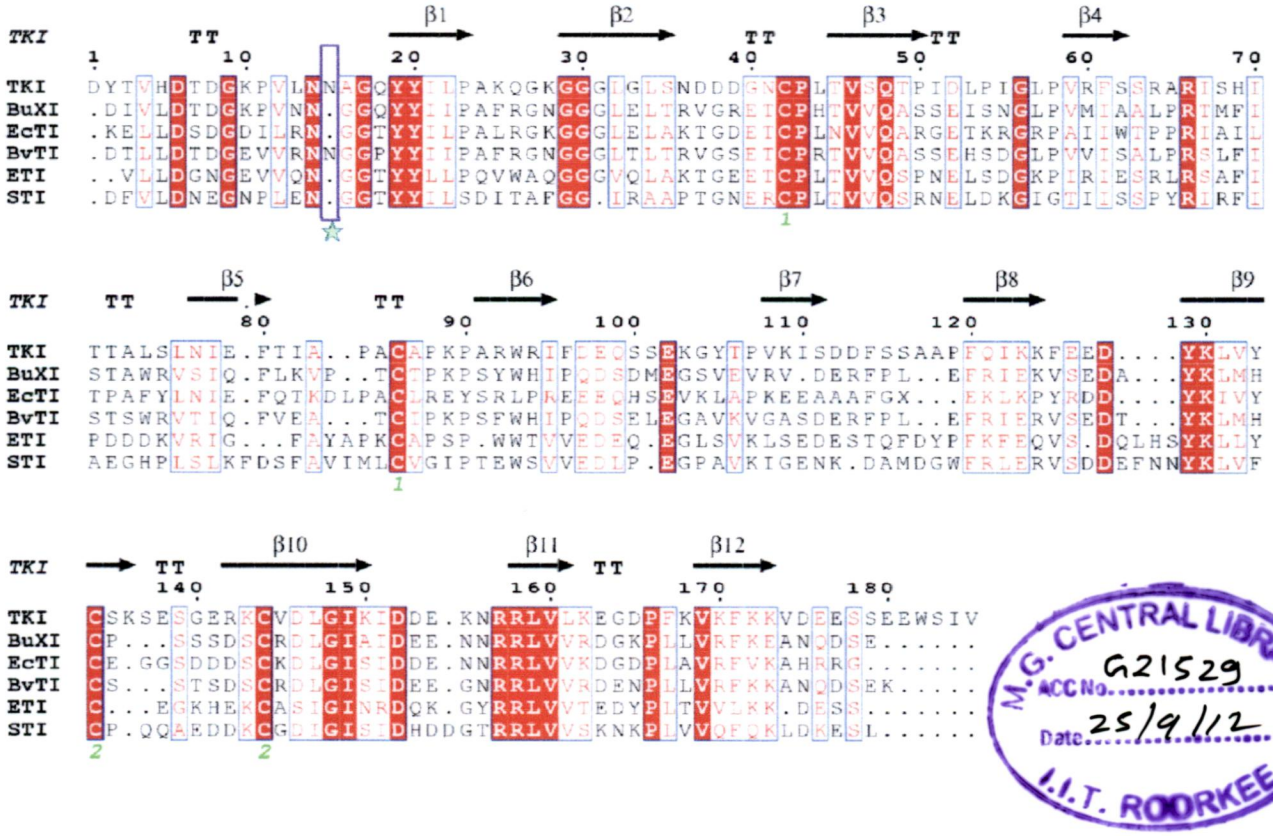


Figure 3.4. Multiple sequence alignment of TKI with other Kunitz STI inhibitors. Arrows are for β sheets and TT are for turns. Absolute conserved residues among all proteins are shown red background. Cysteine residues involved in forming disulfide bonds are represented in green numbering. Insertion of one residue between Kunitz signature sequence shown by brown star. TKI, tamarind trypsin inhibitor; BuXI, factor Xa inhibitor from *Bauhinia ungulata*; EcTI, *Enterolobium contortisiliquum* trypsin inhibitor; BvTI, *Bauhinia variegata* trypsin inhibitor; ETI, *Erythrina caffra* trypsin inhibitor; STI, soyabean Kunitz trypsin inhibitor.

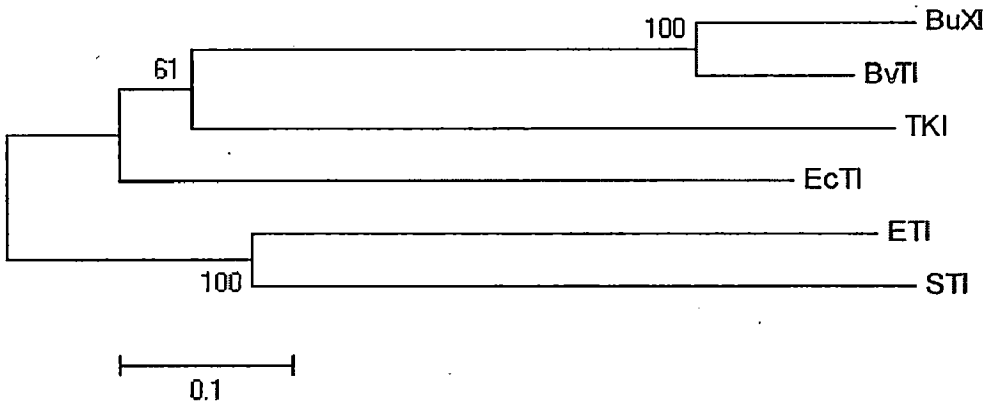


Figure 3.5. Phylogenetic tree of TKI was created using the Neighbor-Joining method in MEGA5 program. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site.

3.3.2. Cloning, expression and purification of TKI

The full length TKI encoding gene was amplified with *Nde*I and *Xho*I digestion site in FWD and REV primers, respectively which showed approximately 550 bp TKI gene (Fig. 3.6). The TKI gene was cloned into pET-28cTEV vector and transformed into DH5α. To confirm TKI insert into pET-28cTEV vector, double digestion of pET-28cTEV vector was done with *Nde*I and *Xho*I that showed TKI insert into pET-28cTEV (Fig. 3.7). It was also further confirmed by amplifying TKI gene using pET-28cTEV vector as a template. For expression of TKI gene, pET-28cTEV-TKI plasmid was transformed into BL21- DE3 competent *Escherichia coli* cells. The rTKI gene was expressed and rTKI protein was soluble but it showed negligible trypsin inhibitory activity which might be due to improper folding and disulfide bonds formed. So, we tried to express TKI in chemically competent *E. coli* K12 cells shuffle T7 which are engineered to form disulfide bonded proteins in the cytoplasm. SDS-PAGE analysis showed that rTKI was expressed at 18 °C as a soluble protein (Fig. 3.8A). Further trypsin inhibitory assay was performed which showed that rTKI has potent trypsin inhibitory activity like native TKI.

Recombinant TKI was purified using affinity chromatography (Ni-NTA) method. The N-terminal His tag protein was bound to Ni–NTA column and eluted with 250 mM imidazole. The SDS-PAGE analysis showed single purified band of rTKI (Fig. 3.8B). The protein was dialyzed to remove excess imidazole and pure rTKI exhibited trypsin inhibitory activity like native TKI.

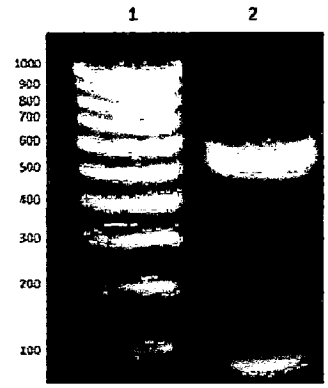


Figure 3.6. Amplification of TKI gene using *Nde*I and *Xho*I digestion site in FWD and REV primers, respectively. Lane 1 -100 bp maker, Lane 2 – TKI gene ~ 550 bp.

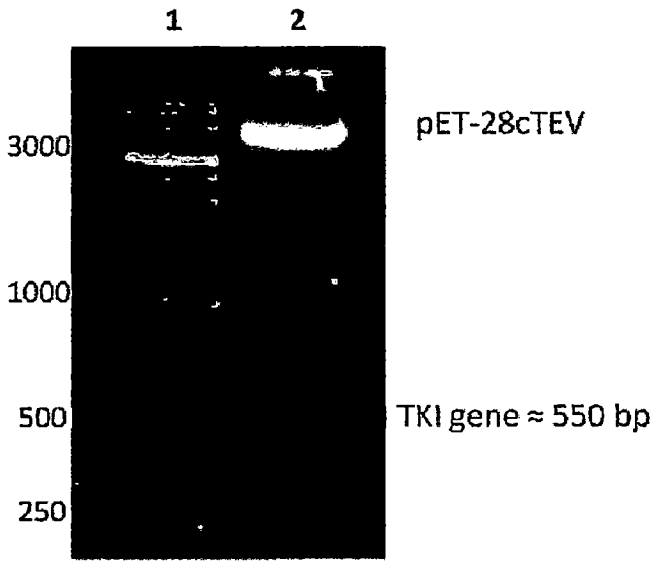


Figure 3.7. Confirmation of TKI gene into pET-28cTEV vector using *Nde*I and *Xho*I restriction enzyme double digestion of recombinant pET-28cTEV plasmid which contains TKI gene insert. Lane 1- broad range DNA marker, Lane2 – RE digseted (*Nde*I and *Xho*I) plasmid showing TKI insert.

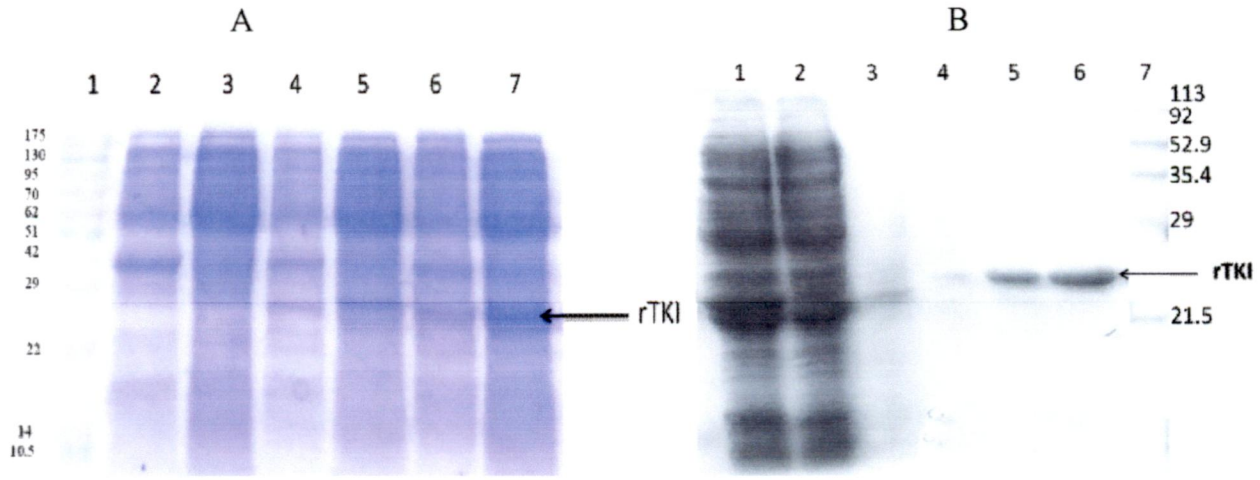


Figure 3.8. A. 15 % SDS-PAGE analysis to check expression and solubility of rTKI. Lane 1- pink plus prestained protein ladder (GeneDireX), Lane 2- uninduced at 37 °C pellet, Lane 3- uninduced at 37 °C sup, Lane 4- induced at 37 °C pellet, Lane 5- induced at 37 °C supernatant, Lane 6- induced at 18 °C pellet, Lane 7- induced at 18 °C supernatant which shows expressed and soluble rTKI. B. 15 % SDS-PAGE analysis of rTKI purification. Lane 1- supernatant, Lane 2- flow through, Lane 3 -4 Wash, Lane 5-6- eluted fraction of rTKI, Lane 7- molecular marker

3.4. Discussion

To deduce amino acid sequence of TKI, total RNA was isolated from tamarind seeds followed by cDNA synthesis. The cDNA was amplified from PCR obtained ~700 bp product which contained the 3' UTR region and poly (A+) tail. A 555 bp ORF obtained after sequencing of the TKI gene coded for a polypeptide of 185 amino acid residues with a calculated molecular mass of 20575 Da. Sequence analysis demonstrated that TKI belongs to Kunitz STI type inhibitors and showed highest sequence identity (45%) with factor Xa inhibitor from *Bauhinia ungulata* (BuXI). The sequence analysis also reveals that TKI has insertion of Asn residue at position 15 within the Kunitz signature sequence and so the Kunitz signature sequence gets distorted. TKI has maintained two characteristics of Kunitz STI family such as having molecular weight ~ 20-22 kDa and conserved two disulphide bonds which was also observed in ETI and STI.

The full length TKI encoding gene was cloned into pET-28cTEV vector and expressed into BL21- DE3 competent *Escherichia coli* cells and showed expression and solubility of TKI protein but have negligible trypsin inhibitory activity which might be due to improper folding and disulfide bonds formed. Formation of correct disulfide bonds and proper folding might be required for TKI to maintain its inhibitory activity, so we expressed TKI in chemically competent *E. coli* K12 cells shuffle T7 which are engineered to form disulfide bonded proteins in the cytoplasm. SDS-PAGE analysis showed that rTKI was expressed as a soluble protein which has potent trypsin inhibitory activity like native TKI. Recombinant TKI was purified using Ni–NTA column. The SDS-PAGE analysis showed single purified band of rTKI and pure rTKI exhibited trypsin inhibitory activity. This result shows that rTKI is a stable protein and able to preserve its functional property and can be used further for biochemical studies.

CHAPTER 4

CRYSTALLIZATION AND STRUCTURE DETERMINATION OF TKI, ITS COMPLEX WITH TRYPSIN AND MOLECULAR DOCKING WITH FACTOR Xa

4.1. Introduction

Detailed information on the structure of plant Kunitz-type inhibitors is useful in understanding the mechanisms underlying their specificity for their cognate proteases and to allow us to investigate region of the protein responsible for its biological activity. Several structures of Kunitz STI type inhibitors have been reported showing that they have β-trefoil fold consisting of 12 anti-parallel β-strands connected by long loops. Their exposed reactive site loop has a characteristic canonical conformation with Arg/Lys at P1 position [41, 96, 120, 144, 158, 187]. Conserved Asn (Asn13 in STI) residue is observed in structures of Kunitz STI inhibitors which maintain the canonical conformation of the reactive site loop [40, 144, 187]. The structure of STI complex with porcine trypsin reveals that P1 residue; Arg63 enters in S1 pocket of trypsin and makes a salt bridge with Asp189 and inhibit it in a substrate like manner. Along with composition and conformation of reactive site loop, other aspects like global architecture which is essential for stability, efficiency and rigidity and important area of contact between two proteins must be considered.

The X-ray crystallographic analysis provides conformational details of protein structure and better understanding of protein-protein interaction at molecular level. To gain the knowledge about TKI structural conformation and its interaction with proteases, we have performed crystallization and structure determination of TKI and its complex with trypsin. Crystals were obtained by sitting drop method in 96-well crystallization plates and the diffraction data was collected at home source. Indexing, integration and scaling of the diffraction data were carried out using HKL2000 program. The phase determination was done by molecular replacement method using the program MOLREP in CCP4i suite. Structure refinement was done using REFMAC5 and model visualization, model building was performed using the molecular graphics program COOT. Structure validation was done by the program PROCHECK which analyse the stereochemistry of polypeptide backbone using the Ramachandran plot. The refined model of TKI was compared with known Kunitz type inhibitors structure and detail interactions of TKI:PPT complex was studied.

To further gain insight into interaction of TKI with factor Xa molecular docking was performed using online protein-protein docking program ClusPro. The detailed interactions between TKI and FXa residues were calculated using PIC server.

4.2. Materials and Methods

4.2.1. Materials

Porcine pancreatic trypsin, MES, HEPES, MPD, PEG 8000, PEG3350 were purchased from Sigma-Aldrich Pvt. Ltd. All other chemicals were purchased from Himedia chemicals, India. Crystallization plates, initial crystallization screen reagent kits and cryoloops were obtained from Hampton Research, USA. Superdex- 200 GL 10/300 column and Gel Filtration Calibration LMW standards kit were obtained from GE Healthcare, USA.

4.2.2. Methods

4.2.2.1. Crystallization of TKI

For crystallization, the purified TKI was used at a concentration of 9 mg/ml in 100 mM Tris–HCl pH 7.4. Crystallization was performed by the hanging-drop vapour-diffusion method in 24-well VDX plates (Hampton Research). The drops were prepared by mixing 4 µl protein solution with the same volume of reservoir solution and were equilibrated against 1000 µl reservoir solution at 293 K. Crystals of tamarind seed proteinase inhibitor were obtained in the presence of 100 mM 2-[4-(2-hydroxyethyl)-1-piperazinyl] ethanesulfonic acid (HEPES) pH 7.0, 100 mM zinc acetate and 10–15% (w/v) polyethylene glycol (PEG) 6000 as precipitant. Afterwards, to obtain better diffracting crystals, initial crystallization experiments were performed via the sitting-drop vapor-diffusion method using Crystal Screen I and II kits and PEG ION screen I and II (Hampton Research, Aliso Viejo, CA). For free TKI sitting drops were prepared by mixing equal volumes (1 µl each) of the protein and reservoir solutions. The better diffracting crystals for free TKI were obtained using 25% (v/v) polyethylene glycol 3350 (PEG 3350) and 0.2 M magnesium formate at pH 7.0.

4.2.2.2. Complex formation of TKI with pancreatic porcine trypsin (PPT) and crystallization

4.2.2.2.1. Complex formation and purification of TKI:PPT

The complex was prepared by mixing 50 µl purified Kunitz-type protease inhibitor at a concentration of 10 mg/ml in a buffer solution containing 100 mM Tris pH 7.4 with 50 µl porcine trypsin at a concentration of 10 mg/ml in a buffer solution containing 100 mM Tris pH 7.4 and 40 mM $CaCl_2$. The protease–inhibitor mixture was then incubated at 277 K for 2 h. For

purification of the protease–inhibitor complex, 100 µl of the mixture was loaded onto a Superdex- 200 GL 10/300 (GE Healthcare) size-exclusion column using a 100 µl sample loop at 0.5 ml/min on an AKTA Purifier FPLC system (GE Healthcare). Elution of the protein was detected by UV absorbance at 280 nm. The size-exclusion column was calibrated with Gel Filtration Calibration LMW standards (GE Healthcare; blue dextran, thyroglobulin, ferritin, aldolase, conalbumin and ovalbumin) for determination of the void volume, construction of the standard curve and estimation of the molecular weight of the complex. The eluted fractions corresponding to the molecular weight of the protease– inhibitor complex were analyzed on 15% reducing SDS–PAGE. The fractions containing the trypsin–TKI complex were pooled and concentrated to 10 mg/ml at 277 K using an Amicon Ultra-4 10 kDa cutoff concentrator (Millipore). The protein concentration was determined with the Bio-Rad protein-assay kit using bovine serum albumin (BSA) as a standard.

4.2.2.2.2. Crystallization of TKI:PPT complex

Crystallization was performed by the sitting-drop method at 293 K using Crystal Screens I and II and Salt Screen (Hampton Research, USA) for initial screening in 96-well crystallization plates (Hampton Research, USA). Drops were prepared by mixing 2 µl protein solution with 2 µl precipitant solution and were equilibrated against 50 µl reservoir solution. Tamarind inhibitor–porcine trypsin complex crystals grew under various conditions, but the diffracting crystals grew in 4 M ammonium acetate and 0.1 M sodium acetate trihydrate pH 4.6. Furthermore, to get higher resolution diffraction data, crystals used for final complex structure determination were obtained using 0.2 M ammonium phosphate dibasic and 20% PEG 3350.

4.2.2.3. Data collection and analysis

Crystals were mounted using 0.1–0.2 mm cryoloops (Hampton Research) and were flash-cooled in a liquid-nitrogen stream at 100 K generated by an Oxford Cryosystem (Oxford, England). Data were collected with a MAR 345 dtb imaging-plate system using Cu Kα radiation generated by a Bruker Microstar-H rotating-anode generator operated at 45 kV and 60 mA and equipped with Helios optics. For free TKI, data were collected as 100 images with a crystal-to-detector distance of 180 mm and 1° oscillation per image. The time of exposure per frame was 5 min. The crystal diffracted to 1.9Å resolution. For TKI:PPT complex crystal, the data were collected as 100 images with a crystal-to-detector distance of 200 mm and 1° oscillation per image. The time of exposure per frame was 10 min. The crystal diffracted to 2.2Å resolution.

Indexing, integration of all the diffraction images and scaling of the diffraction data were carried out in HKL2000 [146] and are summarized in Table 4.1.

4.2.2.4. Structure determination and refinement

The structures of free TKI and its complex with trypsin were solved by molecular replacement method using the program MOLREP in CCP4i suite [7]. The best solution for TKI and complex were obtained with crystal structure of a trypsin inhibitor from *Erythriana caffra* seeds (PDB ID code, 1TIE) [144] and complex structure of soybean (PDB ID code, 1AVW) as a search model respectively [187]. Rigid body refinement was followed by iterative cycles of restrained atomic parameter refinement using REFMAC5 [7, 135]. Further model visualization, refinement, model building and fitting of the electron density map was carried out using the molecular graphics program COOT [49]. The water molecules were located and added in the structure with COOT. The details of data collection, processing and refinement statistics are included in Table 4.1. The final solution of MOLREP yielded two molecules per asymmetric unit for free TKI and one molecule in asymmetric unit for complex. Further, this model was refined as rigid body followed by restraint refinement using REFMAC5 program [135]. The final structures were validated using PROCHECK program [106] as implemented in CCP4.

4.2.2.5. Molecular docking of TKI with factor Xa

As trials of getting crystals of TKI with factor Xa were unsuccessful, to examine the mode of interaction with it, rigid body docking was performed. The fully automated online protein-protein docking program ClusPro [34, 35] was used to model TKI:factor Xa complex with good surface complementarity and low desolvation energies. The PDB file for the factor Xa (PDB code:1FAX) [24] was submitted as a receptor structure, whereas the coordinates of TKI (PDB code:4AN6) as the ligand structure. The server was executed with default parameters. In brief, the top 20,000 complex models were trimmed down to 2000 models based on their desolavation and electrostatic energies. These were then subjected to clustering and ranked according to their cluster size (default value of 9 Å as a radius of clustering). To relieve possible side-chain clashes, these clusters were introduced for short energy minimization. We used the top-ranked complexes from ClusPro for further analysis based on the prior knowledge of active site interactions. Interactions between TKI and FXa residues were calculated using PIC server [198].

4.3. Results

4.3.1. Crystallization of TKI and its complex with trypsin

For free TKI, the better diffraction quality crystals were obtained in 5-8 days in the reservoir solution containing 25% (v/v) polyethylene glycol 3350 (PEG 3350) and 0.2 M magnesium formate at pH 7.0 (Fig 4.1). The crystals belonged to the space group $P2_12_12_1$ and diffracted to 1.9Å resolution in-house. The unit-cell parameters were found to be a = 40.43, b = 60.42, c = 105.53Å, with two molecule per asymmetric unit.

An attempt was made to prepare a protein–protein complex between porcine trypsin and TKI. For complex formation, porcine trypsin and TKI were mixed and the mixture was incubated at 277 K for 2 h. As described above, size-exclusion chromatography was employed to purify the protease–inhibitor complex from the incubated mixture of TKI and trypsin. The molecular weight of the major elution peak from a Superdex-200 GL 10/300 (GE Healthcare) size-exclusion column was calculated using a standard curve and was estimated to be ~44 kDa, which corresponded to the expected molecular weight of the TKI–trypsin protein complex. Additionally, to confirm the presence of the complex in the major elution peak, peak fractions were run on reducing 15% SDS–PAGE and protein bands corresponding to trypsin (~23 kDa) and TKI (~21 kDa) were observed (Fig. 4.2). The fractions containing the complex were subsequently pooled and concentrated to 15 mg/ml. TKI:PPT complex crystals were obtained in 10-12 days in reservoir solution containing 0.2 M ammonium phosphate dibasic and 20% PEG 3350 (Fig. 4.3). The crystals of TKI:PPT complex belonged to the space group $P2_122_1$ and diffracted to 2.2Å resolution in-house. The unit-cell parameters were found to be a = 61.08, b = 67.12, c = 91.61Å, with one molecule per asymmetric unit.

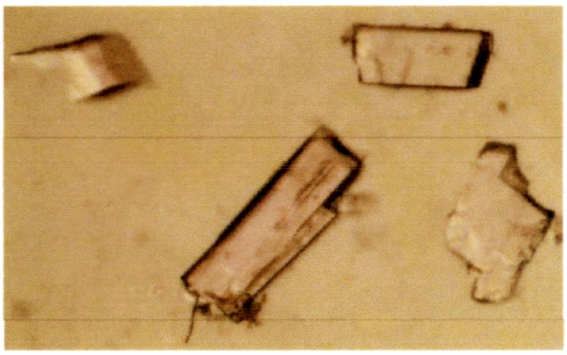


Figure 4.1. Crystals of TKI from tamarind seeds

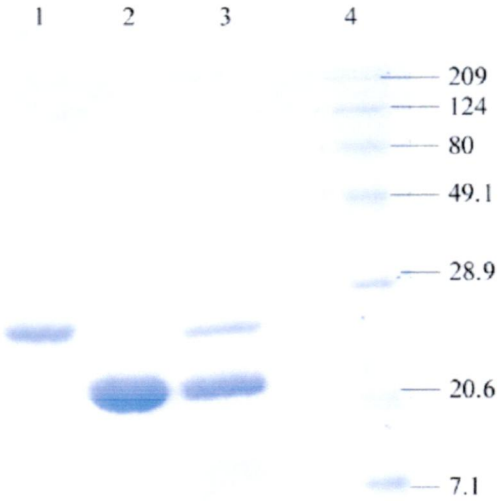


Figure 4.2. 15% SDS–PAGE analysis of purified TKI:PPT complex. Lane 1, porcine trypsin; lane 2, TKI; lane 3, the major gel-filtration peak fraction containing the complex between TKI and trypsin; lane 4, protein molecular-weight markers (kDa)

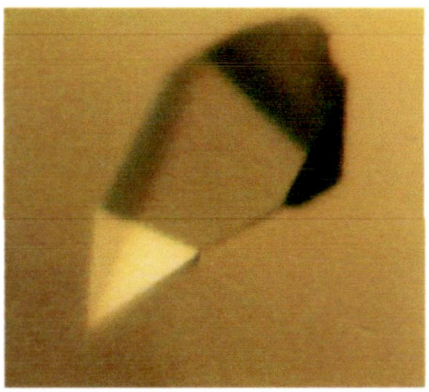
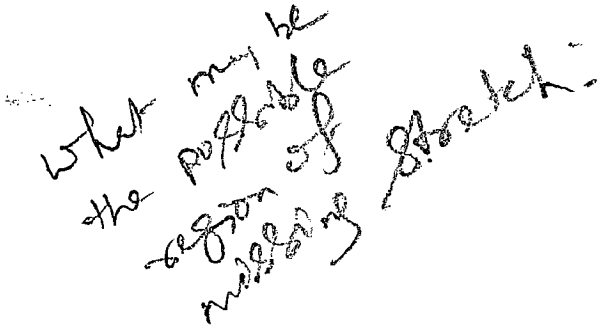


Figure 4.3. A crystal of the complex of TKI with trypsin

4.3.2. Quality of structures

The crystallographic data and refinement statistics for free TKI and TKI:PPT structures are summarized in Table 4.1. TKI structure was determined at 1.93Å resolution and contains two molecules per asymmetric unit. The free TKI structure was solved by molecular replacement method using trypsin inhibitor from *Erythriana caffra* seeds (PDB ID: 1TIE) as search model and refined to an R_{cryst} of 22.7% and R_{free} of 27.0%. The model is well defined by electron density however; the electron density of the residues 1, 103, 104, 153-155 in chain A and residues from 1, 138 to 141 in chain B and C-terminal residues (177-185) from both the chains was not observed. The final model thus consists of 341 residues and 146 solvent molecules in both the monomers. The Ramachandran plot analysis was done using PROCHECK program [106]. The refined model shows that 243 amino acid residues are in most favored regions and 35 are in the additionally allowed regions.

The structure of TKI:PPT complex was determined at 2.2Å resolution and refined to an R_{cryst} of 21.0% and R_{free} of 26.27%. The refined model consists of one molecule per asymmetric unit. The TKI:PPT model lacks three segments (residues 26-28; 96-104 and 136-141), one N-terminal residue and ten residues of C-terminal of TKI. The complex structure consists of 379 residues, 147 solvent molecules and a Ca^{2+} atom. The refined model shows that 268 amino acid residues are in most favored regions and 44 are in the additionally allowed regions of Ramachandran plot.



	Free TKI	TKI :PPT complex
Crystallographic data		
Space group	$P2_12_12_1$	$P2_122_1$
Wavelength	1.54179	1.54179
Resolution	50.0-1.93	91.6-2.22
Cell dimensions		
a (Å)	40.43	61.08
b (Å)	60.42	67.12
c (Å)	105.53	91.61
Unique reflections	19581	17821
Completeness (%) (Last shell)	97.5 (75.7)	92.9 (40.3)
R_{sym}(%)[a] (Last Shell)	4.3(25.7)	7.6(40.0)
I/σ (Last shell)	19.4 (3.6)	11.86 (2.0)
Multiplicity (Last shell)	3.0 (2.4)	3.3 (2.0)
Refinement	29.05–1.94	91.6-2.23
No. of reflections (working/test)	17690 (16726/964)	16883 (15972/911)
No. of residues	341	379
Water molecules	146	147
Resolution range (Å)	1.94	2.23
R_{cryst} (%)	22.6	21.0
R_{free} (%)	27.0	26.2
Average *B*-factors (Å^2)	A 25.27 B 29.39	A 26.41 B 33.03
Water atoms	33.9	41.0
All atoms	27.68	29.8
rmsd on bond lengths (Å)	0.0065	0.007
rmsd on bond angles (Å)	1.55	1.23
Ramachandran plot (%)		
Preferred	86.0	85.6
Allowed	12.4	14.0
Outliers	1.8	0.3

$$R_{sym} = \sum_{hkl}\sum_{i=1}^{n}\left|I_{hkl,i} - \bar{I}_{hkl}\right| \Big/ \sum_{hkl}\sum_{i=1}^{n} I_{hkl,i}$$

Table 4.1. Crystallographic data and refinement statistics for free TKI and its complex with PPT

4.3.3. Overall structure of TKI and its complex with PPT

The overall structure of TKI is similar to other Kunitz-type inhibitors. It is devoid of helical structures and composed of β-strands and turns. The protein shows β-trefoil fold consist of 12 anti-parallel β-strands connected with long loops similar to STI (Soybean Trypsin Inhibitor), which is an archeal member of Kunitz-type trypsin inhibitor [119, 136]. These β-strands form six double stranded β-hairpins. Three β-hairpins form barrel structure and the remaining form cap like structure on barrel (Fig. 4.4A). The TKI structure can be divided into three structurally repeated sub domains, which are related by pseudo-three fold symmetry parallel to the barrel axis of β-barrel. The repetitive sub domain consists of approximately 60 residues, which can be further divided in to four β-stranded motifs residues (18–23, 30–35, 45–49 and 59–62 form β-strands A1–A4 in subdomain A; residues 76–80, 92–94, 109–111 and 120–124 form β-strands B1–B4 in subdomain B and residues 129–137, 140–150, 158–161 and 169–175 form β-strands C1–C4 in subdomain C). TKI contains two disulphide bonds (Cys42-Cys86 and Cys134-Cys144) at the surface also observed in other Kunitz-type of structures [41, 144, 187], which stabilize the three dimensional structure. TKI structure does not have 3_{10}-helix similar to STI due to deletion of three amino acids at the respective position. This type of deletion is also observed in other Kunitz-type inhibitors like ETI, WCI and DrTI [41, 96, 144]. TKI contains two molecules per asymmetric unit and superposition of these two subunits shows variation of the loop regions and also at the reactive site (Fig. 4.4B). The some loops regions of two subunits of TKI also shows high B-factor which could be due to flexibility of theses loops (Fig. 4.5).

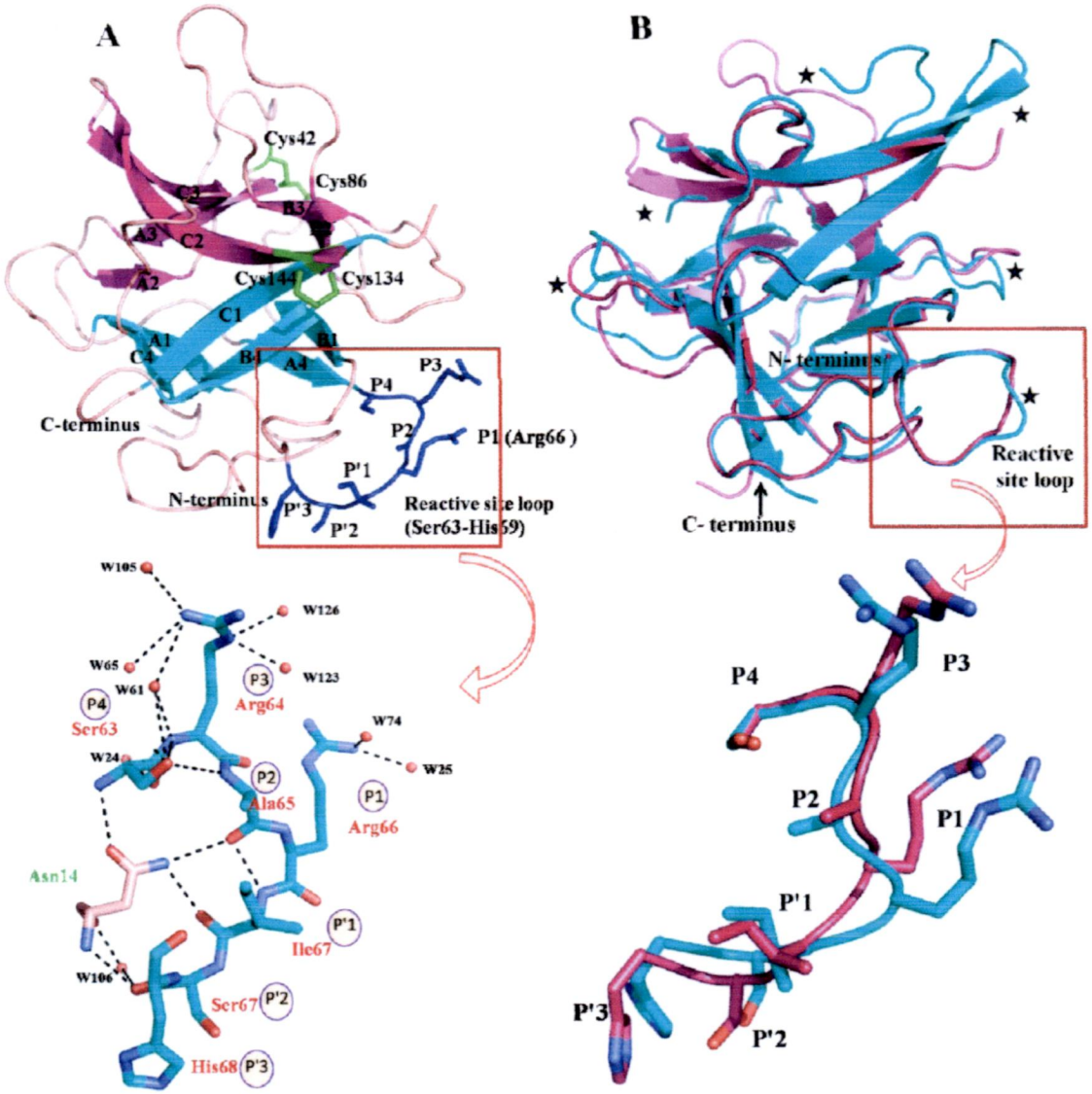


Figure 4.4. A. Cartoon diagram presentation of overall structure of TKI. Magenta and cyan color represents β- strands of lid and β-barrel respectively. Two disulphide bonds (C42-C86, C134-C144) shown in green. The reactive site loop P4-P'3 (Ser63-His69) is represented in blue showing residue Arg66 is P1. The structure was submitted in PDB database (PDB ID code: 4AN6). Architecture of reactive site loop of TKI with the networks of hydrogen bonds interactions around active site residues is highlighted. Reactive site loop residues are shown in cyan carbon sticks. The vital residue Asn14 which hold the canonical conformation is shown in light pink carbon stick. The black dash lines represent hydrogen bonds and water molecules shown in red spheres. B. Conformationally variable loops of TKI subunit A and B are shown by stars. Variable reactive site loop is highlighted showing difference in P1, P3 and P'3 residues.

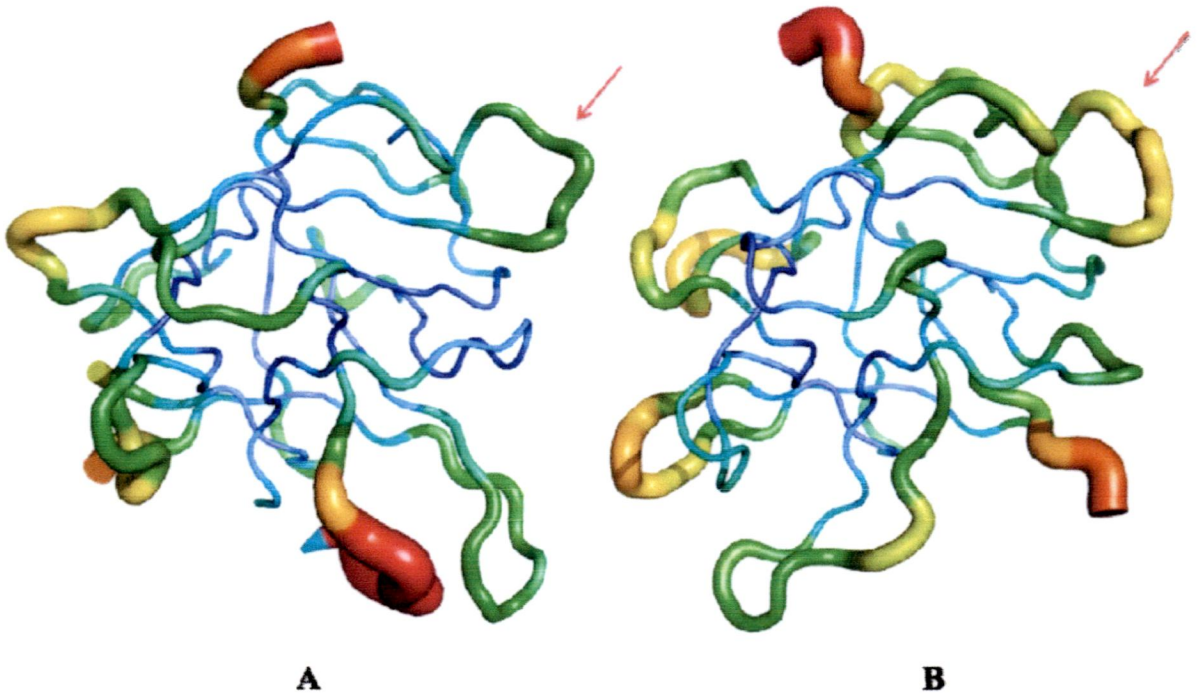


Figure 4.5. High B factor of TKI subunit A and B. Reactive site loop is shown by arrow which also shows difference in B factor.

The overall structure of TKI:PPT complex is similar to the STI:PPT complex structure but the difference lies in the interacting residues of TKI and pattern of stabilization. A conformational change was observed at P1 position after binding to PPT and the temperature factor of reactive site loop decreased. Two water molecules (Wat804 and Wat810) are present at active site in STI:PPT structure, which are engaged in the interaction between STI and PPT are absent in TKI:PPT structure. A water molecule (W34) present between Aβ3, Bβ3 and Cβ3 triangular structure in STI:PPT complex is absent in TKI:PPT complex (18). One another molecule (W220) between the three β-strands in ETI is also absent in TKI:PPT complex [144, 187]. One Ca^{2+} atom is observed in complex model which interacts with OE1 and OE2 of Glu70; carbonyl O of Asn72 and Val75; OE1 of Glu77 and OE2 of Glu80 residues of PPT. This Ca^{2+} atom is also observed at similar position in STI:TKI model. Overall complex structure and detailed interaction of TKI with PPT is shown in Fig.4.6 and Fig. 4.7.

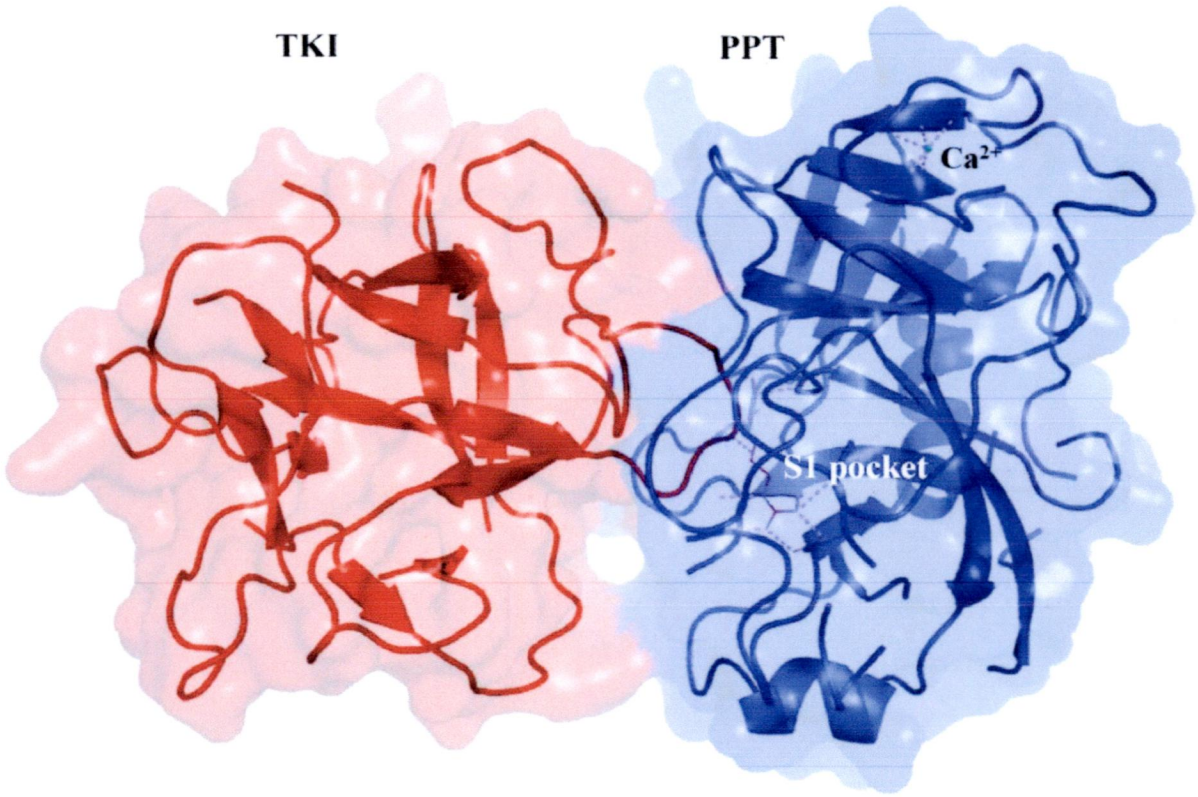


Figure 4.6. Overall structure of TKI:PPT complex. TKI and PPT is shown in red and blue color respectively. The residue P1 enters in S1 pocket and interacts with residues of PPT (purple dashes) and one Ca^{2+} atom is also observed (green) which interacts with PPT residues shown in purple dashes.

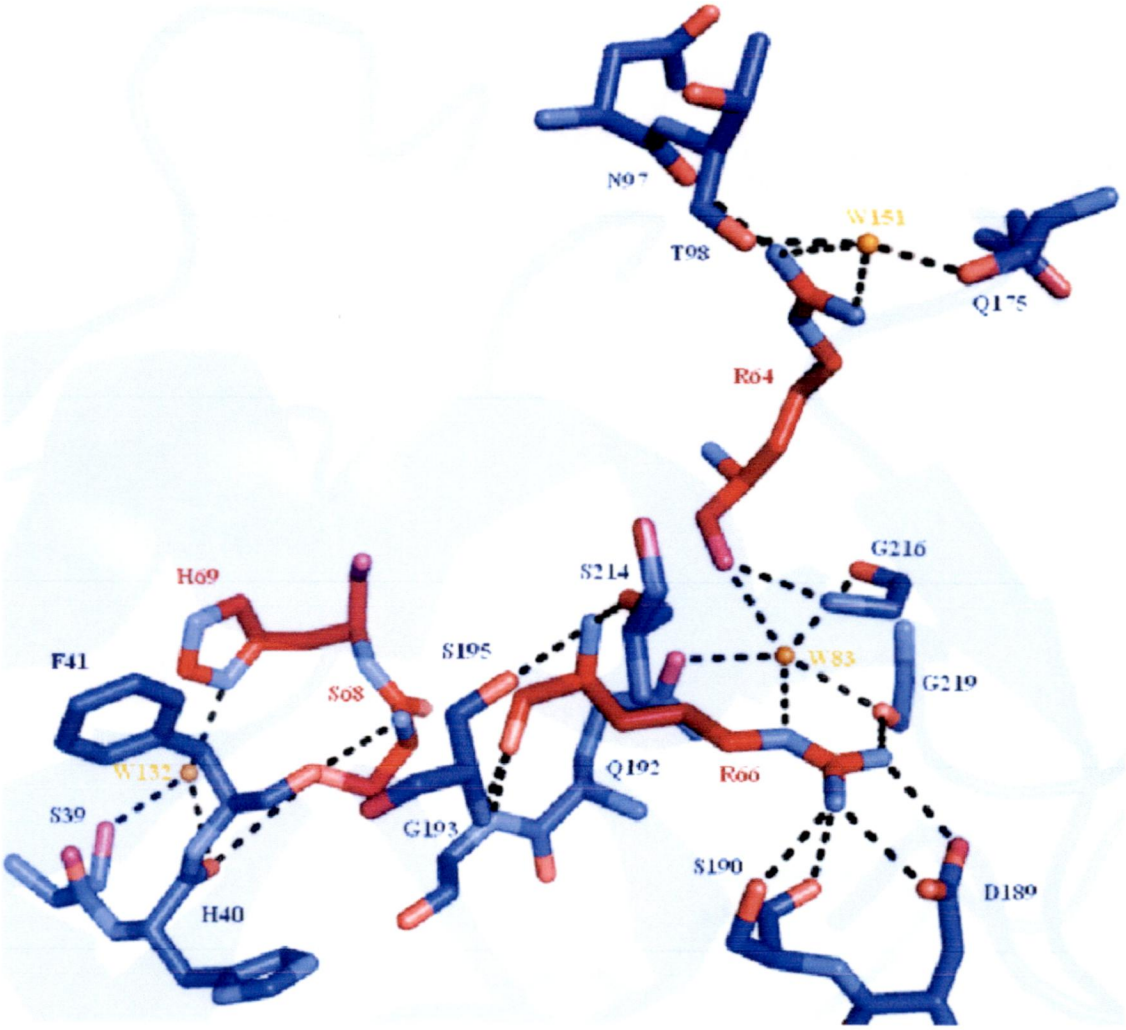


Figure 4.7. Detailed hydrogen bond interactions between TKI and PPT. Interacting residues of TKI are shown and labeled in red. The residues of PPT are shown and labeled in blue. Water molecules are in orange color and hydrogen bonds represented in blacks dashes. Four residues of reactive site loop of TKI P1, P3, P'2 and P'3 also participated in hydrogen bond formation with residues of PPT.

4.3.4. The geometry and conformation of reactive site loop

TKI reactive site loop P4 to P'3 (Ser63-Arg64-Ala65-Arg66-Ile67-Ser68-His69) is devoid of secondary structures and disulphide bonds. Arg66 present at P1, which is identical to its closely related ETI and STI, is defined well in electron density map. The reactive loop of TKI is showing overall high B factor of P1-P4 residues in TKI structure. Asn14 is a conserved residue in STI family, which stabilizes the reactive site loop in a similar manner as shown in other related member of STI Kunitz-type inhibitor family [40, 187]. ND2 atom of Asn14 forms hydrogen bonds with the P2 and P'1 carbonyl O which are present at both side of the scissile bond. Asn14 also interacts with P4 and P'2 and maintains the canonical conformation of the reactive loop which is a characteristic feature of this family. This type of confirmation makes TKI a member of substrate like inhibitor family. This Asn called as spacer Asn residue which is conserved in the Kunitz STI family is also present in the other serine protease families such as, kazal, ecotin, SSI, grasshopper and potato II [40]. The side chains of residues at P1 and P3 and P'3 are flexible which may fit optimally into the active site of cognate proteases by adapting the conformation.

4.3.5. Comparison of free TKI with homologous structures

The structural studies demonstrated that TKI adopt overall β-trefoil fold with two disulphide bonds similar to Kunitz inhibitors. Although the overall fold of TKI show structural similarity with Kunitz STI inhibitor structures, the two subunits in asymmetric unit of free TKI show variation in loop regions (Fig. 4.4B). The two of the connecting loops (L6 residue 95-108 and L7 residue 112-119) of TKI show high degree of deviation among themselves. Apart from connecting loops, the two β-sheets B1 (129–137) and B2 (140–150) exhibit different orientation between the two subunits of TKI.

Erythrina caffra trypsin inhibitor [144], soybean trypsin inhibitor [187], trypsin inhibitor from *Copaifera langsdorfii* seeds (CTI) [95], Kunitz-type winged bean chymotrypsin inhibitor (WCI) [41] and Kunitz type inhibitor from seeds of *Delonix regia* (DrTI) [96] share close similarity with TKI. ETI exhibit the highest sequence identity (39%) with TKI showing 1.05Å r.m.s. deviation between superimposed 152 Cα atoms of TKI polypeptide chain. STI shares 33% sequence identity with r.m.s. deviation 0.9Å while superposing 160 Cα atoms. TKI differs from STI as it does not contain any 3_{10} helix. The sequence identity between TKI and WCI is 31% and the r.m.s. deviation 1.3Å (for 158 Cα atoms). When ETI, WCI and STI models are compared

with each other, the r. m. s. deviation for ETI and STI is 0.8Å, ETI and WCI is 0.5Å and STI and WCI is 0.9Å respectively.

The superposition of Cα atoms of the reactive site loop residues of TKI [P4 to P'3] with the ETI, STI and WCI exhibits r. m. s. deviations of 0.8Å, 0.7Å and 0.7Å respectively. The superposition of the reactive site loop residues of TKI subunit A with subunit B shows different conformations with r. m. s. deviation of 0.8Å. The TKI P1 residue when compared with ETI and STI showed Cα distance between ETI and TKI A and B subunit is 4.6Å and 2.7Å respectively. While, STI P1 Cα distance is 1.3Å and 1.0Å with P1 of TKI A and B subunit. The Cα distance between P1 residue of TKI A and B subunit is 2.1Å, which elucidate that P1 residue is considerably variable and can also interact with other trypsin like proteases.

CTI is an inhibitor of Kunitz [95] family but differs from classical inhibitors of this family due to presence of two polypeptide chains bound non-covalently and a single disulfide bond. Inspite of these differences CTI shows high inhibitory activity against trypsin. Only one disulfide bond between residues C40 and C84 is present and second disulfide bond is absent which is conserved in majority of Kunitz-type inhibitors. It has been confirmed by biochemical studies that even after reducing the second disulfide bond in STI it is able to maintain inhibitory action but loses activity in completely reduced state [45, 95]. CTI shares 41% sequence identity with TKI and r.m.s. deviation of superposition of TKI and CTI is 0.5Å (for 157 Cα atoms). DrTI is a kunitz family member whose reactive loop is distorted due to insertion of Glu68 between P1 and P2. Regardless of that it is still effective against trypsin but with declined inhibitory activity [96]. DrTI shows 36% sequence identity with TKI and r.m.s. deviations of 1.4Å (for 161 Cα atoms). When TKI superimposed onto other Kunitz type inhibitors shows large differences in the connecting loops regions of the inhibitors (Fig. 4.8).

TKI also shares structural similarity of its β-trefoil fold with many proteins families having little sequence similarity and different functions such as tetanus/botulinum neurotoxin [102], Agglutinins [201], ricin B-like lectin [170], Cytokines [47] and Chlorophyll binding protein [75]. This structural similarity among functionally diverse group of protein families suggest that this fold arose from the gene duplication event of a common ancestor in due course of evolution.

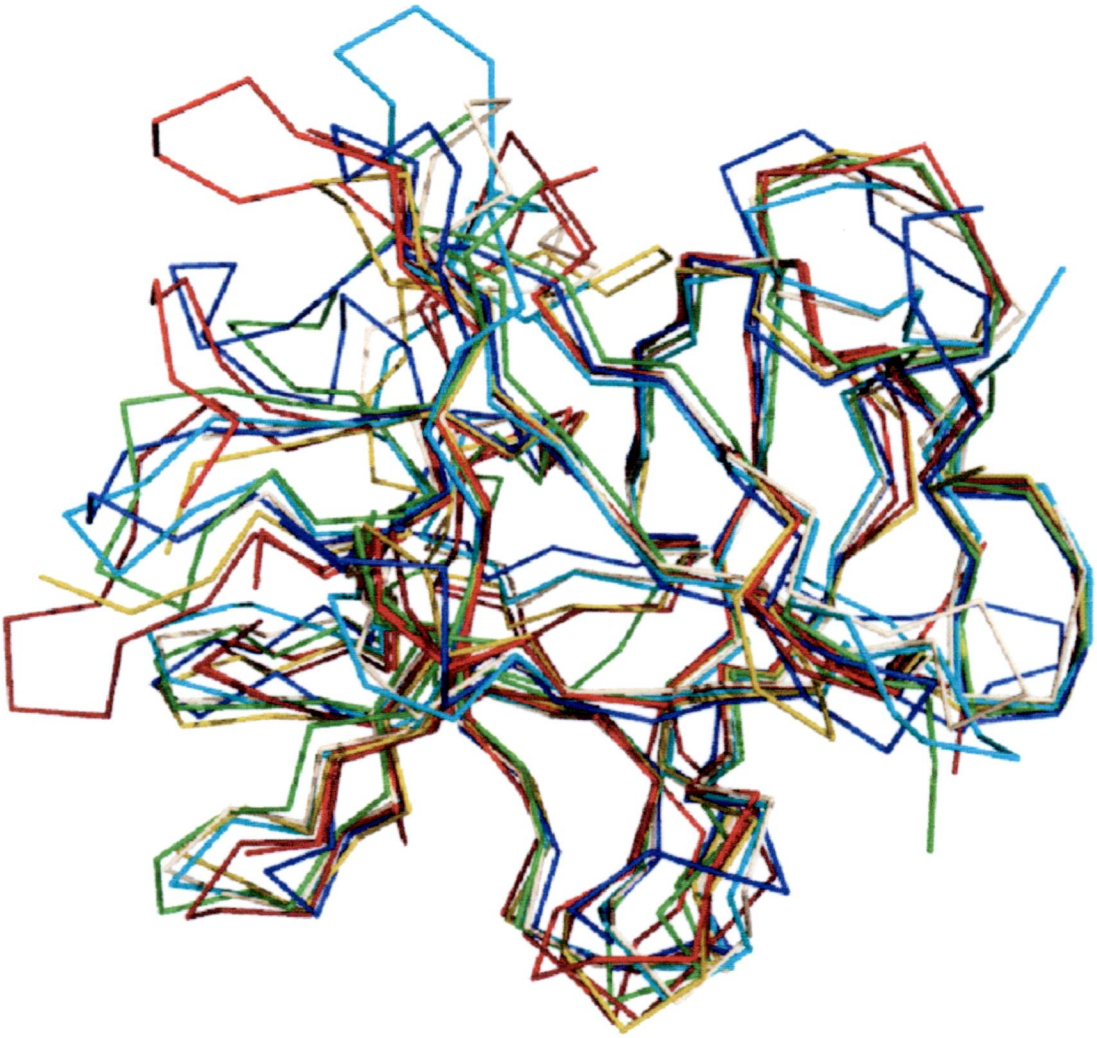


Figure 4.8. Ribbon diagram presentation of superposed TKI structure with other members of Kunitz type inhibitor family.Tamarind Kunitz type inhibitor subunit A and B (4AN6), soybean trypsin inhibitor (1AVU), *Erythriana caffra* trypsin inhibitor (1TIE), wingedbean chymotrypsin inhibitor (4WBC), inhibitor from seeds of *Delonix regia* (1R8N) and trypsin inhibitor from *Copaifera langsdorfii* seeds (1R8O) are shown in red, firebrick, green, wheat, cyan, blue and olive ribbon respectively and shows the overall conserved structure except some loop regions.

4.3.6. Comparison of free TKI and TKI:PPT complex structure

The comparison between free TKI and TKI:PPT complex shows slight conformational changes in the reactive site loop and P1 residue. Superposition of free TKI subunit A and B with TKI:PPT gives an r.m.s deviation of 0.4Å and 0.5Å respectively and its reactive site loop shows decreased in the B-factors compared to both subunits of TKI. Superimposition of TKI:PPT complex with TKI A and B is shown in Fig. 4.9. The reactive site loop of TKI subunit A and B is superimposed onto the TKI:PPT complex showing r. m. s. deviation of 0.6Å and 0.5Å respectively. The relatively flexible P1 residue of reactive site loop of free TKI shows lower B factor when complexed with PPT. This illustrates that P1 becomes rigid after binding with PPT and stabilize the complex by making interactions with residues of S1 pocket and other residues of PPT. Superposition of reactive site loop of TKI:PPT complex with TKI subunit A and B discloses that there is immense divergence at Position P1. The distance between Cα of P1 in TKI:PPT with TKI subunit A and B is 1.4Å and 0.8Å respectively. The difference is also examined at side chain of Arg66 (P1). The distance between NH1 and NH2 of Arg66 of TKI in complex structure and NH1 and NH2 of Arg66 of TKI A is 5.9Å and 5.6Å and with TKI B is 4.2Å and 2.3Å respectively. The details of conformation are shown in Fig. 4.9 (Panel A). Another residue, P3 of reactive site also shows conformational variation. P3 becomes rigid after forming complex with trypsin and participates in stabilization of complex by interacting with residues of PPT. Superposition of TKI subunit A and B on the TKI:PPT complex shows minor change in Cα and the difference in side chain of P3. The distance between NH1 and NH2 Arg64 of TKI in complex structure and NH1 and NH2 Arg64 of TKI A is 3.8Å and 1.4Å and with TKI B is 2.0Å and 2.1Å (Fig. 4.9 Panel B). When compared with free TKI, one another amino acid of reactive site loop of TKI in complex structure shows conformational rotation which is His69. It moves away from Ser68 to prevent short contact and acquires a position where it interacts with Ser39 and His40 of PPT through water molecule (W132) and may contribute for stabilization of the complex. The details of hydrogen bond pattern and conformation of His69 as compared to free TKI is shown in Fig. 4.9 (Panel C).

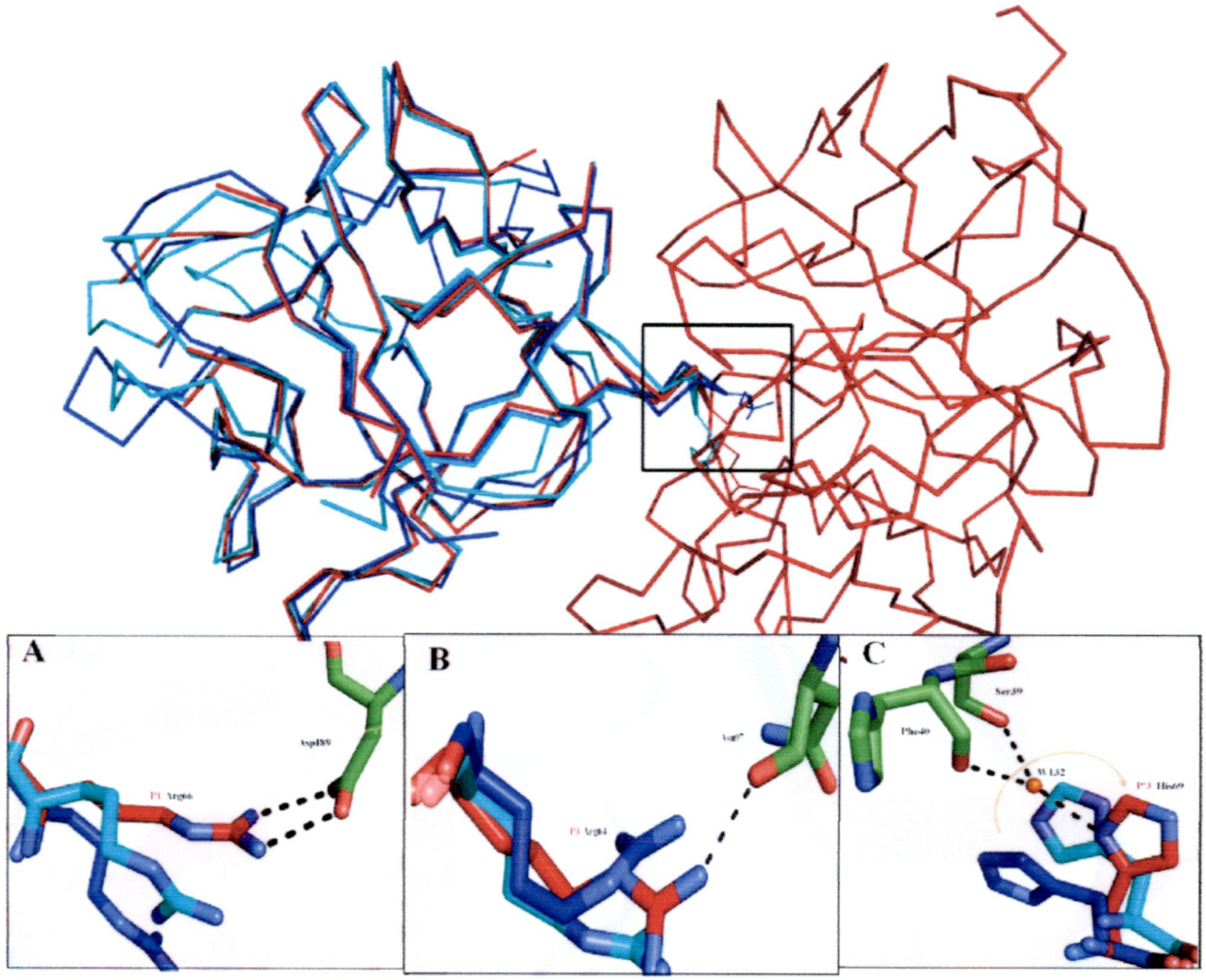


Figure 4.9. Comparison of free TKI and TKI:PPT complex structure. Superposition of free TKI subunit A and B with TKI:PPT represented in blue, cyan and red respectively shows variation at reactive site loop. Conformation of P1 and P3 residue is highlighted shows variation of Cα as well as NH1 and NH2 atoms (panel A and B). Additionally, the conformational rotation of P'3 residue, His69 is also observed. Its moves to a position from where it gets interacted with Ser39 and Phe40 of PPT through water molecule (panel C).

4.3.7. Comparison of TKI:PPT with other Kunitz type inhibitors complex with trypsin

TKI shares 33% sequence identity with STI whose complex structure with trypsin has been solved. TKI:PPT superimposed on the STI:PPT shows r.m.s. deviation of 0.8Å and their reactive site loop shows r.m.s deviation of 0.4Å which is less than the r.m.s. deviation between reactive site loop of free STI and TKI models. As compared to STI:PPT, some water molecules are absent in TKI:PPT model, even at near reactive site residues. The interacting residues of STI and TKI are different with PPT and consequently way of stabilization of complex differs in both cases. Only quite similar interactions are observed in Arg66 (P1) residues. The interactions of Asn15, Arg64 (P3), Ser68 (P'2), His69 (P'3) and Ser116 of TKI with PPT makes it distinct from STI:PPT complex. The detailed interaction of TKI:PPT are discussed below.

TKI also shares 31% sequence identity with Kunitz-type winged bean chymotrypsin inhibitor (WCI) [41]. The complex structure with bovine trypsin of engineered trypsin inhibitor (F64Y/L65R) (pdb id: 3I29) and (L68R) (pdb id: 2QYI) of WCI superimposed on TKI:PPT structure shows r.m.s.d. of 1.5Å and 1.8Å. The interactions of residue Arg at P1 are nearly similar but the water molecules present near active site in L68R mutant inhibitor complex are absent in TKI:PPT complex. Double-headed arrowhead protease inhibitor A (API-A) has less sequence identity (28%) with TKI and its complex with two trypsin molecules (pdb id: 3E8L) superimposed with TKI:PPT showed 1.1Å r.m.s. deviation. API-A has two reactive sites with Leu87 and Lys145 at P1 position. The RS1 having Leu87 adopts novel conformation and fits well in S1 pocket of trypsin [10].

4.3.8. Mode of interaction between TKI and PPT

In TKI:PPT complex structure, six residues of TKI (Thr3, Arg64 (P3), Arg66 (P1), Ser68 (P'2), Ser75 and Ser116) make interaction with PPT residues and four residues of TKI (Asn15, Gly17, Arg60 and His69 (P'3)) interact through water with residues of PPT (Fig. 4.7). Total 10 residues of TKI interact with 20 residues of PPT and form dense network of hydrogen bonds and produce stable complex. Detailed interactions between TKI and PPT are shown in Fig. 4.7 and Table 4.2. Four residues of reactive site of TKI form 21 hydrogen bonds directly and through water. The P1 residue Arg66 interacts with PPT and forms large number of comprehensive hydrogen bonds, accounting 11 in total (Table 4.2). The side chain of Arg66 enters in the S1 pocket of PPT and interacts with residues of S1 pocket. The guanidinium group of Arg66 forms a salt bridge with the carboxylate group of Asp189 of PPT. The Arg66 CO interacted with

Gly193 N and Ser195 N; Arg66 N atoms with Ser195 OG and Ser214 CO; Arg66 NH1 with Ser190 OG and CO, Asp189 OD1 and Arg66 NH2 with Asp189 OD2, Gly218 CO of PPT. Arg66 also interacted with Gln192 and Gly216 of PPT through water (W83). The scissile bond was found uninterrupted in electron density. The P3 residue Arg64 CO and NH2 interacts with Gly216 N and Asn97 CO of PPT. Arg64 also interacts with Thr98, Gln175, Gln192 and Gly218 through water. The Ser68 at P'2 position also participate in making hydrogen bonds with 2 residues of PPT. Ser68 OG and N interacts with His40 CO and Phe41 CO of PPT, respectively. The P'3 residue, His69 shows slight conformation as compare to free TKI and move toward the water molecule (W132) and interacts with Ser39 OG and Phe40 CO atoms of PPT.

TKI	PPT	Water	No. of interactions	Distance (Å)
Thr3	**Lys60**		1 (OG1-NZ)	2.8
Asn15		W25	(NH2 and CO)	3.0, 3.3
	Gln192		NE2	2.6
		W25	(NH2 and CO)	3.0, 3.3
	Asn143		ND2	3.2
		W25	(NH2 and CO)	3.0, 3.3
		W154		3.0
	Lys145		CO	3.0
		W25	(NH2 and CO)	3.0, 3.0
		W154		3.0
	Gly148		CO	3.5
Gly17		W140	N	3.2
	Ser149		OG	3.0
Arg60		W140	NH2	3.4
	Ser149		OG	3.0
Arg64	**Asn97**		1 (NH2-CO)	2.8
	Gly216		1 (CO-N)	3.1
		W151	(NH1 and NH2)	3.0, 3.4
	Thr98		CO	2.7
		W151	(NH1 and NH2)	3.0, 3.4
	Gln175		CO	2.7
		W83	CO	3.2
	Gln192		OE1	2.9
		W83	CO	3.2
	Gly219		CO	2.8
Arg66	**Gly193**		1 (CO-N)	2.7
	Ser195		2 (CO-N)	2.9
			(N-OG)	3.0
	Ser214		1 (N-CO)	3.0
	Ser190		2(NH1-OG)	2.9
			(NH1-CO)	3.0
	Asp189		2 (NH1-OD1)	2.9
			(NH2-OD2)	2.8
	Gly219		1 (NH2-CO)	3.2
		W83	NE	3.2
	Gly216		CO	2.9
		W83	NE	3.2
	Gln192		OE1	2.9
Ser68	**His40**		1 (OG-CO)	3.5
	Phe41		1 (N-CO)	3.2
His69		W132	ND1	2.6
	Ser39		OG	2.9
		W132	ND1	2.6
	His40		CO	3.0
Ser75	**Asn97**		1 (OG-ND2)	3.3
Ser116	**Tyr217**		1 (OG-OH)	3.4

Table 4.2. Total interactions of residues of TKI with PPT

4.3.9. Molecular docking of TKI with factor Xa

As we did not get the crystals of TKI complexed with factor Xa, we performed rigid-body docking using program ClusPro in order to analyze the possible binding of TKI with factor Xa. Fig. 4.10 shows detailed interactions of one of the 30 highest ranking structures calculated by ClusPro for TKI-factor Xa complex. The active site of factor Xa is divided into four sub pockets as S1, S2, S3 and S4 [69, 156]. The S1 sub pocket is responsible for substrate selectivity and binding while the S4 sub-pocket consists of three ligand binding domains, namely the "hydrophobic box", the "cationic hole" and the water site. Most of the factor Xa inhibitors generally bind to their substrate in L-shaped conformation, where one group of the ligand occupies the anionic S1 pocket marked by residues Asp189, Ser195, and Tyr228, and another group of the ligand occupies the aromatic S4 pocket lined by residues Tyr99, Phe174, and Trp215 [156].

The docking studies of both the subunits of TKI with factor Xa showed the interactions with S1 and S4 pocket of factor Xa in a classical L-shaped substrate like conformation (Fig. 4.10B). The active site residues (P1 and P3) of subunit B have more number of favorable stabilizing interactions with S1 and S4 pocket of factor Xa than subunit A. various types of interaction of residues of TKI with factor Xa residues were calculated by PIC server. Mainly three loops residues interact with FXa residues namely loop β1, loop β4β5 and loop β7β8. Residues Gln18, Arg60, Arg66, His69, Glu79 and Lys136 of TKI subunit B forms salt bridges with Lys148, Glu147, Asp189, Glu39, Arg222 and Glu97 of FXa, respectively. Out of seven residues of reactive site loop, six residues interact with FXa residues and participate in stabilization of complex. The detailed interactions are shown in Table 4.3 and Fig. 4.10C. The Arg66 (P1) and Arg64 (P3) of B subunit involved in the hydrogen bonding network with eleven residues of factor Xa with total sixteen hydrogen bonds whereas subunit A involved in seven hydrogen bonds formation with five factor Xa residues. The Arg66 (P1) of TKI subunit B intrudes inside the substrate binding pocket S1 of factor Xa making direct interactions with three marking residues of S1 pocket. The Arg66 form three hydrogen bonds with Asp189, two with Ser195 and one with Tyr228. In addition to that the Arg66 also form stabilizing interactions with other residues like hydrogen bonds with Ala190, Gly218, Gly193, Asp194 and Ser214.

The positively charged side chain of Arg64 (P3) of TKI subunit B accessed the S4 pocket of factor Xa parallel to the π electron rich aromatic ring of Phe174 of factor Xa and may

participate in cation-π interaction (Fig.4.10B). The cation–π interaction can take place either in parallel, perpendicular or T-shaped orientation [109, 123]. PIC server predicts cation-π interaction with Tyr99, Phe174 and Trp215 of S4 pocket. The Arg64 further stabilize the complex by forming hydrogen bonds with Glu97, Thr98 and Gly216. Other reactive site loop residues Ser63 (P4), Ala64 (P2), Ser68 (P'2) and His69 (P'3) also interact with FXa residues and may form stable complex. Interaction of Ser63 with Tyr99 (one of the important residue for specificity at S4 subsite of FXa [161]), Gln192; Ala64 with Gln192; Ser68 with Glu39, Phe41, Arg143 and His69 with Glu39 indicates that reactive site residues of TKI favours tight binding to FXa. Gln192 which is reported as a vital residue in FXa for specificity and binding [162, 214] interacts with four important residues of TKI forming 6 hydrogen bonds (Table 3). Two of them are reactive site loop residues Ser63 (P4) and Ala65 (P2); one is spacer residue Asn14 which is important for stabilization of reactive site loop and maintain the canonical conformation of the reactive site loop and one is insertion residue, Asn15 which is observed in TKI. Some loop residues also take part in stabilization of complex such as Ala73 and Ser75 of β4β5 loop interact with Lys96 and Glu97 of FXa respectively. One of flexible loop β7β8 residues Ser116 and Ala118 also make hydrogen bonds with Ser173 and Glu97 of FXa respectively. Other loop β1 residues interactions are, Asn14 with Gln192; Asn15 with Gln192, Glu147 and Gly17, Gln18 with Lys148 of FXa respectively (Table 4.3).

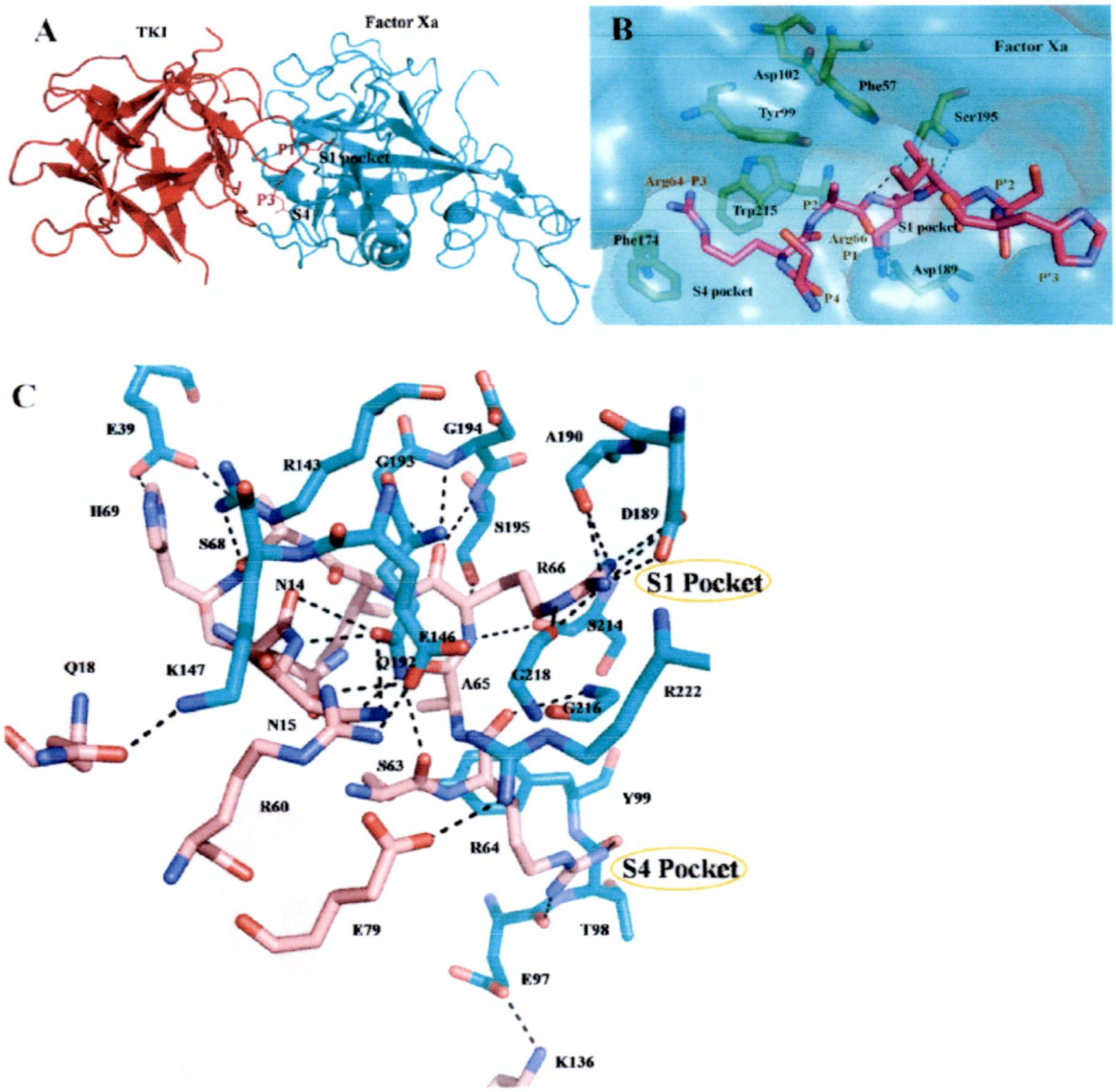


Figure 4.10. Molecular docking of TKI B with factor Xa. A. Overall fold of TKI B:FXa showing P1 blocking S1 pocket of FXa. TKI B and FXa are shown in red and cyan cartoon respectively B. The surface view of TKI:FXa complex showing Arg66 and Arg64 of TKI B intruding inside the S1 and S4 pocket of FXa respectively in L-shaped substrate like manner. The side chain of Arg64 of TKI is parallel to the aromatic ring of Phe174 of FXa and may participate in cation-π interaction. C. The detailed hydrogen bond interactions in TKI B:FXa docked model is represented in black dashes and red and cyan carbon sticks for TKI B and FXa residues respectively.

TKI B	FXa	d (Å)
A) Salt bridges		
Gln18 OE1	Lys148 NZ	2.6
Arg60 NH2	Glu147 OE1	3.9
NH2	OE2	2.7
NH1	OE2	2.7
Arg66 NH2	Asp189 OD1	3.3
NH2	OD2	2.6
NH1	OD1	3.6
His69 ND1	Glu39 OE1	3.0
Glu79 OE2	Arg222 NH1	2.8
Lys136 NZ	Glu97 OE2	2.7
B) The reactive site loop (P4 – P'3) interactions		
Ser63 OG	Tyr99 OH	3.4
CO	Gln192 NE2	3.2
Arg64 NH1	Glu 97 CO	2.7
NH2	Thr 98 CO	2.6
CO	Gly216 N	3.0
	Phe174	cation-π interaction
Ala65 CO	Gln192 NE2	3.0
Arg66 NH1	Asp 189 OD1	3.5
NH2	OD2	2.7
NH2	OD1	3.3
NH1	Ala 190 CO	3.1
NH2	CO	2.7
CO	Gly193 N	2.7
CO	Asp194 N	3.2
CO	Ser 195 N	2.9
N	OG	3.3
N	Ser 214 CO	3.3
NH2	Gly218 CO	2.7
NE	CO	2.9
Ser68 OG	Glu39 OE2	3.0
OG	Phe41 CO	3.3
CO	Arg143 NH2	2.7
His69 ND1	Glu39 OE2	2.9
C) Gln192 of FXa interaction with TKI residues		
Ser63 NE2	Gln192 CO	3.2
Ala65 NE2	CO	3.0
Asn14 OE1	CO	3.5
Asn15 OE1	ND2	2.9
NE2	OD1	2.9
OE1	N	2.7
D) Additional interactions		
Ala73 CO	Lys96 NZ	2.8
Ser75 OG	Glu97 OE1	2.9
Ser116 OG	Ser173 CO	2.8
Ala118 N	Glu97 OE2	3.1
Asn15 ND2	Glu147 OE2	3.2
Gly17 CO	Lys148 NZ	2.6

Table 4.3. TKI interactions with Factor Xa

4.3.10. One of most vital inserted residue: Asn15

The sequence analysis reveals that TKI has insertion of Asn residue at position 15 within the Kunitz signature sequence and consequently, TKI becomes a signature sequence less inhibitor of Kunitz STI family. In free TKI structure, Asn15 shows variation in conformation. The C_α distance between free TKI subunit A and B is 1.7Å and their side chain is 3.0Å apart from each other (Fig. 4.11A). The interactions of Asn15 of free TKI subunit A and B with other residues are also different due to flexibility. Superimposition of TKI subunit A and B with STI and ETI shows that Asn15 is not superimposed with any residue of ETI and STI, on the contrary it points outwards from an exposed loop at the protein surface. The question arises at this stage that what is possible role of this inserted residue? The answer lies in the TKI:PPT complex structure and modeled structure of TKI:FXa. In TKI:PPT complex structure, Asn15 ND2 and carbonyl O forms hydrogen bond with Gln192 NE2 and Asn143 ND2 through WAT25 and further with Lys145 and Gly148 via second water molecule (WAT154).

Consequently formation of network of hydrogen bonds by Asn15 through two water molecules makes its interaction with PPT stronger and contributes to the stabilization of the TKI:PPT complex (Fig.4.11B and Table 4.2). Moreover, in modeled complex of TKI with factor Xa, the Asn15 also plays a vital role as it forms three hydrogen bonds with Gln192 of factor Xa. Previous report shows that Gln192 in factor Xa is a key determinant for specificity of substrate/inhibitor [162, 214]. Formation of hydrogen bonds with OD1, ND2 and N of Gln192 by Asn15 demonstrates firm interaction between these residues (Fig. 4.11C). Specific recognition of Asn15 by Gln192 of factor Xa and role in TKI:PPT complex, leads to the assumption that Asn15 is a key residue of TKI which can be selected by it cognate proteases for inhibitor recognition. Asn15 interacts not only with Gln192 but also with Glu147 of autolysis loop (secondary binding site) of factor Xa and these observations gives confirmation that Asn15 is also takes part in stabilization of the complex. Together, these results indicate that TKI has got a key residue insertion which is important in recognition by its proteases and stabilization of complexes. The insertion of Asn is also observed in BVTI (Fig. 3.4) but due to unavailability of its structure the role of Asn in BVTI cannot be postulated at this moment. CTI, a Kunitz type unusual inhibitor having two polypeptide chains also has insertion of Asp at the same place of Asn (Fig. 4.11D). From the interaction of Asn15 of TKI, we postulate that the Asp of CTI may also take part in formation of complex with trypsin and stabilize the complex by interacting with residues of trypsin.

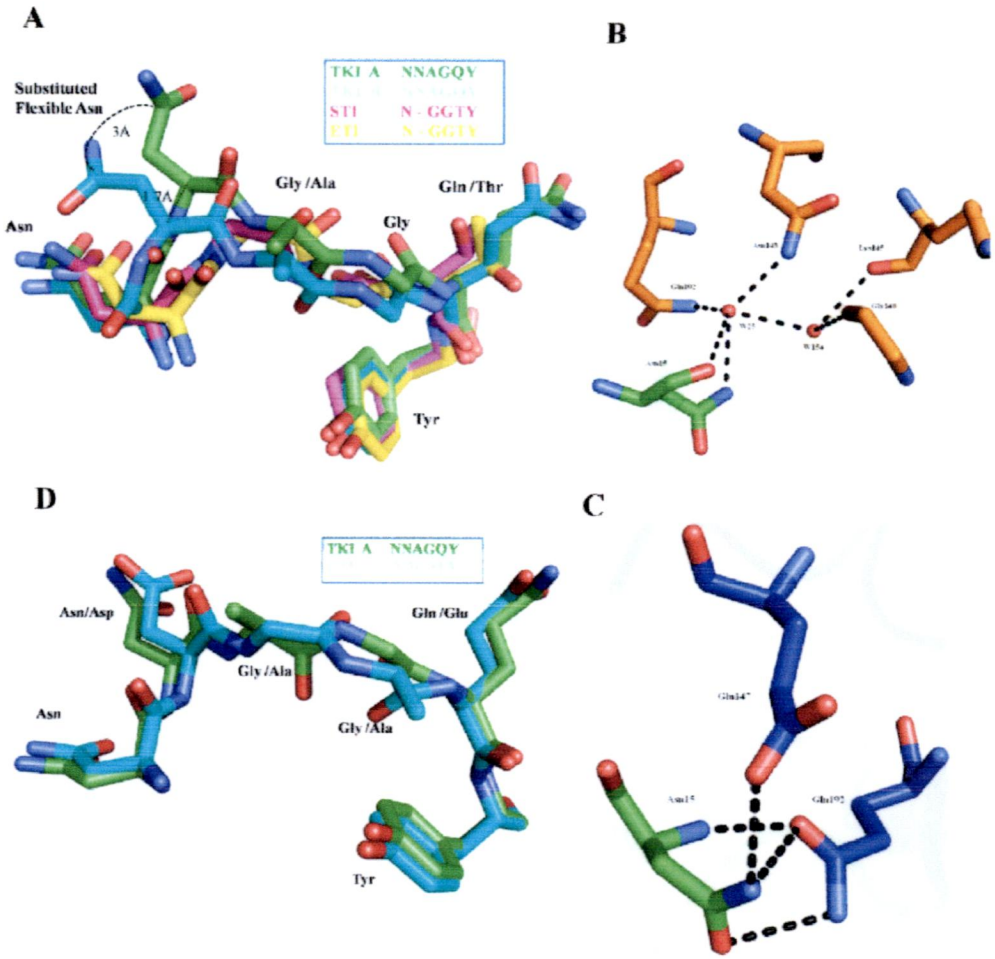


Figure 4.11. Insertion of Asn15 and its role in complex stabilization. Insertion of Asn15 within Kunitz signature sequence of TKI and its role in mediating interaction between inhibitor and its cognate proteases. A. Superposition of TKI subunits A and B with STI and ETI shows that Asn15 is not superimposed with any residue of ETI and STI. The Asn15 residue in both TKI subunits A and B are not superimposed on each other due to conformational variation and their side chains are 3Å apart. B. Interactions of Asn15 with Gln192 via WAT34 and with Asn143 and Lys145 of PPT via WAT34 and WAT49 are shown. Hydrogen bonds and water molecules are presented in black color and red spheres respectively. C. In modeled TKI: FXa complex, formation of three hydrogen bonds with a key determinant for specificity towards its substrates/inhibitor, Gln192 of factor Xa and one hydrogen bond with Glu147 of factor Xa demonstrate that Asn participates in rigid interaction with factor Xa. D. Superposition of TKI A with CTI shows that CTI also has insertion of Asp instead of Asn and it may also participate in complex formation with its cognate proteases.

4.3.11. Flexible loop β7β8 interacts with PPT and FXa

Superposition of TKIA on TKIB shows conformational variation at the loop β7β8 (Asp112-Pro119). The main chain C_α distance between them is 3.9Å. This loop is drastically different from respective loop of STI, around 11Å apart from each other. Loop β7β8 of TKI is closer to reactive site loop and its flexible nature indicates that it may interact with residues of proteases to which TKI binds and may stabilize the complexes. In the complex structure of TKI:PPT, it was observed that the Ser116 of the loop β7β8 interacts with residue Tyr217 of PPT. Furthermore, in modeled structure TKI:FXa, two residues Ser116 and Ala118 interact with Ser173 and Glu97 of FXa respectively (Table 4.3) and take part in stabilization of complex. From this observation it is concluded that the flexible loop β7β8 is also one of the important part in formation of complex. Such type of interaction is not observed in STI:PPT complex as the corresponding loop is far away from the reactive site loop in STI consequently it is not able to interact with PPT. In the complex structure with bovine trypsin of engineered trypsin inhibitor (F64Y/L65R) (pdb id: 3I29) of WCI, the corresponding loop residue Glu111 interacts with Ser217 and Lys224 of bovine trypsin. This loop may not be very much important in complex formation with PPT, but might be crucial for making rigid complex with other serine proteases. In modeled TKI-FXa complex, two residues of TKI interact with FXa residues. One of them interacts with Glu97 which is S4 pocket residue of FXa demonstrating that this loop may be one of the crucial loop engaged in complex formation with FXa and can further enhance affinity of TKI towards FXa.

4.3.12. Versatile reactive site loop

The reactive site loop in free TKI exhibit great mobility however complex formation with PPT results into rigidification of this loop. The reactive site of Kunitz STI type inhibitor shows diversification in its amino acids content as shown in Fig. 4.12A. The only similarity is observed at P1 residue which is basic in nature. P1 residue fits well at it cognate proteases S1 pocket in substrate like manner and forms salt bridge with carboxylate group of Asp189 of trypsin. The pattern of stabilization of complex with trypsin is different in both TKI and STI as they have different reactive site residues. The residue at P4 is serine which does not show any interaction with PPT but may be very crucial in FXa complex formation as in modeled complex with FXa, Ser63 interacts with two vital residues Tyr99 and Gln192 which defines specificity of FXa [161,

162, 214]. The Arg residue at P3 which is very uncommon in Kunitz STI inhibitors, interacts with Asn97, Thr98, Gln175, Gln192 Gly216 and Gly219 of PPT whereas in FXa modeled complex, it occupies the S4 hydrophobic pocket formed by Tyr99, Phe174 and Trp215 residues and interact with Glu97, Thr98 and Gly216 as its side-chain position towards an electronegative cavity of cation-binding hole of S4. Arg64 can also make cation-π interaction with residues of the S4 pocket of FXa. As Arg64 side chain is parallel to hydrophobic ring of Phe174 it can possibly to form cation-π interaction consecutively blocking S4 pocket of FXa (Fig. 4.10B). In TAP-FXa complex structure Arg503 occupies the S4 pocket and side-chains of Phe174, Tyr99 and Trp215 may develop favorable interactions with the hydrophobic side-chain moieties of Arg503. Moreover, it also interacts with Glu97, Thr98 and Gly216 and induces rearrangement of the side-chain of Glu97 and aromatic ring rotation of Tyr99 [216]. In the docked model of antistasin with factor Xa, Arg32 (P3) plays important role as it interacts with the carbonyl oxygens of Lys96 and Glu97 of the cation hole of factor Xa and also with the Glu97 carboxylate [104]. From these observations, we believe that the Arg64 (P3) of TKI is one of the most significant residues which favor key interactions with FXa and it may induce rearrangement of the side chains of S4 pocket residues like Arg503 of TAP.

Alanine at P2 site of TKI which is also unconserved in Kunitz STI inhibitors didn't show any interaction with PPT residues but it may interact with vital residue at S1 pocket, Gln192 of FXa. P1 residue Arg which is conserved among Kunitz STI inhibitors and has common mechanism of inhibition, enters into deep S1 pocket and its guanidinium group forms salt bridge with the carboxylate group of S1 residue, Asp189 and also forms hydrogen bond with Ser195. Similar interaction was observed in TKI:PPT complex in which P1 was making tight interactions with residues of S1 pocket of PPT such as two hydrogen bonds with Asp189, Ser190 and Ser195. Arg further established tight interaction by forming hydrogen bonds with Gln192, Gly193, Ser214, Gly216 and Gly219. Similarly P1 residue is able to make contacts with residues of S1 of FXa. The difference is only that trypsin has Ser190 and FXa has Ala190 even though Arg forms two hydrogen bonds with each residue. Ile at P'1 which forms scissile peptide bond didn't make any contact with PPT and FXa. Ser at P'2 interact with His40 and Phe41 of PPT and in case of complex with FXa it plays as one of the key determinant of specificity of TKI. Ser68 interact with Glu39 and Phe41 of acidic 36 loop and Arg143 of autolysis loop of FXa. Further, formation of salt bridge by Gln18 of TKI with Lys148 of autolysis loop of FXa indicates that residues of TKI will interact with positively charge residues autolysis loop of FXa and may facilitate stable

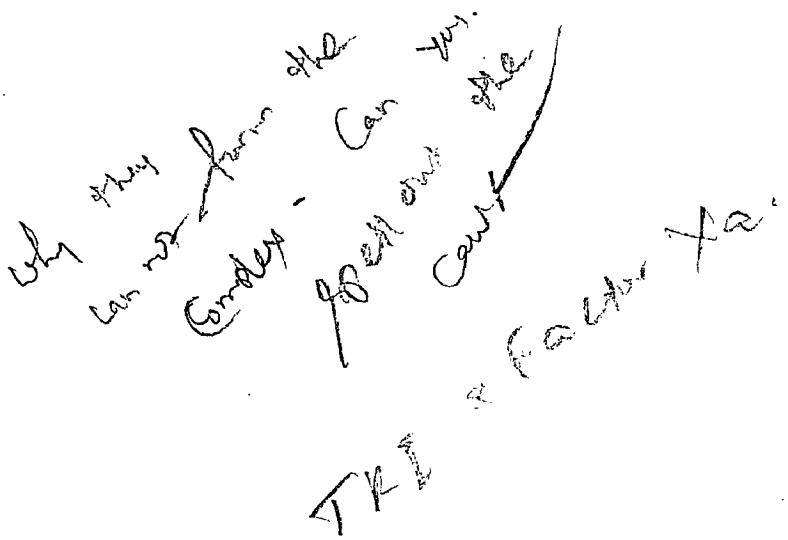
complex formation with FXa. The basic residues of autolysis loop and residues of acidic 36 loop of FXa play an important role in interacting with ZPI, a physiological inhibitor of factor Xa and substitution of these residues with Ala decreases the inhibitory potency (2 to 3 fold) of ZPI [217]. In the crystal structure of antithromobin-S195A factor Xa-heparin complex, residues of antithrombin forms salt bridges and hydrogen bonds with the 36-loop and the autolysis loop residues [84]. In the tick anticoagulant peptide-bovine FXa complex structure, the positively charged residues of the autolysis loop interact with the residues of TAP by forming hydrogen bonds and salt-bridges [216]. The modeled complex of antistasin-FXa and second Kuntiz domain of TFPI-FXa shows that the residues of inhibitor are able to interact with basic residues of autolysis loop of FXa [26, 104, 117].

His at P'3 is also one of the key residue that interacts with Ser39 and His40 of PPT via water and in FXa complex it was observed that His makes a salt bridge with Glu39 of acidic 36 loop. Other Kunitz type inhibitors such as STI, ETI, DrTI, CTI, BuXI, BvTI, EcTI and WCI do not possess Ser and His at P'2 and P'3 site respectively (Fig. 3.4). Conversely these sites were found with nonpolar amino acids in most of Kunitz inhibitors. The Arg residues at P1 and P3 has L-shaped conformation which possibly blocks the S1 and S4 pocket in similar way as of DX-9065a, a potent synthetic factor Xa inhibitor [24, 156]. This indicates that reactive site residues favor formation of stable complex with FXa due to its diverse nature which was not observed in other Kunitz STI inhibitors. Along with versatile and distinctive reactive site residues; formation of six salt bridges; expanded interactions of residues of β1, β4β5 and β7β8 loop residues of TKI with three exosites also called as secondary binding site (residues of acidic 36 loop, autolysis loop and sodium binding 220 loop) and perfectly suitable scaffold of TKI which fit well in FXa reactive site cavity (Fig. 4.12B) indicate that TKI is able to bind and can form steady complex with FXa. Mutagenesis studies will further enlighten their effects on inhibitory activity and will provide an approach to design an improved inhibitor in future which can act as anticoagulant in vivo.

TKI do not have any sequence identity with other factor Xa inhibitors except BuXI, which is Kunitz type factor Xa inhibitor. In fact, TKI shows highest sequence identity with BuXI among the Kunitz inhibitor but the reactive site residues are distinct. Two methionines at Met58 and Met66 (P'2) and Thr65 (P'1) might play important role in factor Xa inhibition in BuXI [4, 141] so the mechanism of inhibition pattern of FXa may be different for BuXI and TKI. Comparison of reactive site of TKI with other physiological or non physiological inhibitors of

FXa such as antithrombin (AT), protein Z-dependent protease inhibitor, tissue factor pathway inhibitor (TFPI), BuXI, antistasin and tick anticoagulant peptide shows that it does not have any similarity with reactive site of above factor Xa inhibitors except Arg at P1 (TAP and ZPI has Tyr at P1) and Arg at P3 same as TAP and antistasin. These results suggest that TKI has got distinctive reactive site which do not have much similarity with Kunitz type inhibitor as well as other factor Xa inhibitors.

The obvious question comes in mind that which residue of TKI gives selectivity towards FXa along with trypsin inhibition. To address this question, we compared the interacting residues of TKI with porcine trypsin and modeled complex with human factor Xa. Human FXa shows 37% sequence identity with porcine trypsin and has similar residues at S1 pocket except Ala190 instead of Ser190 in FXa. The most prime determinant for selectivity was observed at S4 pocket, TKI residues interact with four residues of S4 pocket which is most important binding region of FXa along with S1 [116, 156]. Arg64, which interacts with cation hole and it side chain occupies the aryl binding of S4 is the most important residue of TKI which may provide specificity towards FXa. Besides Arg64, Ser63, Ala73, Ser75, Ala118 and Lys136 may also interact with residues of S4 and may participate in construction of compact binding at S4 pocket. Such type of interactions was not observed in case of trypsin as '96-KETY-99' in FXa for specificity was unavailable (see Fig. 4.12C). Additionally, TKI reactive site and β1 loop residues are close enough to two exosites (acidic 36 loop and autolysis loop) which may form hydrogen bonds and salt bridges with residues of these loops (Fig. 4.12C). So conclusively, TKI residues are able to interact with S1, S4 pockets and two exosites of FXa and confer specificity for FXa along with trypsin.



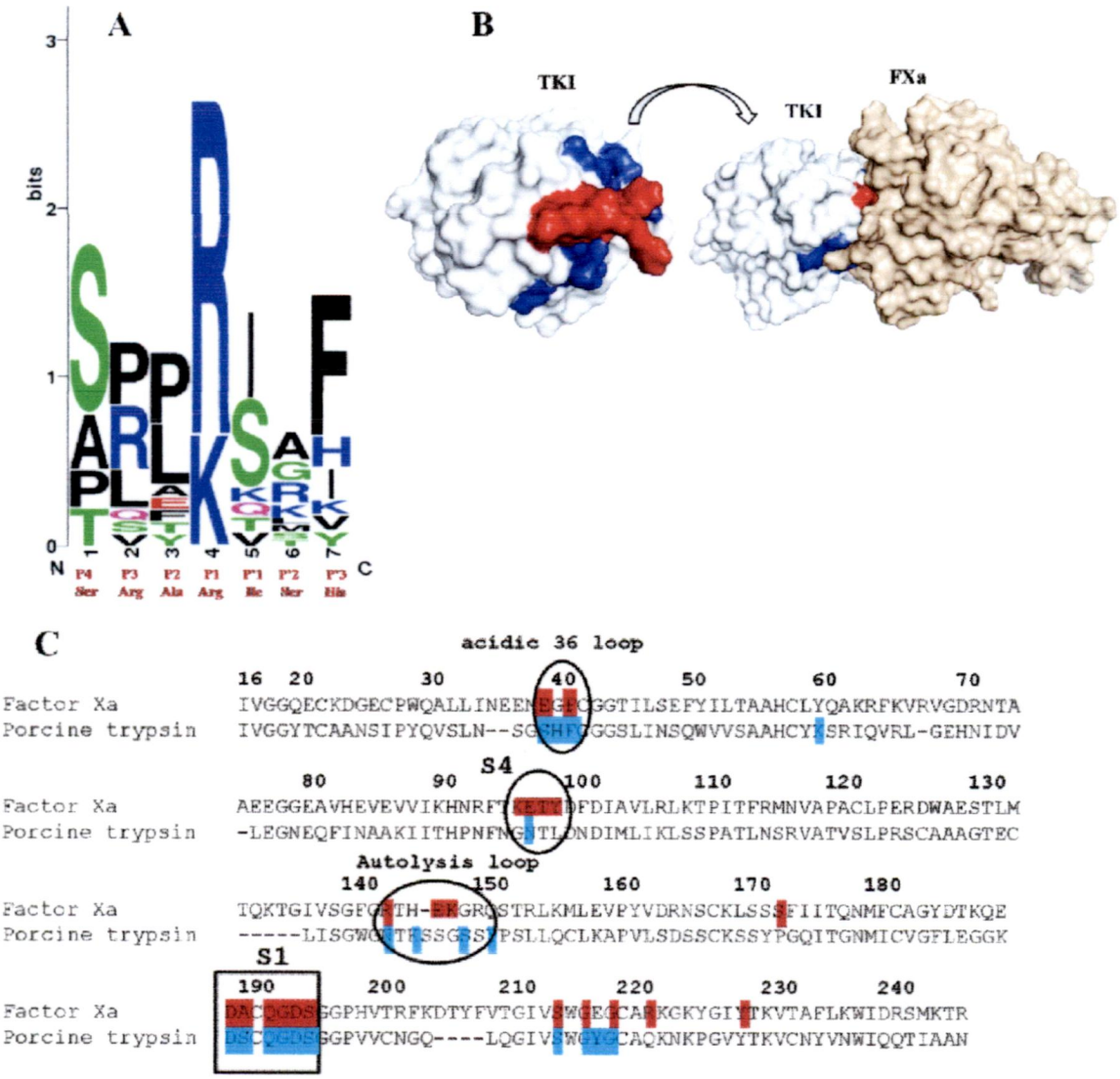


Figure 4.12. Versatility of reactive site residues and its interavtion with proteases. A. Web logo representation showing variation in residues reactive site in TKI as compared to other Kunitz STI type inhibitors [37]. The Y-axis represents sequence conservation in bits and X axis represent reactive site (P4-P'3). Respective reactive site residues of TKI are shown in red color. B. Scaffold of TKI is shown with interacting reactive site (red) and other residues (blue) of TKI with FXa. TKI reactive site enters in cavity of S1 pocket of FXa and forms stable complex. C. Interaction of residues of TKI with porcine trypsin and factor Xa are compared (chymotrypsin numbering). The specificity of TKI towards FXa may be obtained due to its interaction of TKI with S4 and two exosites namely autolysis and acidic 36 loop (shown in oval shape) and the difference in FXa and trypsin residues at this sites gives selectivity towards them. Aromatic S4 pocket formed by residues Tyr99 (Leu99), Phe174 (Gly174) and Trp215 of FXa which may form cation- π interaction with it substrate/inhibitor is absent in trypsin. S1 pocket is shown in square which has same residues except Ala190 in FXa instead of Ser190 of trypsin.

4.4. Discussion

The structure of TKI which display β trefoil fold with two disulphide bridges and an exposed reactive site loop having Arg at P1 follows the classical characteristics of Kunitz STI type inhibitors. Apart from these typical characteristics, TKI has insertion of residue Asn15 at the Kunitz signature motif which was clearly visible in electron density. To examine the interaction of TKI with proteases, we carried out formation and crystallization of TKI:PPT complex. The stable complex formed by TKI blocked the S1 pocket of PPT by its exposed reactive site and stabilized the complex by making interaction with the subsites of PPT. However, the interaction of P1 residues is similar as of STI:PPT complex but the pattern of stabilization is different than STI:PPT complex. The flexible residues of reactive site P1, P3 and P'3 become rigid after complex formation. As TKI also shows factor Xa inhibitory activity, to analyze the possible binding of TKI with factor Xa, molecular docking was performed. The Arg residues at P1 and P3 have L-shaped conformation which possibly blocks the S1 and S4 pocket of factor Xa. TKI could form six salt bridges with factor Xa, six residues of reactive site out of seven could interact with residues of FXa, interaction of Gln192 and Tyr99 which are important for specificity of FXa with residues of TKI and interaction of TKI residues at secondary binding site such as acidic 36 loop and autolysis loop along with S1 and S4 pocket of FXa confer stable complex formation with FXa. TKI residues interact with four residues of S4 pocket which is most important binding region of FXa along with S1 [116, 156] and the prime determinant for selectivity. Arg64, which interacts with cation hole and it side chain occupies the aryl binding site of S4, is the most important residue of TKI which may provide specificity towards FXa.

Structure of TKI:PPT complex and docking studies with FXa display the role of inserted residue, Asn15 in stabilization of complexes. Asn15 interact with four residues of PPT via water molecules and two residues of FXa and participate in stabilization of complex. Furthermore, it interacts with those residues which are important for specificity of FXa which point towards its role in specific recognition of proteases. Evaluation of reactive site residues shows that they do not have similarity with those of reactive sites of known Kunitz type inhibitors as well as physiological and non-physiological factor Xa inhibitor. This demonstrates that TKI has got novel reactive site residues and some of these residues are versatile. TKI is also reported as weak inhibitor of porcine pancreatic elastase [5] so it could be believed that the diverse and versatile reactive site of TKI is able to block PPE S1 pocket. It is unusual for PPE to take Arg at small S1 pocket even though, rBBCI mutant V66R also show activity towards PPE [65] which confirms

the assumption that TKI reactive site can also block S1 pocket of PPE. Taken together, TKI is not only restricted to inhibit trypsin but also FXa, PPE and probably other serine proteases through its versatile, unique reactive site and could play multifunctional roles.

CHAPTER 5

PURIFICATION, BIOCHEMICAL CHARACTERIZATION, CRYSTALLIZATION, STRUCTURE DETERMINATION OF TAMARIND CHITINASE LIKE LECTIN (TCLL) AND ITS COMPLEX WITH GlcNAc

5.1. Introduction

Carbohydrate binding proteins from plants are usually called lectins, agglutinins, or hemagglutinins. Plant lectins are heterogeneous and highly diverse class of non-immune origin (glyco) proteins because of their carbohydrate-binding specificity, differences in molecular structure and biochemical properties [39]. For specific recognition, lectins have at least one non-catalytic domain, which reversibly bind to sugars or glycans of glycoproteins and glycolipids and do not modify the structure of carbohydrates [39, 54]. Lectins were first discovered in plants and later also identified in organisms from all kingdoms of life [42]. Plant lectins are classified into twelve diverse families of evolutionary and structurally related lectin domains by Van Damme *et al.*, [204]. Whether these twelve families are a closed group or expandable is an open question. There is no precise description for the biological function of plant lectins because of their diverse classes and carbohydrate specificities. Lectins with homologous sequences also have different biological roles and their function cannot be generalized. Rigorous investigations need to be carried out for an understanding of the biological roles of individual lectins. Several functions for particular lectins reported include antifungal [87, 139, 221], antisecticidal [171, 207, 223], antiviral [8, 89, 221], antiproliferative and apoptosis-inducing [22, 27, 54, 112, 151, 177], symbiosis mediating between nitrogen fixing microorganisms and legume plants [15, 42]. Lectins shows extensive structural diversity and considering the structural folds, plant lectins are grouped into seven folds [29]. Previous reports showed that plant lectins are also members of chitinase family in fold while they might possess or lack chitinase activity [28, 203].

In an effort to gain further insight into structure-function relationship of lectins in general, A N-acetyl glucosamine binding chilectin from tamarind seeds has been named "TCLL" and investigated biochemically and structurally. In this chapter, the novel lectin; TCLL from seeds of tamarind (*Tamarindus indica*), a member of Leguminosae family, was purified and crystallized. We have determined a high resolution crystal structure of TCLL which displays conserved $(\beta\alpha)_8$ barrel fold typical of GH18 family but the active site residues for chitinase activity are substituted. On the other hand, biochemical data and the structure of TCLL complexed with GlcNAc disclose that TCLL is a lectin possessing novel binding site for the glycan. X-ray structure of free TCLL and its complex with GlcNAc suggests the protein to be a member of a new class of plant lectins as it does not show significant structural similarity to known lectins from the classification. The amino acid sequence of TCLL has been proposed using the combination of high resolution (1.4Å) crystal structure and by mass spectrometry. It,

shows scant similarity to known lectins and a distinct evolutionary relationship from them support a new classification as well. TCLL not only provides insight into evolution of lectins from class III chitinases but also structural affirmation that legume plants are capable of building a domain or site for specific identification of carbohydrate using existing scaffolds.

5.2. Materials and Methods

5.2.1. Materials

Affi-Gel Blue matrix, MacroPrep DEAE and MacroPrep CM (carboxymethyl) support, Molecular weight standards, protein assay kit (concentration estimation), were obtained from Biorad (Bio-Rad Laboratories, Hercules, CA, USA). PEG 8000, PEG3350 were purchased from Sigma-Aldrich Pvt. Ltd. All other chemicals were purchased from Himedia chemicals, India. Crystallization plates, initial crystallization screen reagent kits and cryoloops were obtained from Hampton Research, USA. Superdex- 75 GL 10/300 column and Gel Filtration Calibration LMW standards kit were obtained from GE Healthcare, USA.

5.2.2. Methods

5.2.2.1. Purification of TCLL

Tamarind seeds were obtained from fruit collected from plants growing locally. The fruit pulp was removed manually and the seeds were soaked in buffer A (100 mM Tris pH 7.4) for ~6 h at room temperature. A razor blade and a pair of forceps were used to remove the seed coat. The crude extract was prepared by homogenizing the seed kernels thus obtained in buffer A at room temperature using a blender. The prepared crude extract was then subjected to centrifugation at 14000g for 20 min at 277 K and the supernatant was collected. The obtained supernatant was applied onto an Affi-Gel Blue matrix (Bio-Rad Laboratories, Hercules, California, USA) pre-equilibrated with buffer A. The column was washed with buffer and eluted using a stepwise gradient of NaCl (0.1, 0.2, 0.3, 0.5, 0.8 and 1.0 M) in buffer A. A chitinase-activity assay was performed on each fraction according to the reported method using Calcofluor dye [186]. Fractions with chitinase activity were pooled, concentrated and dialyzed extensively against buffer A at 277 K using a 3.5 kDa dialysis membrane (Pierce, USA). The dialyzed sample was loaded onto a weak cation-exchange matrix [MacroPrep CM (carboxymethyl) support, BioRad Laboratories, Hercules, California, USA] pre-equilibrated with buffer A. The chitinase activity was detected in the unbound sample.

The unbound sample was subsequently loaded onto a weak anion-exchange matrix [MacroPrep DEAE (diethylaminoethyl), BioRad Laboratories, Hercules, California, USA] pre-equilibrated with buffer A. Protein elution was carried out using a stepwise gradient of NaCl (0.05, 0.1, 0.2, 0.3 and 0.5 M). The chitinase activity was detected in fractions eluted with 0.1 M NaCl. Eluted fractions were analyzed on 15% SDS–PAGE stained with Coomassie Brilliant Blue dye. Highly pure fractions containing chitinase activity were pooled and concentrated to a concentration of 14 mg/ml using Amicon Ultra-15 (10 000 Da cutoff, Millipore). Protein concentration was determined with the BioRad protein assay kit using BSA as a standard and the protein yield was consequently estimated to be 5 mg per gram of seeds. The N-terminal amino-acid sequence was obtained using the Edman degradation method on a Shimadzu automated protein sequencer (PPSQ-20).

5.2.2.2. Chitinase assay

The protein purified in 0.1 M Tris-HCl pH 7.4 was extensively dialyzed and subsequently buffer exchanged into sodium acetate 0.05 M pH 5. The chitinase assay was done according to the protocol given by Imoto and Yagishita [80]. Briefly, a 0.05% solution of glycol chitin was prepared in 0.1 M sodium acetate buffer pH 5 and incubated with increasing range (0-80 µg) of the enzyme at 40 °C for 30 min. This was followed by incubation with 0.2 ml of potassium ferricyanide (Potassium ferricyanide prepared in 0.5 M sodium carbonate) and the reaction was terminated by boiling at 90 °C for 15 min. Chitin hydrolysing activity was monitored at 420 nm.

The Calcofluor white dye based assay for chitinase was also performed [186]. For this assay, 0.4 ml of 0.05% glycol chitosan was incubated with 0.5 ml of protein solution and 0.6 ml of the sodium acetate buffer pH 5 for 1 h at 40 °C. To the control tube, all the constituents were added sequentially except the enzyme. The control and the experimental tubes were boiled simultaneously for 15 min and placed in cold water. To these tubes, 0.01% calcofluor white freshly prepared in distilled water was added and incubated for 1 h at room temperature. The calcofluor white binds to unhydrolysed chitin and gives a characteristic peak at 406 nm. The absorbance was monitored spectrophotometrically at 406 nm.

5.2.2.3. Carbohydrate estimation assay

The carbohydrate content of TCLL was quantitatively determined according to the phenol-sulphuric acid method [46] by plotting the standard curve using dextrose (0-25 μg). Experiment was done in triplicate and the percentage of carbohydrate content was estimated.

5.2.2.4. Determination of oligomerization state

To check oligomerization state, TCLL was loaded onto precalibrated Superdex- 75 GL 10/300 (GE Healthcare) size-exclusion column with gel filtration calibration LMW standards on an AKTA Purifier system (GE Healthcare). The column was equilibrated with 100 mM Tris buffer, pH 7.4 and 1 ml fractions were collected. Peak fractions were run on 15% SDS-PAGE under non-reducing condition.

5.2.2.5. Hemaglutination and hemaglutination inhibition assay

Fresh 1 ml of human blood sample of different blood groups (A, B, O) were collected in centrifuge tubes containing 6% EDTA as anticoagulant in 3 ml of phosphate buffer saline (PBS), pH 7.2 and erythrocytes were obtained. For hemagglutinating assay, 3% (v/v) suspension of human red blood cells in phosphate-buffered saline (pH 7.2) was prepared. Serial dilution of purified TCLL was taken in microtiter U-plates (50 μl) and was mixed with 50 μl of diluted erythrocytes cells. The plate was incubated at room temperature for 1 h. Specific activity was expressed as the minimum TCLL concentration (μg/ml) showing detectable hemaglutination activity. For the hemagglutination inhibition assay, N-acetyl glucosamine, N-acetyl galactosamine, cellobiose, carboxymethylcellulose, dextrose, glucose and galactose were used. Final concentrations of the additives were maintained at 10 mM before addition of red blood cells. The assay was monitored under microscope.

5.2.2.6. Fluorescence quenching assays

Binding of the monosaccharides to TCLL was analyzed by ligand induced quenching of intrinsic tryptophan fluorescence. Fluorescence measurements were carried out in a Varian Cary Eclipse fluorescence spectrometer. Emission spectra were recorded from 291- 500 nm upon excitation at 290 nm. Excitation and emission slits were maintained at 5 nm and the scan speed was set at 100 nm/min. Standard reaction mixtures were prepared using 0.5 μM solution of protein in 25 mM Phosphate buffer saline, pH 7.2 to a final volume of 1 ml. To study the binding

affinity of monosaccharides to the enzyme, the additives were pre-incubated for 20 min and then spectra were obtained. The spectrum was corrected for background emission generated by buffer and ligands and repeated in duplicates.

5.2.2.7. Crystallization of TCLL

Crystallization was performed by the sitting-drop vapour-diffusion method in 96-well crystallization plates (Hampton Research) at 293 K. Drops were prepared by mixing 2 µl protein solution in buffer A with the same volume of reservoir solution. The final protein concentration in the drops was 7 mg/ml. Initial crystallization conditions were obtained using Crystal Screen (Hampton Research, USA). To obtain higher resolution crystals, various permutations and combinations of crystallization conditions were explored. Crystal with maximum dimension was obtained within three days at 20 °C using the precipitant solution containing 20 mM calcium chloride, 100 mM sodium acetate (pH 5) and 10-30% MPD with the drop size of 1:1 and 1:2 with the reservoir buffer. The crystals of TCLL were directly flash frozen in liquid nitrogen. For the complex generation, a soaking solution was prepared containing 20 mM calcium chloride, 100 mM sodium acetate (pH 5), 30% MPD and 20 mM N-acetyl glucosamine (GlcNAc). Complex was obtained by incubating TCLL crystals in soaking solution for about 20 min at room temperature. Different sugars were soaked in similar manner with TCLL crystals but all trials were unsuccessful.

5.2.2.8. Data processing and refinement

X-ray diffraction data were collected at 1.49Å for free TCLL and 1.6Å TCLL complex with GlcNAc at home source. Indexing, integration and scaling of the diffraction data were carried out in HKL2000 [146] and are summarized in Table 1. The structure of TCLL was solved by molecular replacement method using hevamine as a template model (PDB code: 2HVM). The transformed coordinates were subjected to 50 cycles of rigid body refinement followed by restrained refinement with refmac 5.2 from the CCP4i package [33, 135]. The manual model building was carried out with graphics programs COOT [49, 50]. The Fo - Fc difference Fourier map generated indicated the presence of one MPD, a sodium acetate ion and a glycosylated moiety in the free TCLL crystal. The complexed TCLL structure indicated presence of one additional GlcNAc moiety. The three dimensional models for MPD, acetate and GlcNAc were generated from PRODRG [175]. The positions of water molecules were determined using Arp/ warp solvent program from CCP4i suite. After several cycles of rigid-body refinement followed

by restrained refinement, acceptable R_{cryst} and R_{free} were obtained. The models were evaluated using the program PROCHECK [106]. Figures were generated using the program PyMOL [44] and ESPRIPT program [61].

5.2.2.9. Analysis of crystallographic TCLL sequence

Primary amino acid sequence of TCLL was proposed from the crystal structure solved at 1.49Å. The programe BLAST was used for detecting similar sequences available in database [2]. Multiple sequence alignment was performed using CLUSTAL W [196] with default parameters. Proposed sequence was analyzed for N-linked and O-linked glycosylation site using the Oxford Protein Analysis Linker (OPAL) available publicly through the Oxford Protein Production Facility UK (OPPF) Website http://www.oppf.ox.ac.uk/opal/

5.2.2.10. Internal sequence determination using mass spectrometry

For confirmation of amino acid sequence obtained by X-ray crystallography, LC MALDI analysis of trypsin and Glu-C (V-8) endopeptidase digested TCLL protein was performed. Peptides obtained by proteolytic digestion were separated on C-18 column and subjected to MALDI fragmentation for MS/MS sequence determination. Coomassie blue-stained protein bands were excised from the corresponding SDS-polyacrylamide gels and destained with 50% acetonitrile in 50 mM ammonium bicarbonate (ABC) for 1 h with three intermittent changes of the solution. The supernatants were replaced with 10 mM DTT (Sigma) in 50 mM ABC solution to reduce the proteins for 15 min at 56 °C. The supernatant was discarded and 20 mM iodoacetamide (Sigma) in 50 mM ABC was added for 15 min at 25 °C in the dark to alkylate the proteins. The gel pieces were collected, washed three times with 200 µl of 50 mM ABC for 15 min at RT and then dried in vacuo. The gel pieces were rehydrated with 3.0 µl of 12.5 ng/µl of sequencing grade modified trypsin gold (Promega) and 3 µl of 50 ng of V8 protease in 50 mM ABC in two separate reactions and incubated for 60 min at 25 °C. Further 50 µl of 10% acetonitrile in 50 mM ammonium bicarbonate was added and the digestion of proteins was continued for 18 h at 37 °C with agitation. The supernatant was collected and the gel pieces were extracted for 40 min at 37 °C with successive 50 µl aliquots of 0.1% trifluoroacetic acid and 50% acetonitrile. The combined extracts were concentrated using Speed Vac.

The peptide extracts were reconstituted in 40 µl of 98% water, 2% acetonitrile and 0.1% TFA. The aliquot of 12 µl of digested peptide was separated on Chromolith CapRod Monolithic

capillary column (150 mm × 0.1 mm RP-18 endcapped). The LC gradient for the separation of the peptides was prepared from two solvent systems: solvent A (98% water, 2% acetonitrile and 0.1% trifloroacetic acid) and Solvent B (98% acetonitrile, 2% water and 0.1% trifluoroacetic acid). Trypsin and V8 protease digested peptides were desalted with solvent A for 30 min on Cap-trap C18 column at a flow rate of 20 µl in the reverse direction. After desalting, the peptides were eluted from the column over 60 min using gradients of solvent A and B in following steps: 5-10% B in 7 min, 10-35% B in 18 min, 35-50% B in 10 min, 50-90% B in 2 min followed by a washing step for 5 min at same gradient. Most of the peptides eluted between 10 min to 30 min. All the eluted peptides were simultaneously mixed with α-cyano-4-hydroxycinnamic acid (CHCA) matrix at a ratio of 1:1 and spotted after every 7 sec on LC MALDI plate. The flow rate of the solvent was 1.2 µl /min.

The LC MALDI plates spotted with peptides as above were analyzed on AB Sciex 4800 Plus TOF/TOF analyzer in reflector ion mode. To identify the peptide and its PTM modification peptide masses and their MS/MS sequences derived from the mass spectrometric analysis were searched for matches using the Protein Pilot 2.0 software against given X-ray crystal structure sequence. In the Protein Pilot search, biological modification and amino acid substitution was incorporated as ID focus factor along with the gel based ID search. Detected protein threshold was fixed at confidence score of 99.9%.

5.2.2.11. Phylogenetic analysis

Phylogenetic tree of TCLL was constructed using MEGA version 5 program using the Neighbor-Joining method. The evolutionary distances were computed using the Poisson correction method [193].

5.3. Results

5.3.1. Purification of TCLL

TCLL was purified to homogeneity by a three-step chromatographic procedure involving a combination of affinity and ion-exchange chromatography. The enzyme activity was analyzed after each step of purification by performing chitinase assays showing very negligible activity. In the first step, Affi-Gel Blue matrix was used and fractions were eluted using a stepwise gradient of NaCl. Protein with chitinase activity was eluted in the 0.5 M NaCl fraction. The eluted fraction was dialyzed to remove salt and applied onto CM matrix. The unbound fraction of the CM matrix contained the protein as indicated by the chitinase assay and was loaded onto DEAE matrix. The fraction eluted from the DEAE column with 0.1 M NaCl was highly pure (Fig. 5.1) and showed chitinase activity. The purity of the sample was confirmed by the presence of a single band on a 15% SDS–PAGE stained with Coomassie Brilliant Blue dye (Fig. 5.2). The molecular weight of the purified TCLL was calculated by running protein molecular-weight markers along with the purified protein on SDS–PAGE. The relative mobility of the protein markers was plotted against log molecular weight. The molecular weight obtained was ~34000 Da. Gel filtration analysis also showed that TCLL is a monomeric ~34000 Da protein.

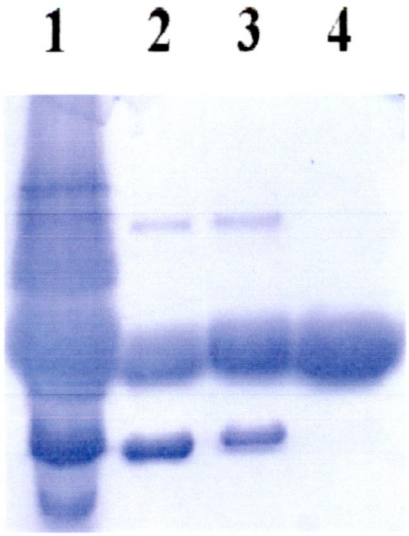


Figure 5.1. Elution profile of TCLL on SDS-PAGE. Lane 1: crude extract; Lane 2: Affi-gel blue (0.5 M elute); Lane 3: CM flow through; Lane 4: pure TCLL on DEAE elution (0.1 M elute).

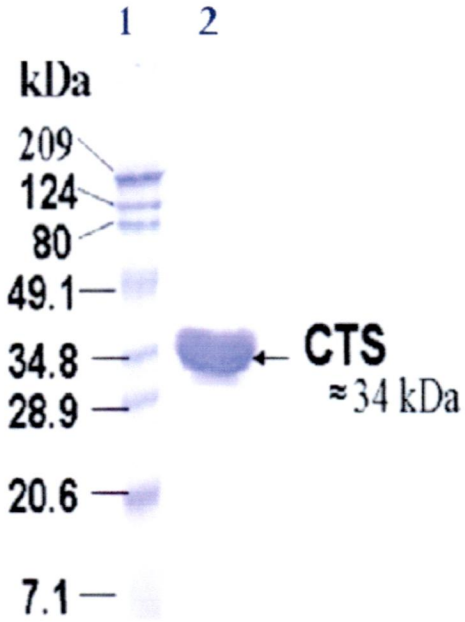


Figure 5.2. 15% SDS–PAGE analysis of TCLL. Lane 1: protein markers and Lane 2: purified TCLL.

5.3.2. Hemaglutination and fluorescence quenching assays

TCLL was tested for its hemaglutinating activity. TCLL shows lectin-like activity on human erythrocytes from three blood groups (A, B, O). TCLL shows lattice formation of the erythrocytes of all the blood groups with specific activity of ~45 μg/ml examined under microscope. The formation of lattice was inhibited by 10 mM GlcNAc but not with other sugars tested. Further, binding studies of the sugar moieties was determined by exploiting the intrinsic tryptophan fluorescence property of the protein. It was observed that addition of GlcNAc (1-20 mM) to a solution of protein resulted in quenching of the fluorescence between 310-320 nm without any shift of wavelength emission maxima (Fig. 5.3). The fluorescence quenching occurred till 10 mM GlcNAc and beyond this concentration there was no detectable change in the spectra (Fig. 5.3). No other sugar showed any noteworthy change in fluorescent intensity indicating that TCLL has affinity specifically for GlcNAc moiety.

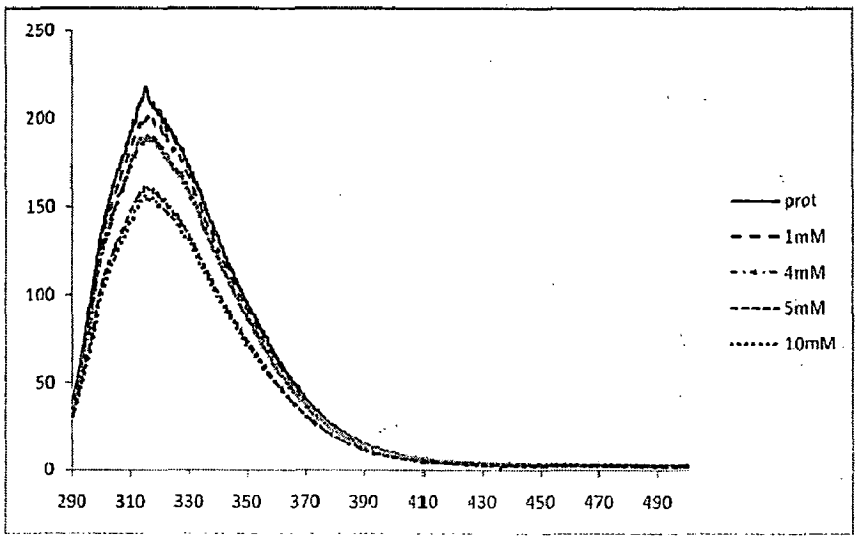


Figure 5.3. Effect of GlcNAc on the intrinsic fluorescence of TCLL. Addition of GlcNAc to 0.5 μM solution of protein in 25 mM Phosphate buffer saline, pH 7.2 at increasing concentration resulted in quenching of spectra. The emission spectra were recorded between 291-500 nm upon excitation at 290 nm. No shift in wavelength maxima from 316 nm was observed.

5.3.3. Quality of model

Free TCLL structure was solved at 1.49Å resolution with primitive tetragonal space group $P4_32_12$ with cell dimensions of $a = b = 100.096$, $c = 81.713$ Å by molecular replacement method using hevamine (2HVM) [205] as a search model. The crystals have one molecule in an asymmetric unit. TCLL structure was refined with acceptable stereochemistry and final R_{cryst} and R_{free} values of 19.4% and 20.3%, respectively (Table 5.1). The final model showed coordinates for residues from 9 to 187 and 189 to 272 and lacked residue 188 due to inappropriate electron density. A total of 514 water molecules were identified in the model. Structure of TCLL in complex with GlcNAc was determined at 1.6Å having same space group as free TCLL. The model was refined to final R_{cryst} and R_{free} of 16.5% and 19.8%, respectively. GlcNAc moiety was clearly visible and it did not alter the overall conformation of the protein compared to free structure. The crystallographic data and refinement statistics for free and complex structures are summarized in Table 5.1.

	TCLL	TCLL complex with GlcNAc
Crystallographic data		
Space group	$P4_32_12$	$P4_32_12$
Wavelength	1.54179	1.54179
Resolution	50-1.49	50-1.6
Cell dimensions		
a (Å)	100.09	100.09
b (Å)	100.09	100.09
c (Å)	81.71	81.84
Unique reflections	67001	54128
Completeness (%) (Last shell)	98.2 (70.8)	97.8 (56.3)
R_{sym}(%)[a] (Last Shell)	4.3 (52.9)	3.4 (38.4)
I/ σ (Last shell)	21.1 (2.0)	49.0 (3.0)
Multiplicity (Last shell)	5.5 (3.7)	7.7 (5.3)
Refinement	20-1.49	20-1.6
No. of reflections (working/test)	63044 (59689/3355)	51052 (48315/2737)
No. of residues	263	265
Water molecules	514	477
Resolution range (Å)	20-1.49	20- 1.61
R_{cryst} (%)	19.3	16.3
R_{free} (%)	20.3	19.8
Average *B*-factors (Å^2)	16.7	19.9
Water atoms	42.7	41.6
All atoms	21.8	24.3
rmsd on bond lengths (Å)	0.006	0.04
rmsd on bond angles (Å)	1.05	3.1
Ramachandran plot (%)		
Preferred	89.9	89.2
Allowed	10.1	10.8
Outliers	0.0	0.0

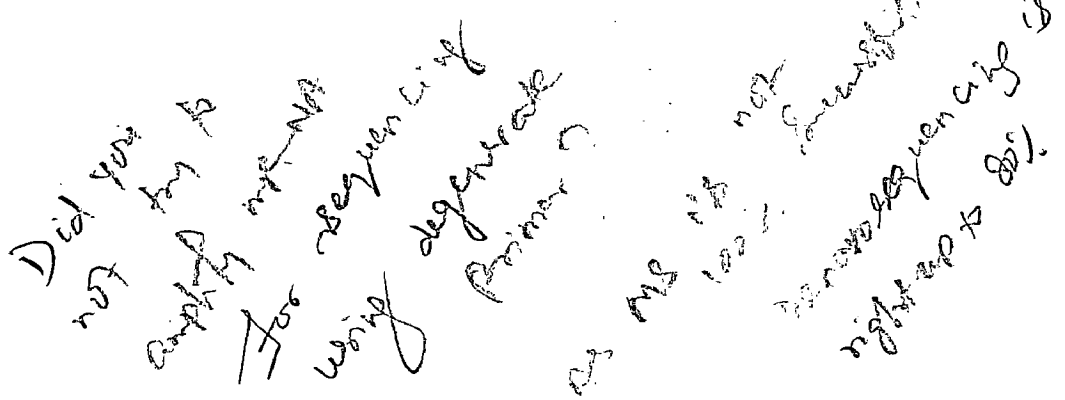
$$R_{sym} = \sum_{hkl}\sum_{i=1}^{n}\left|I_{hkl,i} - \bar{I}_{hkl}\right| \Big/ \sum_{hkl}\sum_{i=1}^{n} I_{hkl,i}$$

Table 5.1. Crystallographic data and refinement statistics for free TCLL and complex with GlcNAc.

5.3.4. Sequence analysis

TCLL sequence is proposed from crystal structure of TCLL solved at 1.4Å and by mass spectrometry. Protease digested peptides were subjected to MALDI TOF TOF analysis and the crystallographic amino acid sequence was used as search template for the MS/MS sequence. Typical MS/MS spectra and the consequent sequence of representative peptides are shown in Fig. 5.4. Sequence coverage of ~73% was obtained when results from both the proteases were combined (Fig. 5.5). Post-translational modifications may be responsible for lower sequence coverage. No major differences were observed in the mass spectrometry based amino acid sequence and from X-ray crystallographic sequence.

TCLL shows significant sequence similarity to the reported class III chitinase GH18 hevamine in the pdb database using NCBI BLAST. The proposed amino acid sequence of TCLL, however, does not show any signature motif of family 18 chitinases which comprise of a sequence DXXDXDXE/ [LIVMFY]-[DN]-G-[LIVMF]-[DN]-[LIVMF]-[DN]-x-E. This is due to natural variation in the signature motif in TCLL. A protein sequence BLAST against pdb database showed that TCLL shares sequence identity of 36%, 35%, 34%, 33% and 33% with hevamine (2HVM) from latex of *Hevea brasiliensis* [205], PPL2 (2GSJ) from *Parkia platycephala* seeds [28], concanavalin B (1CNV) from *Canavalia ensiformis* [174], xylanase inhibitor protein I (XIP-I) (1OM0) from *Triticum aestivum* [148] and xylanase and alpha-amylase inhibitor protein (XAIP) (3MU7) from *Scadoxus multiflorus* [98], respectively. It also shows 25% sequence identity with fungal "plant-type" family 18 chitinase, scctsl from *Saccharomyces cerevisiae* (2UY2) [79] (Fig. 5.6). Phylogenetic analysis also shows that TCLL is closely related to class III chitinases from plants of GH18 family and distinctly related to other chilectins from animals of GH18 (Fig. 5.7). Proposed TCLL sequence analyzed for N-linked glycosylation, predicted 3 glycosylation sequon starting at positions 54 (NCT), 92 (NLS) and 146 (NSS) (Fig. 5.8). One O-linked glycosylation site was also predicted at position 90 (Fig. 5.9).



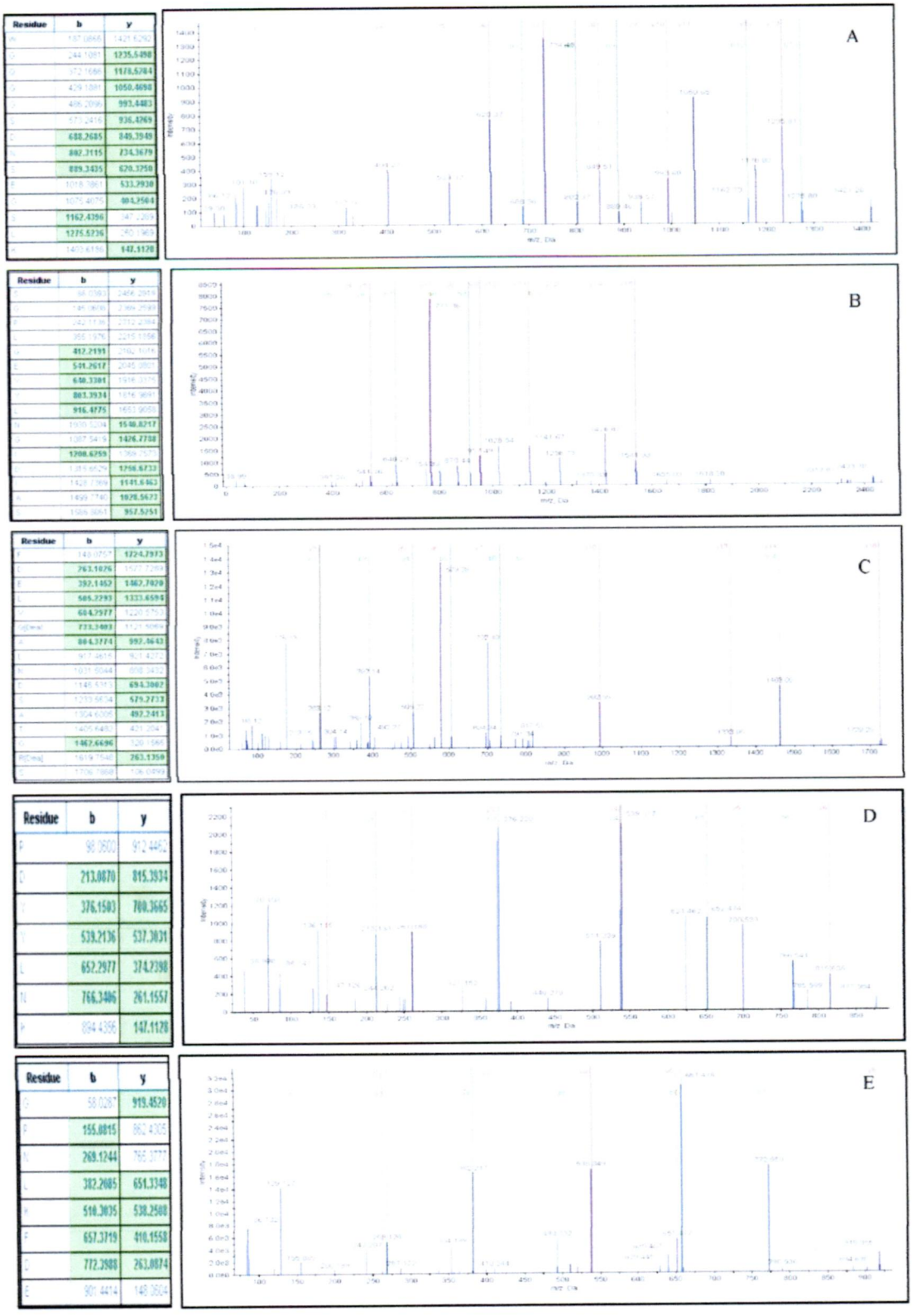


Figure 5.4. Typical MS/MS spectra and the consequent sequence of representative peptides of TCLL

CHT TRYPSIN 66.9% RED
CHT GLU-C 23.9%
BOTH COMBINED 197/272= 72.42%
MUTATIONS UNDERLINED

GSLKEACKTGHYDQIVLEFLITYDNGRDPDLNLGAH
CVNCTSLQSEIKYCQSKLIKILIQIGQVIPTKSLIKDLIIKDLSQYLDSNFFSGKSGPL-
GLVQALNDSATGRRIYLSASPNCVYPDYYLDKLDFLFVQFFYGL
PCTYTQGLPEDLEQAMKTWTSNVPE DLNGYIPPRVLNKAILPAVTQASNFAGVMIIDRYFD
RFRKYSSKIKR

Contrib	Conf	Sequence	Modifications	\Mass	Prec MW	Spectrum
2.00	99	GLNLKFDE	Leu->Pro *a* 2	-0.0332	918.4116	1.CHT GLU C.7.73.1
2.00	99	FDELVQALNDSATGRR	Arg->Ser *a* 16	-0.0582	1723.7318	1.CHTTrypsin.7.108.1
2.00	99	PDYYLDK	Asp->Asn *a* 6	-0.0038	911.43	1.CHTTrypsin.7.131.3
2.00	99	SGPLGEVYLDGIDIASVPEGLNLK	Asp->Asn *a* 10	-0.0439	2455.2407	1.CHTTrypsin.7.127.2

Figure 5.5. Proposed amino acids sequence of TCLL which was obtained from crystal structure of TCLL solved at 1.4Å and confirmed by mass spectroscopy.

Figure 5.6. Sequence alignment of TCLL with other highest matching members of GH18 family such as hevamine, PPL2, concanavalin B, XIP-I, XAIP and SCCTS1. The conserved residues are represented in red background. The key active site residues for chitinase activity are shown by green arrows.

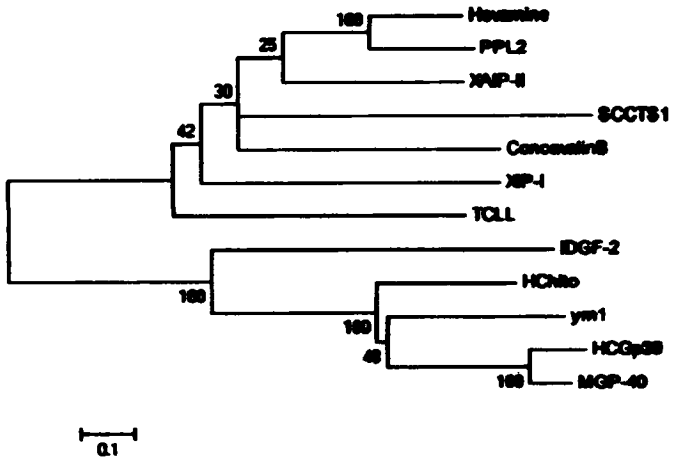


Figure 5.7. Phylogenetic analysis of TCLL was created using MEGA5 program.

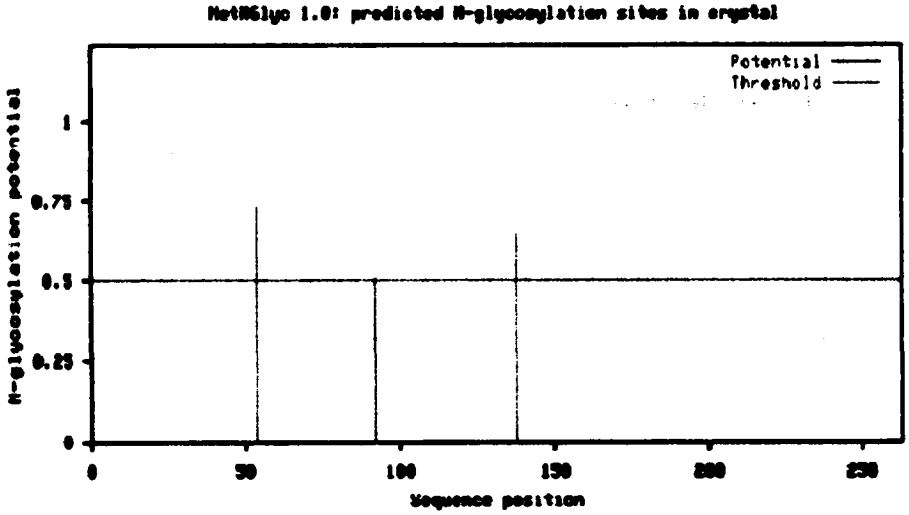


Figure 5.8. Predicted potential N–glycosylation sites in the sequence of TCLL by NetNGlyc 1.0

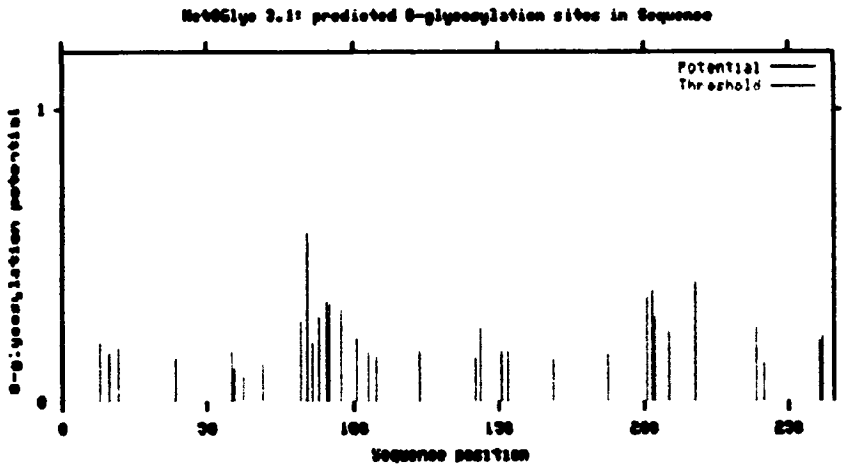


Figure 5.9. Predicted potential O–glycosylation sites in the sequence of TCLL by NetOGlyc 3.1

5.3.5. Overall structure of TCLL

The overall structure of free TCLL which comprises of a single polypeptide chain consists of 266 amino acids, MPD, GlcNAc and sodium acetate. The TCLL model has acceptable overall geometry (Table 5.1) and no residues in disallowed region of the Ramachandran plot. The structure displays a $(\beta\alpha)_8$ barrel topology as observed in the other GH18 chitinase family members [28, 174, 205]. Fig. 5.10 illustrates nomenclature of α-helices and β-strands including the additional strands. Loops connecting carboxy terminus of strand to amino terminus of helix (8-12 amino acids long) and carboxy terminus of the helix to amino terminus of strand (4-8 amino acids long) were termed as βxαx and αxβx+1 loops, respectively, to maintain consistency in the nomenclature. The $(\beta\alpha)_8$ barrel is made up of eight core parallel β sheets (residues 12-14 (β1), 37-40 (β2), 79-84 (β3), 124-127 (β4), 155-158 (β5), 180-184 (β6), 216-222 (β7), 252-256 (β8)) surrounded by eight α-helices (26-32 (α1), 64-76 (α2), 97-101 (α3), 138-146 (α4), 170-175 (α5), 198-211 (α6), 233-246 (α7), 266-269 (α8)). Additionally, some distinct structural features in TCLL are: one 23 amino acids long loop, β2α2, which contain an antiparallel β-hairpin (β2' (43-46) & β2'' (53-55)) stabilized by three hydrogen bonds. This extended loop is unusual for the $(\beta\alpha)_8$ topology. (2) Another loop β3α3 also contains insertion of β-sheet termed as β3' (86-88) which makes parallel β-sheet conformation with β2' of loop β2α2 (Fig. 5.10) and this conformation is stabilized by two hydrogen bonds. Typical βαβ folding pattern is disrupted by additional short α-helix α1' (20-24) & α8' (258-264) (Fig. 5.10). The insertion of additional α-helix was observed in several $(\beta\alpha)_8$ barrel topologies [53, 174].

Three disulphide bonds (Cys30-Cys73, Cys60-Cys63 and Cys162-Cys191) were observed which are also present in hevamine, PPL2 and concanavalin B structures. These disulfide bridges help in maintaining the integrity of the enzyme. The helices α1 and α2 are connected by Cys30–Cys73 bridge, loops β5α5 and β6α6 by Cys162–Cys191 whereas Cys60-Cys63 bridge forms hairpin like structure in the loop β2α2. TCLL structure contains three well defined cis-peptides of which one is cis-proline (Tyr164-Pro165) and the other two cis-peptides are Glu41-Glu42 and Phe256-Asp257. Glycosylation was also observed clearly in electron density with N-acetyl glucosamine binding to Asn146, one of the predicted glycosylation sites (Fig. 5.11). TCLL glycosylation was also confirmed by phenol-sulphuric acid method (data not shown). Glycosylation in plant GH18 chitinases of class III is very rare. Chitinase from suspension-cultured bamboo [101] and xylanase inhibitor protein I (XIP-I) [148] from wheat belonging to this family shows glycosylation. The structure of XIP-I (pdb id: 1OM0) shows two

glycosylation sites at Asn89 and Asn265. There are some reports on glycosylation of mammalian chitinase like proteins (MGP40, HCGP-39) [76, 130] and *Drosophila melanogaster* protein (IDGF-2) [208] of GH18 family.

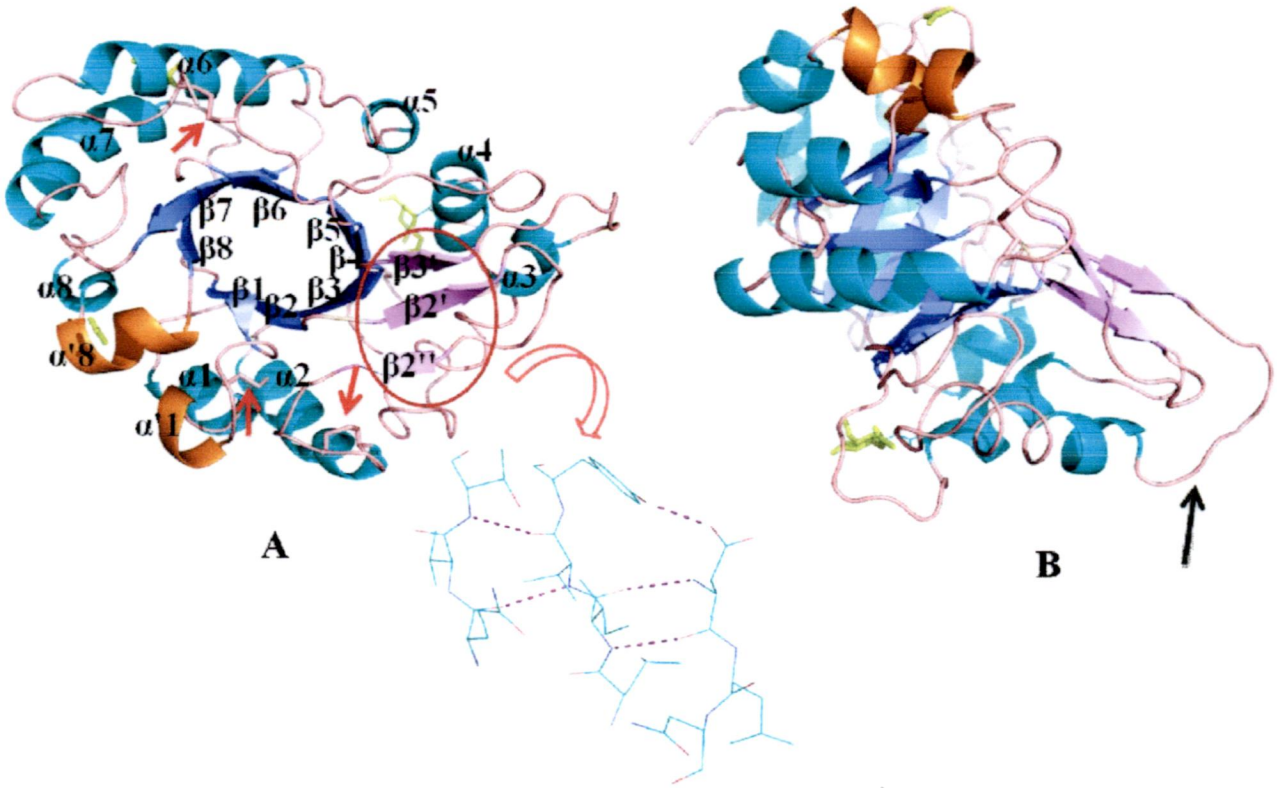


Figure 5.10. Cartoon diagram representation of TCLL in (A) top view orientation and (B) side view orientation. The α-helices (α1 to α8) are shown in cyan and extra helices (α1 and α8) in orange. The β-strands (β1 to β8) are shown in TV blue and extra strands (β2', β2'' and β3') in violet. Connecting loops are in wheat color. GlcNAc which is N-glycosylated at Asn146, MPD and sodium acetate are shown in green color. Three disulphide bonds are indicated by red arrows. Unusual loop that protrudes out from domain are shown by black arrow in (B). The hydrogen bond network formed by extra strands is highlighted.

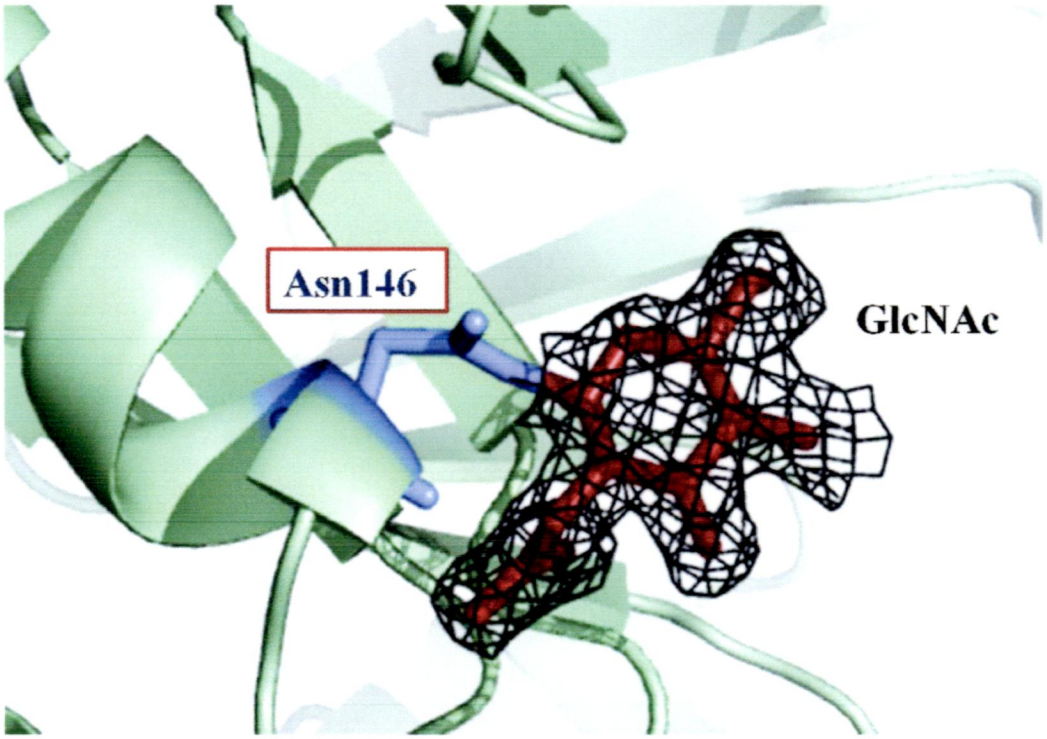


Figure 5.11. N-glycosylatation site in TCLL. Glycosylation was observed clearly in electron density with N-acetyl glucosamine (GlcNAc) binding to Asn146 which is one of the predicted glycosylation site from the sequence of TCLL.

5.3.6. N-acetyl glucosamine binding site

TCLL crystals soaked in mother liquor containing 20 mM N-acetyl glucosamine upto 20 min displayed clear interpretable electron density for a single molecule of N-acetyl glucosamine. In the complex structure, the GlcNAc binding site is located in a shallow groove surrounded by two loops, β4α4 and β5α5. The GlcNAc molecule accommodated well in the pocket and was delimited by mostly polar amino acids. The pocket lines three aromatic amino acids Phe132, Tyr161 and Tyr162. Six water molecules in the cavity form hydrogen bonds with GlcNAc moiety, which is stabilized by interactions with protein residues. The most important interactions of GlcNAc are with Glu126 and Tyr161. O4 and O5 of GlcNAc are engaged in hydrogen bond with carbonyl O and OH of Tyr161. O3 and O7 of GlcNAc interact with N and carbonyl O of Glu126. O3 and O4 of GlcNAc interact with Val124, Glu126 and Tyr162 through W38. O6 is hydrogen bonded with Lys165 by W189. N2 and O7 interact with W148, which is further stabilized by making interaction with carbonyl O of Gly127. A water molecule, W93, seen in contact with O4 of GlcNAc mediates the hydrogen bond with OE2 of Glu126. Detailed interactions of pocket residues with GlcNAc are shown in Fig. 5.12 and Table 5.2. Structural comparison between the free and the GlcNAc bound complex indicate that TCLL does not display any major conformational change upon binding of GlcNAc. Superposition of the complex structure and the free structure provides an r.m.s. deviation of 0.1Å (244 to 244 atoms).

5.3.7. Substrate binding channel and active site architecture for chitinase activity

Since TCLL showed very negligible chitinase activity and attempts to get crystals of the protein in complex with chitotrise and chitotetraose were unsuccessful, the substrate binding site was evaluated and compared with the other GH18 family members. These subsite grooves are formed in TCLL by joining the carboxyl terminal residues of the β-strands with their subsequent loops. Compared to hevamine, the architecture of the subsites are different with some residues that interact with oligosachharide being substituted. The subsite -4 in TCLL formed by loops β1α1 and β2α2 contains Gly18, Asn55 and Ala58 similar to hevamine residues Gly11, Asn45 and Gly48 indicating that this subsite in TCLL might accommodate the substrate. However, immediately above Gly18, one extra helix α1' (Asp20-Glu23) exists which might hamper the entry of chitooligosaccharide. The side chains of Ser19 and Asp20 protrude in the pocket with Gly18 hidden beneath Ser19 so that it can fail to attain the N-acetyl glucosamine moiety. The subsite -3 is present on the same loops as -4 has and include Gln16 and Asn55 (Gln9 and Asn45

in hevamine). Two striking differences were observed in the TCLL -4 and -3 subsites as compared to hevamine: one is that these two subsites are very close to each other (The Cα distance between Gly18 and Gln16 is 5.5Å compared to 6.6Å of corresponding hevamine residues) and second that the cavity formed by loops is wider than hevamine so probably it will not grip two GlcNAc moieties. As for subsite -2, it is the most disturbed site. The loop forming this groove - β3α3 - is very different in TIM barrel topology with extra β-strand (β3'). It protrudes straight away from core barrel rather than forming pocket and looks like flung out from $(\beta\alpha)_8$ fold. However, in hevamine strand β3 moves little inside in the barrel and the corresponding loop is projected in the groove and arranged in such a way that it holds chitin molecule by interacting with it. The residues Val87 and Thr88 on one side of the groove on the extra β3 strand and Phe256 on the opposite side of the groove from this subsite in TCLL gets substituted (natural variants) as compared to hevamine Gly81, Ile82 and Trp255 residues. The Cα distance between Val87 and Phe256 is 21.4Å, which is very high as compared to that in hevamine (11.9Å) and forms very shallow and broad cavity. The residue Thr88 participates in hydrogen bond interaction with Val45 of strand β2' which is parallel to β3' and is involved in stabilization of strand rather than participation in accepting GlcNAc moiety. Thr88 is also engaged in hydrogen bonding with Asp47 so that the overall this groove is completely unsuitable and may not hold chitin molecule.

The most crucial subsite -1 contains Glu184, Arg187, Lys227 and Val87 that correspond to Gln181, Asn184, Ala224 and Gly81 of hevamine. Arg187 protrudes in the cavity and reduces the ability to bind GlcNAc. The architecture of the loop that contains Ala224 in hevamine is in such a way that it bends toward the core barrel and is involved in making cavity while Ala224 form hydrogen bond with O6 of GlcNAc moiety. This type of interaction is not possible in the TCLL structure because the loop moves away from barrel and the residue Lys227 is not able to interact with O6 of GlcNAc and also has rigid β-turn conformation. The Cα distance between Val87 and Lys224 is 25.5Å, whereas the corresponding distance in hevamine is 12.3Å. Overall, this subsite is also not favorable for binding to chitin molecule. The key residues for catalytic activity are Asp125, Glu127 and Tyr183 in hevamine. Glu127 is considered to donate a proton to glycosidic bond [21, 194, 195]. Asp125 stabililizes the transition state during hydrolysis and Tyr183 assist Asp125 in making hydrogen bond to carbonyl oxygen of the N-acetyl group. It has been shown that mutation of the very important catalytic residue Glutamate (Glu127) in hevamine impaired the chitinase activity [21, 215]. The respective residues in TCLL are Ala128,

Val130 and Phe186. Neither Ala128 nor Phe186 side chain is able to make hydrogen bond with GlcNAc residue to stabilize the oxazolium intermediate. The most important residue Glutamate is substituted with Valine which cannot donate a proton to the scissile bond. The loop that bears Ala128 and Val130 has different orientation than the corresponding loop in hevamine. Corresponding residues of the active site and substrate binding site of TCLL and hevamine are shown in Fig. 5.13. The differences explained here in TCLL structure rules out any possibility of chitin binding as well as chitinase activity.

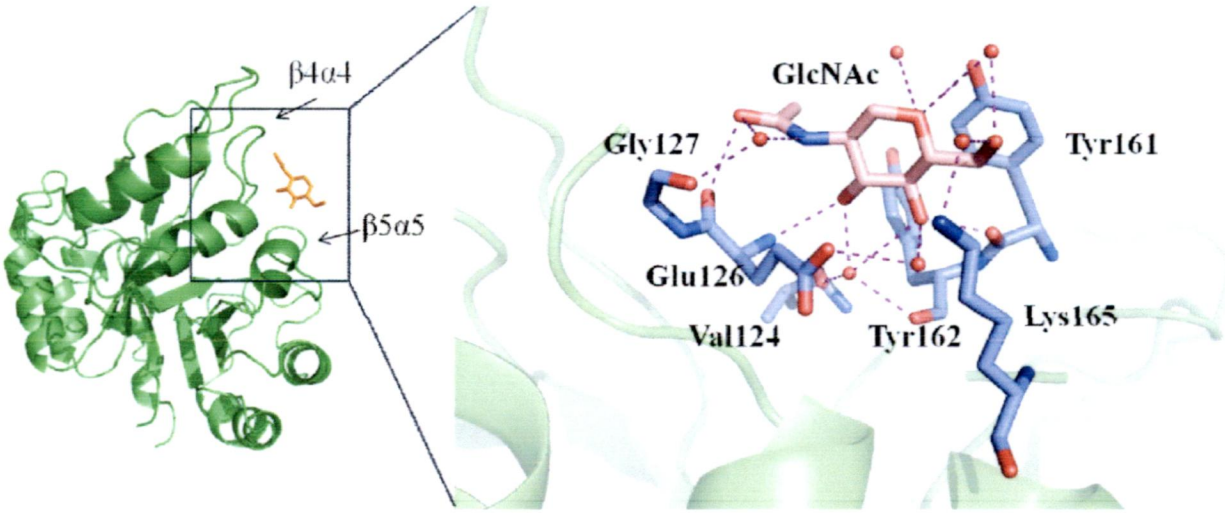


Figure 5.12. Overall TCLL structure with GlcNAc bound in the pocket formed by loops β4α4 and β5α5. GlcNAc binding site is focused and interactions are shown. GlcNAc is stabilized through hydrogen bonds with residues Val124, Glu126, Gly127, Tyr161, Tyr162, Arg165 and six water molecules in the pocket. Hydrogen bonds are shown in deep purple dotted lines, water molecules in red spheres, GlcNAc is shown as stick with orange carbon. Residues interacting with GlcNAc are shown as stick with blue carbons.

GlcNAc atoms	TCLL	Water	TCLL atoms
O3	Glu126 N		
		W38	Val 124 O
O4		W38	Tyr162 O
	Tyr161 O		
		W93	Glu126 OE2
O5	Tyr161 OH		
		W226	
		W301	
O6		W301	
		W189	Arg165 NZ
O7	Glu126 O		
		W148	Gly127 O
N2		W148	

Table 5.2. Hydrogen bond network of GlcNAc with TCLL

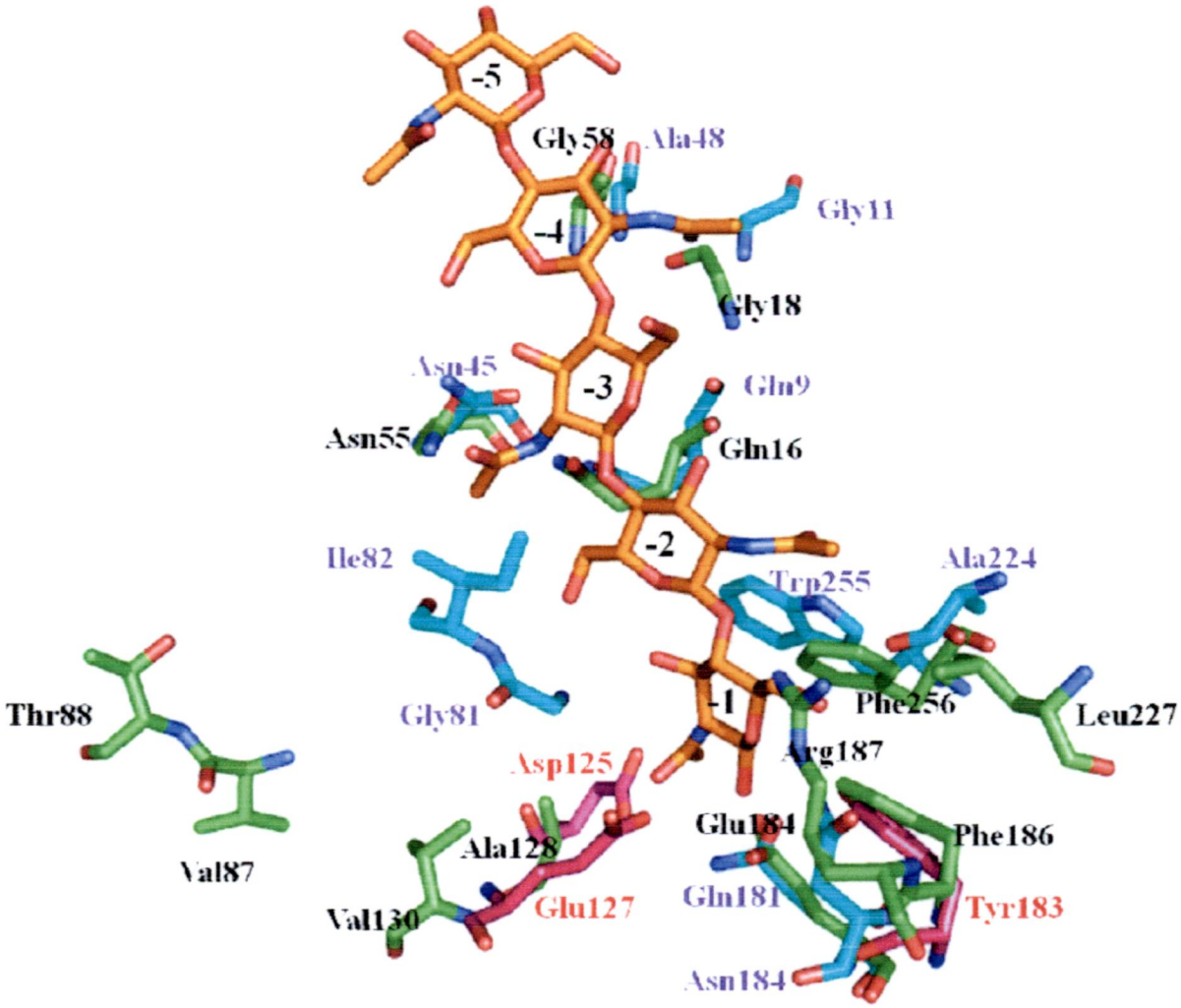


Figure 5.13. Substrate binding subsites and active site residues of TCLL with respect to hevamine. Hevamine (2HVM and 1KQY) was superimposed onto TCLL and the interacting residues with GLcNAc moiety are shown. Hevamine residues are in cyan sticks with purple label and the corresponding TCLL residues are in green sticks with black label. The active site residues of hevamine are shown in pink sticks with red label and corresponding residues of TCLL are Ala128, Val130 and Phe186 which shows that mutations of the vital catalytic residues.

5.3.8. Structural comparison with other GH18 'plant type' chitinases

The three dimensional structure of TCLL superimposed onto those of hevamine, PPL2, concanavalin B, XIP-I, XAIP and sccts1 shares 36%, 35%, 34%, 33%, 33% and 25% sequence identity and exhibit root mean square deviation (r.m.s.d.) for all Cα atoms of 0.81, 0.77, 0.78, 0.959 and 0.86Å, repectively. This suggests that TCLL displays an overall similar fold to those of the above structures except some loop deviations (Fig. 5.14). Compared to hevamine, some structural features in TCLL are different - such as glycosylation at Asn146, substitutions in the substrate binding site and the active site, smaller sizes of strand β1 and helix α3 than corresponding hevamine strand and helix, different conformation and extra strands of loops β2α2 and β3α3, loop β3α3 obtrusion from $(\beta\alpha)_8$ fold forming hinge like conformation and some major divergence of loops β4α4, β7α7 and α3β4. Asp123 & Trp179 are strictly conserved in all members of family 18 plant chitinases [205]. While Aspartate is conserved, Trptophan is replaced by Phenylalanine in TCLL. Unlike PPL2 from *Parkia platycephala* seeds which is an endochitinase with N-acetyl glucosamine binding hemagglutinin activity, TCLL shows glycosylation and lacks endochitinase activity due to substitution of active side residues. Nevertheless TCLL is N-acetyl glucosamine–binding lectin that shows lattice formation of human erythrocytes. Comparison of GlcNAc binding site of TCLL with PPL2 demonstrates variation in loop β4α4 which is longer and form cavity. The residues which hold the GlcNAc are also substituted in PPL2. The corresponding cavity in PPL2 is shallow and might not interact with GlcNAc.

Concanavalin B from from *Canavalia ensiformis*, xylanase inhibitor protein I (XIP-I) from *Triticum aestivum* and xylanase and alpha-amylase inhibitor protein (XAIP) from *Scadoxus multiflorus* belongs to GH18 family as they preserved some characteristic features of this family [98, 148, 174] but are devoid of chitinase activity due to substitution of key residues of active site. Narbonin from the seeds of *Vicia narbonensis* and *Vicia pannonica* is also reported as a member of this family that lacks chitinase activity [36, 71, 72]. Narbonin has Glu132 at similar position to hevamine catalytic glutamate and still do not show chitinase activity as Glu132 is engaged with Arg87 in forming salt bridge. One extra disulphide bond in concanavalin B Cys41-Cys93, which is engaged in stabilization of the loops β2α2 and β3α3 is observed that is unusual for the GH18 family [174] whereas, XIP-I and XAIP shows only two disulphide bonds [98, 148]. When compared to XIP-I which is a dual inhibitor of xylanase belonging to GH10 and GH11 family [149], loop α4β5 in TCLL is shorter in length. The elongated loop observed in XIP-I enter into the cavity of GH11 xylanase and residue Arg149 interact direct with active site residues

Glu85 and Glu176 while such interacting residues are missing in TCLL. GH 10 xylanase also has a TIM barrel topology and is inhibited by XIP-I helix α7 (232-245) which blocks central part of substrate binding groove (-3 to +2). The residues Lys234 interact directly with Glu131 and through water with Glu239. In TCLL, the residue Lys234 is conserved but other interacting residues are not. Other interacting residues 225 to 232 display different conformations and this loop moves towards core barrel rather than protrude out. Residues 189 to 205 attained some amino acid insertions and the resultant longer loop adopts β-turn conformation that obstructs the interaction with helix α7. Overall, TCLL might not function like XIP-I. XAIP shows GH11 xylanase and GH13 α-amylase inhibition. Kumar *et al*, (2010) [98] proposed loop α3β4 for xylanase inhibition and loop β6α6 and helix α7 for α-amylase inhibition. In case of TCLL, loop α3β4 is shorter in length and loop β6α6 has different conformation and proposed residues of helix α7 interacting with α-amylase are substituted. One another member of this family is a 'plant type' fungal chitinase, scctsl from *Saccharomyces cerevisiae,* which shares 25% identity with TCLL. Alignment of TCLL with scctsl structure shows difference at some loop regions and the active site region [79]. Although, several features typical of family 18 are preserved in TCLL structure such as the $(\beta\alpha)_8$ barrel fold with the two family-18 consensus regions (but with some mutations at these regions), presence of two non-proline cis-peptide bonds (Table 5.3), three disulphide bonds and elongated loop β2α2 having one antiparallel β-hairpins.

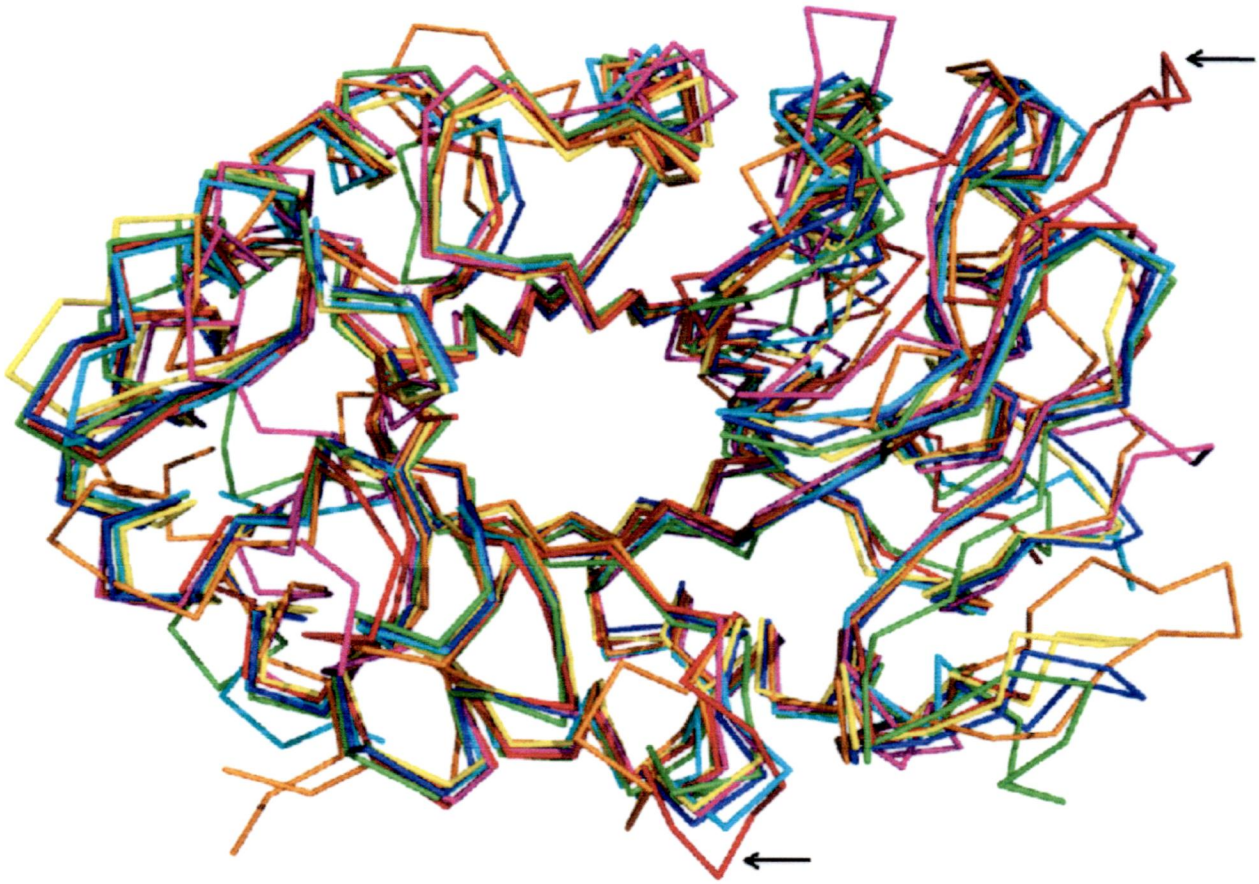


Figure 5.14. Superposition of TCLL with homologous structures of the GH18 family. Ribbon diagram showing the superposition of C_α atoms of TCLL (red), hevamine (blue) (2HVM) from latex of *Hevea brasiliensis*, PPL2 (yellow) (2GSJ) from *Parkia platycephala* seeds, concanavalin B (green) (1CNV) from *Canavalia ensiformis*, xylanase inhibitor protein I (magenta) (XIP-I) from *Triticum aestivum*, xylanase and alpha-amylase inhibitor protein (cyan) (XAIP) from *Scadoxus multiflorus* (3MU7) and sccts1 from *Saccharomyces cerevisiae* (orange) (2UY2). The superposition shows that the overall structure is conserved except some loop regions shown by arrows.

	Consensus sequence region β3		Consensus sequence region β4		two non-proline cis-peptide bonds β2		two non-proline cis-peptide bonds β8	
TCLL	78	IDILLQIGQVT	123	DGIDIASV	41	EE	256	FD
Hevamine	72	IKVMLSLGGGI	120	DGIDFDIE	31	AF	255	WS
PPL2	72	IKVMLSIGGGA	120	DGVDFDIE	31	AF	255	WD
ConB	77	VKVFLALGGPK	124	DGIHFDIQ	34	SF	265	WN
Xaip-Ii	68	VKVLLSIGGPA	119	DGIDFHIE	33	GF	253	WD
XIP-I	71	VPVSLSIGGYG	121	DGVDLFLE	36	SF	256	WD
Scctsl	101	KKVLLSLGGAS	150	DGFDFDIE	59	SF	285	WD

Table 5.3. Alignment of the two consensus regions in GH18 family plant chitinase and two conserved non-proline cis-peptide bonds

5.3.9. Structural comparison with other GH18 chitinase like proteins

A DALI search [74] shows structural similarity of TCLL with those of GH 18 chitinase like proteins (CLPs) or chi-lectins except for some loops and α/β insertion domains. Several reports demonstrated that CLPs show high sequence similarity to GH18 family but lack chitinase activity. They are thought to perform variety of functions and interact with sugars [31, 76, 88, 147, 178]. There was lack of significant sequence similarity between any CLPs and TCLL and structure of TCLL superimposed on those of HCgp-39 [76], Yml [190], MGP-40 [130], and IDGF-2 [208] showed overall similar structural fold but quite higher r.m.s. deviation. However, CLPs share some common features with TCLL like conserved three cis-peptide bonds and mutations in signature motif DXXDXDXE. TCLL is a glycoprotein like most of CLPs, however; glycosylation is not common to GH18 and the glycosylation site in TCLL is different than observed in CLPs like HCgp-39, MGP-40 and IDGF-2.

TCLL structure compared to the structure of HCgp-39 (1NWT, 1HJW) [55, 76], a human cartilage glycoprotein, in complex with chitin fragment shows differences in chitin binding groove. The residues that putatively interact with chitin are substituted in TCLL. The groove formed in HCgp-39 is long and deep allowing it to accommodate chitin fragment faultlessly. However, in case of TCLL, chitin binding cavity is shorter and very superficial. Moreover, the cavity forming loops on the opposite side are very apart thus resulting in an open cavity (Fig. 5.15). The two important residues, Trp99 and Trp352, of HCgp-39 which engage in hydrophobic stacking interactions with hydrophobic sides of sugars as well as makes hydrogen bonds with sugars [55] and contribute in ligand recognition are missing in TCLL. The structure of chitin fragment bound HCgp-39 suggests that chitin, which is structural element of nematodal and

fungal pathogen to human, can be physiological ligand of HCgp-39. Further, HCgp-39 is known to switch on innate immunity by sensing chitin of pathogen and cooperating with macrophage [76].

Yml, a secretory novel protein from murine activated peritoneal macrophages upon infection of nematode *Trichinella spiralis*, also shares structural similarity with TCLL. Yml has glucosamine and heparin/heparan sulfate binding ability but other significant biological functions have not yet been revealed. The chitin binding groove of Yml is neither distinct (Fig. 5.15) nor conserved and yet it binds glucosamine and the binding site is observed inside the core of TIM barrel at the carboxy terminal of the β-strands. In case of TCLL, this site is not properly defined, shallow and interacting residues are not conserved. One another member of this family is imaginal disc growth factor-2 (IDGF-2) from *Drosophila melanogaster*. The structure of IDGF-2 shows partly blocked binding cavity that lacks a proper configuration of residues for binding to oligosaccharides (Fig. 5.15). However, it has been proposed that it might promote cell proliferation by assisting to insulin. These structural comparisons reflect that GH18 chitinase like proteins preserve the TIM barrel domain of the family with no chitinase activity. Additionally, they have evolved to participate in diverse functions. TCLL has TIM barrel domain as well without a well defined chitin binding groove or catalytic ability. This explains evolutionary progression from chitinase to chitinase like lectin that binds to GlcNAc. The biological significance of GlcNAc binding of TCLL, however, needs to be explored.

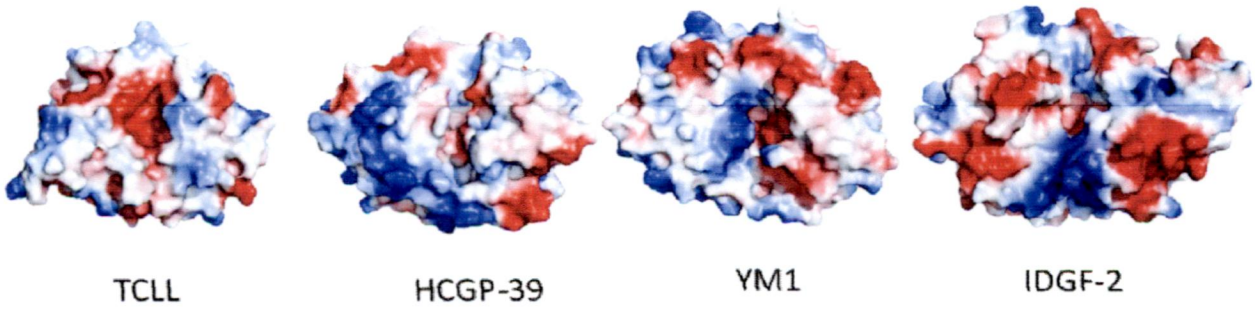


Figure 5.15. Electrostatic surface potential map of the chilectins: TCLL, HCGP-39 (1HJW), YM1 (1E9L), and IDGF-2 (1JND) displaying chitin binding groove. Electrostatic potential was calculated by Pymol and is colour-coded on the surface from blue (~63) to red (~63). Only HCGP-39 has appropriate groove and chitin fragment binds at this groove. TCLL displays more negative site and has deep pocket like structure. YM1 and IDGF-2 do not have well defined cavity.

5.3.10. Comparison with other GlcNAc binding lectins

As TCLL is GlcNAc binding chilectin, we evaluated the structure of other GlcNAc binding proteins. For example, *Agaricus bisporus* lectin (ABL) (PDB: 1Y2X) [27], *Boletus edulis* lectin [45] (PDB: 3QDV) [22], *Sclerotium rolfsii* lectin (SRL) (PDB: 2OFE) [108], *Ulex europaeus* lectin II (PDB: 1QOO) [113], *Psathyrella velutina* lectin (PVL) (PDB: 2C4D) [32], wheat germ agglutinin (WGA) (PDB: 2UVO) [176], and two lectins which are specific for saccharides containing GlcNAc namely, *Urtica dioica* agglutinin (UDA-VI) (PDB: 1EHH) [66] and *Phytolacca americana* lectin (PL-D2) (1ULM) [68]. Structurally, TCLL does not display any similarity with the aforesaid GlcNAc binding proteins. Further, the GlcNAc interacting residues and environment of the cavity are compared here. ABL has two diverse binding sites that differentiate the different configurations at a single epimeric hydroxyl. The GlcNAc interacted mainly with residues Asp79, Ile80, Thr82, Arg103 and Tyr113 of ABL. BEL has two sites: site 1 that hold the GalNAc and T-antigen disaccharide while binding site 2 is most likely chitin binding site with preference for GlcNAc and chitobiose. The interacting residues of BEL with GlcNAc are Asp78, Val79, Thr81, Arg102 and Tyr113. Asp78 defines specificity for binding to GlcNAc as its carboxylate oxygens form hydrogen bonds with O4 and O6 of the GlcNAc. SRL is similar to ABL and BEL with two distinct carbohydrate-binding sites, a primary and a secondary site. GalNAc is attached at the primary and GlcNAc prefers only the secondary site. Residues Asp77, Ile78, Thr80, Arg101, and Tyr112 are involved in hydrogen bonds with GlcNAc moiety. ABL, BEL and SRL have common interacting residues which delineate specificity to GlcNAc.

Ulex europaeus lectin II contains promiscuous carbohydrate-binding site with triad of Asp86/Gly106/Asn136 that accommodates GlcNAc, galactose and fucosylgalactose. PVL binds GlcNAc at six different sites with minor differences in binding mode. Two hydrogen bonds are formed by the side chain of one asparagine (or aspartate) residue with the O3 and O4 of GlcNAc. The O3 also develops a hydrogen bond to a conserved tryptophan. Hydrophobic contacts with GlcNAc are made by conserved histidine and tyrosine at each site. WGA has four hevein domains in each polypeptide. Structure of WGA complexed with GlcNAc displays five binding sites and residues (Asp or Glu/Ser/Ser/Tyr) of cavity built by hevein domains of 36-kDa homodimer interact with GlcNAc moiety. UDA-VI has two hevein analogous domains. The complex structure of UDA-VI with NAG3 displays two binding sites of UDA-VI molecule. One NAG3 is sandwiched between two molecules of UDA-VI. The hevein domain residues Trp21/Trp23and Ser19/Cys24/Tyr30 of domain 1 and His67/Trp69 and Ser65/Tyr76 of domain 2

recognize and interact with NAG3 molecule. PL-D2 also has two hevein domains (I and II) and the interacting residues in its complex with NAG3 are Ser20, Trp22, Tyr24 and Tyr31 of domain I and Trp43, Ser61, Tyr63, Trp65 and His72 of domain II.

Remarkably, comparison with other GlcNAc binding lectins shows that TCLL has different architecture of cavity and novel binding mode which has not been documented so far in plant chilectins. A relevant question is the biological role that TCLL plays in the seeds and the significance of the GlcNAc binding. It is documented that several plant lectins serve as plant defence proteins due to their carbohydrate specificity and fight against phytophagous insects and fungus [150, 168, 169, 207]. It was also observed that lectins from seeds of Leguminosae family might play role in symbiosis with nitrogen fixing bacteria [42, 169]. Additionally, in case of TCLL it was observed that it is degraded during germination as was also observed in most of seed lectins [169]. From this observation, proposing TCLL as just seed storage protein may not be reasonable considering the presence of specific carbohydrate binding site. Additional exploration is required to ascertain the physiological function of the GlcNAc-binding site in the TCLL.

5.4. Discussion

In the present study, we have purified and crystallized chitinase like lectin from tamarind seeds. We proposed amino acid sequence for TCLL from our high resolution structure data and further confirmed it by internal sequencing. By considering the amino acid similarity, TCLL becomes a member of the GH18 family [http://afmb.cnrsmrs. fr/CAZy/index.html; The CAZy (Carbohydrate-Active enZYmes) database describes the families of structurally-related catalytic and carbohydrate-binding modules (or functional domains) of enzymes that degrade, modify, or create glycosidic bonds]. TCLL displays 36% sequence identity with hevamine, a class III chitinase of GH18 family isolated from the latex of *Hevea brasiliensis* [205]. It also shares significant sequence identity with the other members of this family such as PPL2 from *Parkia platycep*hala seeds [28], concanavalin B from *Canavalia ensiformis* [174], xylanase inhibitor protein I (XIP-I) from *Triticum aestivum* [148] and xylanase and alpha-amylase inhibitor protein (XAIP) from *Scadoxus multiflorus* [98]. All the proteins mentioned above, possess the $(\beta\alpha)_8$ TIM barrel topology with the two consensus regions characteristic of the GH18 family. Among this class, hevamine and PPL2 are active chitinase whereas, concanavalin B, XIP-I and XAIP are bereft of chitinase activity due to mutation of key residues of active site.

TCLL showed neglegible chitinase activity which led us to conclude that it may be a chitinase. However, further structural studies revealed that all the critical active site residues have been mutated to hydrophobic residues. Evaluation of the chitin binding cavity of TCLL showed that it is shorter, very superficial, deep pocket like instead of groove and cavity forming loops that are marginally apart creating an open cavity. It also further confirmed that TCLL has inappropriate chitin binding cavity as TCLL co-crystallization and soaking with chitotrise and chitotetraose were unsuccessful. Now the question arises that from where the chitinase activity appeared. Rao *et al.*, also reported the 34kDa class III chitinase from tamarind seeds called as tamarinin which has very low chitinase activity [157] and later same group deposited its sequence. Interestingly, TCLL sequence proposed from high resolution crystallographic data and by mass spectroscopy shares 39% sequence identity with tamarinin. Surprisingly, biochemical studies and the 3D structure of TCLL with GlcNAc revealed that TCLL is lectin. The GlcNAc binding site observed in the TCLL complex structure is novel, which is created by loops, β4α4 and β5α5. The interacting residues and binding environment are different than other GlcNAc binding lectins. From these observations, we hypothesize that there might be evolution of lectin that has occurred in tamarind sequentially. Tamarinin is at the phase where it has little chitinase activity left and TCLL is in a phase where it is a GlcNAc binding lectin and minor chitinase activity of TCLL is probably due to minute presence of tamarinin in it.

Mammalian lectins HCGP 39 [76] and Yml [190] which are also the members of GH18 family with $(\beta\alpha)_8$ barrel fold are inactive chitinases. The human chi-lectin HCGP 39 binds chitin fragment and Yml, the murine chitinase like lectin, comprise specificity towards GlcN and heparin/heparan sulfate [190]. Hitherto the physiological significance of these finding is not clear. A lectin (RobpsCRA) reported from bark of *Robinia pseudoacacia* shares 50% sequence identity with plant class V chitinases of the GH18 family. This protein also lacks chitinase activity but interacts with high-Man N-glycans. The experimental structure has not been solved albeit by molecular modeling it has been shown to have a TIM barrel domain [203]. Recognition of TCLL as a lectin, which evolved from chitinase illustrates acquiring of new activity by losing chitinase activity. Adoption of a new function by inactive chitinases of class III of GH18 family was also reported earlier. For example, XIP-I which is inactive chitinase was reported to inhibit fungal xylanases of class GH10 and GH11 [149]. XAIP is documented as inhibitor of xylanases of family GH11 and α-amylases of family GH13 [98]. These evolved proteins still have same fold as that of parent proteins but have developed a different site or conformation of interacting loops with ligands. Interestingly, these are not just inactive chitinases but proteins having precise

biological functions. So, utilization of family GH18 existing scaffold is common to conquer new function, as also exemplified by TCLL.

Due to diversity and evolution, classification of plant lectins is a challenging task. Previously, plant lectins were classified into seven classes based on the carbohydrate recognition domains but recently Van Damme *et al*, [204] classified plant lectins into twelve families based on sequence, structural and evolutionary relatedness. For example, RobpsCRA, a lectin from black locust bark having 50% sequence identity with class V chitinase, accordingly is classified as 'Class V chitinase homolog' in the existing class of plant lectins. TCLL structure reveals that it has a TIM barrel domain and is evolutionary related to class III chitinase as its shows highest sequence identity with hevamine of class III of GH18 family. So TCLL cannot be categorized into previous lectin classification and we propose a new class of lectin which is evolutionarily related to class III chitinase. Moreover, the existing 'Class V chitinase homologs' and our proposed class 'Class III chitinase homologs' have same fold. To avoid number of classes in lectin classification, we propose the new class "TIM barrel domain" rather than existing one and these two classes can be grouped into TIM barrel domain as one subclass. One cannot exclude the possibility that the glycosyl hydrolase family members or other families having TIM barrel topology from plants will not emerge as lectins. So the TIM barrel domain will be the main class and the other classes will be subclasses.

This is the first structural report of this new family of "TIM barrel domain" of plant lectins which binds GlcNAc monomer at different site than observed in chitin binding site of GH18 family. However, the biological significance of carbohydrate binding of TCLL awaits further study. It is speculated that it might work as plant defense protein and/or may take part in symbiosis like some other plant lectins. However, a thorough investigation is required to shed further light on possible biological function of TCLL.

Purification and biochemical characterization of TKI

- The tamarind Kunitz type inhibitor was purified to homogeneity by three step procedure. The first step involved usage of an affinity matrix affi-gel blue gel followed by anion CM and cation DEAE column.
- The SDS-PAGE analysis under both reducing and non-reducing conditions confirmed that TKI is a single polypeptide chain with approximate molecular mass of 21 kDa.
- The trypsin inhibitory activity and N-terminal amino acid sequence analysis revealed the isolated protein is a protease inhibitor belonging to Kunitz type inhibitor family.
- The N-terminal amino acid sequence of TKI showed 79% homology with factor Xa inhibitor (BuXI) from *Bauhinia ungulata* and 64% with Trypsin inhibitor (BvTI) from *Bauhinia variegata.*
- TKI showed significant prolongation of coagulation time of partial activated thromboplastin time (APTT) and prothrombin time (PT). The TKI (10 μM) extended the APTT of normal plasma 2.6 folds and PT 5 folds.
- TKI showed inhibition of factor Xa.
- TKI caused dose dependent inhibition of cell viability of cancer cells. Among these cells MCF-7 and HepG2 were more sensitive whereas PC-3 was least sensitive. After 24 h of treatment TKI caused about 14 % and 13 % reduction in cell viability of breast (MCF-7) and liver (HepG2) cancer cells respectively at 125 μM concentration whereas there was only 6% reduction in cell viability of prostate cancer (PC-3) cells at the same dose.

cDNA synthesis, cloning, expression and purification of TKI

- Total RNA was isolated from tamarind seeds followed by cDNA synthesis using oligo$(dT)_{18}$ adaptor primer.
- The cDNA was amplified which yielded a ~700 bp amplified product containing a 3' UTR region and a poly (A+) tail.
- A 555 bp ORF obtained after sequencing of the TKI gene codes for a polypeptide of 185 amino acid residues with a calculated molecular mass of 20575 Da.
- Sequence analysis demonstrated that TKI belongs to Kunitz STI type inhibitors and shows highest sequence identity (45%) with factor Xa inhibitor from *Bauhinia ungulata* (BuXI).

- The Kunitz signature sequence gets distorted due to insertion of Asn residue at position 15 within the Kunitz signature sequence.
- Sequence analysis showed that TKI has conserved two disulphide bonds and Arg at P1 site.
- TKI gene was cloned into pET-28cTEV vector and expressed in chemically competent *E. coli* K12 cells shuffle T7. The expressed TKI showed good expression and solubility and possessed potent trypsin inhibitory activity like native TKI.
- Recombinant TKI (rTKI) was purified using Ni-NTA column giving single purified band of rTKI on SDS-PAGE. The purified rTKI exhibited trypsin inhibitory activity. This result show that rTKI is a stable protein and able to preserve its functional property like native TKI.

Crystallization and structure determination of TKI, its complex with trypsin and molecular docking with factor Xa

- The structure of TKI displays β trefoil fold with two disulphide bridges and an exposed reactive site loop having Arg at P1.
- TKI has insertion of residue Asn15 at the Kunitz signature motif which is clearly visible in electron density.
- The TKI:PPT complex structure shows that the exposed reactive site of TKI block the S1 pocket of PPT and stabilized the complex by making interaction with the subsites of PPT.
- The flexible residues of reactive site P1, P3 and P'3 become rigid after complex formation.
- The modeling studies of TKI with factor Xa showed that Arg residues at P1 and P3 have L-shaped conformation which could possibly blocks the S1 and S4 pocket of factor Xa.
- The most important residue of TKI that may provide specificity towards FXa is Arg64. It interacts with cation hole and it side chain occupies the aryl binding site of S4.
- Structure of TKI:PPT complex and docking studies with FXa display the role of inserted residue, Asn15 in stabilization of complexes. Asn15 interact with four residues of PPT and two residues of FXa and participate in stabilization of complex. Furthermore, Asn15 interacts with residues of FXa which are important for specificity, pointing towards its role in specific recognition of proteases too.

- Evaluation of reactive site residues of TKI shows that it has novel reactive site residues as they don't have similarity with reactive site of known Kunitz type inhibitors as well as physiological and non-physiological factor Xa inhibitors.
- Taken together, TKI through its versatile, unique reactive site could inhibit other serine proteases also and may act as multifunctional inhibitor.

Purification, biochemical characterization, crystallization, structure determination of tamarind chitinase like lectin (TCLL) and its complex with GlcNAc

- The tamarind chitinase like lectin was purified to homogeneity by three step procedure. The first step involved usage of an affinity matrix affi-gel blue gel followed by anion CM and cation DEAE column.
- The SDS-PAGE analysis under both reducing and non-reducing conditions confirmed that TCLL is a single polypeptide chain with approximate molecular mass of 34 kDa.
- TCLL showed neglegible chitinase activity and hemaglutination assay showed lectin-like activity. Further hemaglutination inhibition by sugar GlcNAc and fluorescence studies showed that TCLL is a GlcNAc binding lectin.
- TCLL structure was solved at 1.4Å which revealed that it possess the $(\beta\alpha)_8$ TIM barrel topology.
- Furthermore, structural studies revealed that all the critical active site residues have been mutated to hydrophobic residues. Evaluation of the chitin binding cavity of TCLL showed that it is shorter, very superficial, deep pocket like instead of groove and cavity forming loops that are marginally apart creating an open cavity.
- We proposed amino acid sequence for TCLL from our high resolution structure data which was further confirmed by internal sequencing. By considering the amino acid similarity, TCLL displays highest sequence identity with members of class III chitinase of GH18 family.
- Biochemical studies and the 3D structure of TCLL with GlcNAc revealed that TCLL is a lectin. The GlcNAc binding site observed in the TCLL complex structure is novel, which is created by loops, β4α4 and β5α5.
- Recognition of TCLL as a lectin, which evolved from chitinase illustrates acquiring of new activity by losing chitinase activity

- TCLL does not fit into classification of lectin proposed by Van Damme *et al,* 2008, so we propose a new class of lectin which is evolutionarily related to class III chitinase.
- This is the first structural report of this new family of "TIM barrel domain" of plant lectins which bind GlcNAc monomer at different site than the chitin binding site of GH18 family.

Tamarind Kunitz type inhibitor (TKI)

- Site directed mutagenesis at reactive site residues can further give more detailed information regarding individual role of each amino acid of reactive site.
- Mutation studies of Asn15 will shed light on role of inserted Asn15 on inhibitory action of TKI and importance in stabilization of compleses with porteases
- Detailed studies on anticancer activity can be performed to find out the target of TKI
- Structure determination of rTKI can be done to see any difference with native TKI structure.
- Structural studies on complex with factor Xa can be carried out to attain knowledge about molecular basis of inhibition of factor Xa by TKI.
- It is possible to design inhibitors against its cognate proteases

Tamarind chitinase like lectin (TCLL)

- cDNA cloning of TCLL will provide complete amino acid sequence which can be further used for further studies.
- Site directed mutagenesis can be performed at active site residue to study whether these changes could bring about hydrolysis of chitin
- Studies on complex glycan could provide information that may have biotechnological value.
- Functional studies of TCLL can be performed to know its biological role.

REFERENCES

1. Adhikari, P., Bachhawat-Sikder, K., Thomas, C. J., Ravishankar, R., Jeyaprakash, A. A.,Sharma,V., Vijayan, M. and Surolia, A. Mutational Analysis at Asn-41 in Peanut Agglutinin. *J. Biol. Chem.* 276(44):40734-40739 (2001).
2. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search program. *Nucleic Acids Res.* 25(17):3389-3402 (1997).
3. Ambrosi, M., Cameron, N. R. and Davis, B. G. Lectins: tools for the molecular understanding of the glycocode. *Org. Biomol. Chem.* 3(9):1593-1608 (2005).
4. Andrade, S. A., Santomauro-Vaz, E. M., Lopes, A. R., Chudzinski-Tavassi, A. M., Juliano, M. A., Terra, W. R., Sampaio, M. U., Sampaio, C. A. M. and Oliva, M. L. V. *Bauhinia* proteinase inhibitor-based synthetic fluorogenic substrates for enzymes isolated from insect midgut and caterpillar bristles. *Biol. Chem.* 384(3):489-492 (2003).
5. Araújo, C. L., Bezerra, I. W. L., Oliveira, A. S., Moura, F. T., Macedo, L. L. P., Gomes, C. E. M., Barbosa, A. E. A. D., Macedo, F. P., Souza, T. M. S. and Franco, O. L. *In vivo* bioinsecticidal activity toward *Ceratitis capitata* (fruit fly) and *Callosobruchus maculatus* (cowpea weevil) and *in vitro* bioinsecticidal activity toward different orders of insect pests of a trypsin inhibitor purified from tamarind tree (*Tamarindus indica*) seeds. *J. Agric. Food Chem.* 53(11):4381-4387 (2005).
6. Asthagiri, D., Liu, T., Noodleman, L., Van Etten, R. L. and Bashford, D. On the role of the conserved aspartate in the hydrolysis of the phosphocysteine intermediate of the low molecular weight tyrosine phosphatase. *J. Am. Chem. Soc.* 126(39):12677-12684 (2004).
7. Bailey, S. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* 50:760-763 (1994).
8. Balzarini, J., Schols, D., Neyts, J., Van Damme, E., Peumans, W. and De Clercq, E. Alpha-(1-3)-and alpha-(1-6)-D-mannose-specific plant lectins are markedly inhibitory to human immunodeficiency virus and cytomegalovirus infections in vitro. *Antimicrob. Agents Chemother.* 35(3):410-416 (1991).
9. Banerjee, R., Das, K., Ravishankar, R., Suguna, K., Surolia, A. and Vijayan, M. Conformation, protein-carbohydrate interactions and a novel subunit association in the refined structure of peanut lectin-lactose complex. *J. Mol. Biol.* 259:281-296 (1996).
10. Bao, R., Zhou, C. Z., Jiang, C., Lin, S. X., Chi, C. W. and Chen, Y. The ternary structure of the double-headed arrowhead protease inhibitor API-A complexed with two trypsins reveals a novel reactive site conformation. *J. Biol. Chem.* 284(39):26676-26684 (2009).
11. Barriault, D., Lépine, F., Mohammadi, M., Milot, S., Leberre, N. and Sylvestre, M. Revisiting the regiospecificity of *Burkholderia xenovorans* LB400 biphenyl dioxygenase toward 2,2'-dichlorobiphenyl and 2,3,2',3'-tetrachlorobiphenyl. *J. Biol. Chem.* 279(46):47489-47496 (2004).
12. Batista, I. F. C., Oliva, M. L. V., Araujo, M. S., Sampaio, M. U., Richardson, M., Fritz, H. and Sampaio, C. A. M. Primary structure of a Kunitz-type trypsin inhibitor from *Enterolobium contortisiliquum* seeds. *Phytochemistry* 41(4):1017-1022 (1996).
13. Batista, I. F. C., Ramos, O. H. P., Ventura, J. S., Junqueira-de-Azevedo, I. L. M., Ho, P. L. and Chudzinski-Tavassi, A. M. A new Factor Xa inhibitor from *Amblyomma cajennense* with a unique domain composition. *Arch. Biochem. Biophys.* 493(2):151-156 (2010).
14. Bhadoriya, S. S., Ganeshpurkar, A., Narwaria, J., Rai, G. and Jain, A. P. *Tamarindus indica*: Extent of explored potential. *Pharmacogn. Rev.* 5(9):73-81 (2011).
15. Bhuvaneswari, T. V., Pueppke, S. G. and Bauer, W. D. Role of lectins in plant-microorganism interactions: I. Binding of soybean lectin to rhizobia. *Plant Physiol.* 60(4):486-491 (1977).

16. Bilgrami, S., Tomar, S., Yadav, S., Kaur, P., Kumar, J., Jabeen, T., Sharma, S. and Singh, T. P. Crystal Structure of Schistatin, a Disintegrin Homodimer from Saw-scaled Viper (*Echis carinatus*) at 2.5A Resolution. *J. Mol. Biol.* 341(3):829-837 (2004).
17. Birk, Y. Proteinase inhibitors from legume seeds. *Methods Enzymol.* 45:697-700 (1976).
18. Birrane, G., Varma, A. K., Soni, A. and Ladias, J. A. A. Crystal structure of the BARD1 BRCT domains. *Biochemistry* 46(26):7706-7712 (2007).
19. Bode, W. and Huber, R. Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem.* 204(2):433-451 (1992).
20. Bode, W. and Huber, R. Structural basis of the endoproteinase-protein inhibitor interaction. *Biochim. Biophys. Acta* 1477(1-2):241-252 (2000).
21. Bokma, E., Rozeboom, H. J., Sibbald, M., Dijkstra, B. W. and Beintema, J. J. Expression and characterization of active site mutants of hevamine, a chitinase from the rubber tree *Hevea brasiliensis*. *Eur. J. Biochem.* 269(3):893-901 (2002).
22. Bovi, M., Carrizo, M. E., Capaldi, S., Perduca, M., Chiarelli, L. R., Galliano, M. and Monaco, H. L. Structure of a lectin with antitumoral properties in king bolete (*Boletus edulis*) mushrooms. *Glycobiology* 21(8):1000-1009 (2011).
23. Boyd, W. C. and Shapleigh, E. Specific Precipitating Activity of Plant Agglutinins (Lectins). *Science* 119(3091):419 (1954).
24. Brandstetter, H., Kühne, A., Bode, W., Huber, R., von der Saal, W., Wirthensohn, K. and Engh, R. A. X-ray structure of active site-inhibited clotting factor Xa. *J. Biol. Chem.* 271(47):29988-29992 (1996).
25. Brouillard, J. N. P., Günther, S., Varma, A. K., Gryski, I., Herfst, C. A., Rahman, A. K. M., Leung, D. Y. M., Schlievert, P. M., Madrenas, J. and Sundberg, E. J. Crystal structure of the streptococcal superantigen SpeI and functional role of a novel loop domain in T cell activation by group V superantigens. *J. Mol. Biol.* 367(4):925-934 (2007).
26. Burgering, M. J. M., Orbons, L. P. M., van der Doelen, A., Mulders, J., Theunissen, H. J. M., Grootenhuis, P. D. J., Bode, W., Huber, R. and Stubbs, M. T. The second Kunitz domain of human tissue factor pathway inhibitor: cloning, structure determination and interaction with factor Xa. *J. Mol. Biol.* 269(3):395-407 (1997).
27. Carrizo, M. E., Capaldi, S., Perduca, M., Irazoqui, F. J., Nores, G. A. and Monaco, H. L. The antineoplastic lectin of the common edible mushroom (*Agaricus bisporus*) has two binding sites, each specific for a different configuration at a single epimeric hydroxyl. *J. Biol. Chem.* 280(11):10614-10623 (2005).
28. Cavada, B. S., Moreno, F. B. B., Da Rocha, B. A. M., De Azevedo Jr, W. F., Castellón, R. E. R., Goersch, G. V., Nagano, C. S., De Souza, E. P., Nascimento, K. S. and Radis Baptista, G. cDNA cloning and 1.75 Å crystal structure determination of PPL2, an endochitinase and N acetylglucosamine binding hemagglutinin from *Parkia platycephala* seeds. *FEBS J.* 273(17):3962-3974 (2006).
29. Chandra, N. R., Kumar, N., Jeyakani, J., Singh, D. D., Gowda, S. B. and Prathima, M. N. Lectindb: a plant lectin database. *Glycobiology* 16(10):938-946 (2006).
30. Chandra, V., Jasti, J., Kaur, P., Betzel, C., Srinivasan, A. and Singh, T. P. First structural evidence of a specific inhibition of phospholipase A2 by alpha-tocopherol (vitamin E) and its implications in inflammation: crystal structure of the complex formed between phospholipase A2 and alpha-tocopherol at 1.8 A resolution. *J. Mol. Biol.* 320(2):215-222 (2002).
31. Chang, N. C. A., Hung, S. I., Hwa, K. Y., Kato, I., Chen, J. E., Liu, C. H. and Chang, A. C. A macrophage protein, Ym1, transiently expressed during inflammation is a novel mammalian lectin. *J. Biol. Chem.* 276(20):17497-17506 (2001).

32. Cioci, G., Mitchell, E. P., Chazalet, V., Debray, H., Oscarson, S., Lahmann, M., Gautier, C., Breton, C., Perez, S. and Imberty, A. [beta]-Propeller Crystal Structure of Psathyrella velutina Lectin: An Integrin-like Fungal Protein Interacting with Monosaccharides and Calcium. *J. Mol. Biol.* 357(5):1575-1591 (2006).
33. Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 50:760-763 (1994).
34. Comeau, S. R., Gatchell, D. W., Vajda, S. and Camacho, C. J. ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Res.* 32(suppl 2):W96-W99 (2004).
35. Comeau, S. R., Gatchell, D. W., Vajda, S. and Camacho, C. J. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* 20(1):45-50 (2004).
36. Coulson, A. F. W. A proposed structure for 'family 18' chitinases. A possible function for narbonin. *FEBS Lett.* 354(1):41-44 (1994).
37. Crooks, G. E., Hon, G., Chandonia, J. M. and Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* 14(6):1188-1190 (2004).
38. Czapinska, H. and Otlewski, J. Structural and energetic determinants of the S1-site specificity in serine proteases. *Eur. J. Biochem.* 260(3):571-595 (1999).
39. Damme, E. J. M. V., Peumans, W. J., Barre, A. and Rouge, P. Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Crit. Rev. Plant Sci.* 17(6):575-692 (1998).
40. Dasgupta, J., Khamrui, S., Dattagupta, J. K. and Sen, U. Spacer Asn determines the fate of Kunitz (STI) inhibitors, as revealed by structural and biochemical studies on WCI mutants. *Biochemistry* 45(22):6783-6792 (2006).
41. Dattagupta, J. K., Podder, A., Chakrabarti, C., Sen, U., Mukhopadhyay, D., Dutta, S. K. and Singh, M. Refined crystal structure (2.3A) of a double-headed winged bean alpha-chymotrypsin inhibitor and location of its second reactive site. *Proteins* 35(3):321-331 (1999).
42. De Hoff, P. L., Brill, L. M. and Hirsch, A. M. Plant lectins: the ties that bind in root symbiosis and plant defense. *Mol. Genet. Genomics* 282(1):1-15 (2009).
43. De Paula, C. A. A., Coulson-Thomas, V. J., Ferreira, J. G., Maza, P. K., Suzuki, E., Nakahata, A. M., Nader, H. B., Sampaio, M. U. and Oliva, M. L. V. EcTI, a plant proteinase inhibitor, decreases in vitro cell adhesion and invasion by inhibition of Src-FAK signaling pathways. *J. Biol. Chem.* 287(1):170-182 (2012).
44. DeLano, W. L. The PyMOL molecular graphics system. *See* http://www.pymol.org (2002).
45. DiBella, F. P. and Liener, I. E. Soybean trypsin inhibitor. Cleavage and identification of a disulfide bridge not essential for activity. *J. Biol. Chem.* 244(11):2824-2829 (1969).
46. Dubois, M., Gilles, K., Hamilton, J. K., Rebers, P. A. and Smith, F. A colorimetric method for the determination of sugars. *Nature* 168(4265):167 (1951).
47. Dunn, E. F., Gay, N. J., Bristow, A. F., Gearing, D. P., O'Neill, L. A. J. and Pei, X. Y. High-resolution structure of murine interleukin 1 homologue IL-1F5 reveals unique loop conformations for receptor binding specificity. *Biochemistry* 42(37):10938-10944 (2003).
48. Edelman, G. M., Cunningham, B. A., Reeke, G. N., Becker, J. W., Waxdal, M. J. and Wang, J. L. The covalent and three-dimensional structure of concanavalin A. *Proc. Natl. Acad. Sci. U. S. A.* 69(9):2580-2584 (1972).
49. Emsley, P. and Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60(12):2126-2132 (2004).
50. Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* 66(4):486-501 (2010).

51. Erlanger, B. F., Kokowsky, N. and Cohen, W. The preparation and properties of two new chromogenic substrates of trypsin. *Arch. Biochem. Biophys.* 95:271-278 (1961).
52. Fang, E. F., Wong, J. H., Bah, C. S. F., Lin, P., Tsao, S. W. and Ng, T. B. *Bauhinia variegata* var. variegata trypsin inhibitor: from isolation to potential medicinal applications. *Biochem. Biophys. Res. Commun.* 396(4):806-811 (2010).
53. Farber, G. K. An [alpha]/[beta]-barrel full of evolutionary trouble. *Curr. Opin. Struc. Biol.* 3(3):409-412 (1993).
54. Fu, L., Zhou, C., Yao, S., Yu, J., Liu, B. and Bao, J. Plant lectins: Targeting programmed cell death pathways as antitumor agents. *Int. J. Biochem. Cell Biol.* 43(10):1442-1449 (2011).
55. Fusetti, F., Pijning, T., Kalk, K. H., Bos, E. and Dijkstra, B. W. Crystal structure and carbohydrate-binding properties of the human cartilage glycoprotein-39. *J. Biol. Chem.* 278(39):37753-37760 (2003).
56. Gahloth, D., Selvakumar, P., Shee, C., Kumar, P. and Sharma, A. K. Cloning, sequence analysis and crystal structure determination of a miraculin-like protein from *Murraya koenigii*. *Arch. Biochem. Biophys.* 494(1):15-22 (2010).
57. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. and Bairoch, A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31(13):3784-3788 (2003).
58. Georgieva, D., Greunke, K., Genov, N. and Betzel, C. 3-D Model of the bee venom acid phosphatase: Insights into allergenicity. *Biochem. Biophys. Res. Commun.* 378(4):711-715 (2009).
59. Georgieva, D., Kardas, A., Buck, F., Perbandt, M. and Betzel, C. Isolation, crystallization and preliminary X-ray diffraction analysis of L-amino-acid oxidase from *Vipera ammodytes* ammodytes venom. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 64(10):918-921 (2008).
60. Gómez-Gil, L., Kumar, P., Barriault, D., Bolin, J. T., Sylvestre, M. and Eltis, L. D. Characterization of biphenyl dioxygenase of *Pandoraea pnomenusa* B-356 as a potent polychlorinated biphenyl-degrading enzyme. *J. Bacteriol.* 189(15):5705-5715 (2007).
61. Gouet, P., Courcelle, E. and Stuart, D. I. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15(4):305-308 (1999).
62. Gray, J., Janick-Buckner, D., Buckner, B., Close, P. S. and Johal, G. S. Light-dependent death of maize lls1 cells is mediated by mature chloroplasts. *Plant Physiol.* 130(4):1894-1907 (2002).
63. Günther, S., Varma, A. K., Moza, B., Kasper, K. J., Wyatt, A. W., Zhu, P., Rahman, A. K. M., Li, Y., Mariuzza, R. A. and McCormick, J. K. A novel loop domain in superantigens extends their T cell receptor recognition site. *J. Mol. Biol.* 371(1):210-221 (2007).
64. Gustafson, C. L. T., Stauffacher, C. V., Hallenga, K. and Van Etten, R. L. Solution structure of the low-molecular-weight protein tyrosine phosphatase from Tritrichomonas foetus reveals a flexible phosphate binding loop. *Protein Sci.* 14(10):2515-2525 (2005).
65. Hansen, D., Macedo-Ribeiro, S., VerÃssimo, P., Yoo Im, S., Sampaio, M. U. and Oliva, M. L. V. Crystal structure of a novel cysteinless plant Kunitz-type protease inhibitor. *Biochem. Biophys. Res. Commun.* 360(4):735-740 (2007).
66. Harata, K. and Muraki, M. Crystal structures of Urtica dioica agglutinin and its complex with tri-N-acetylchitotriosel. *J. Mol. Biol.* 297(3):673-681 (2000).
67. Hassan, M. I., Bilgrami, S., Kumar, V., Singh, N., Yadav, S., Kaur, P. and Singh, T. P. Crystal Structure of the Novel Complex Formed between Zinc [alpha] 2-Glycoprotein (ZAG) and Prolactin-Inducible Protein (PIP) from Human Seminal Plasma. *J. Mol. Biol.* 384(3):663-672 (2008).

68. Hayashida, M., Fujii, T., Hamasu, M., Ishiguro, M. and Hata, Y. Similarity between protein-protein and protein-carbohydrate interactions, revealed by two crystal structures of lectins from the roots of pokeweed. *J. Mol. Biol.* 334(3):551-565 (2003).
69. Hedstrom, L. Serine protease mechanism and specificity. *Chemical Rev.* 102(12):4501-4524 (2002).
70. Hehemann, J. H., Redecke, L., Murugaiyan, J., von Bergen, M., Betzel, C. and Saborowski, R. Autoproteolytic stability of a trypsin from the marine crab *Cancer pagurus*. *Biochem. Biophys. Res. Commun.* 370(4):566-571 (2008).
71. Hennig, M., Pfeffer-Hennig, S., Dauter, Z., Wilson, K. S., Schlesier, B. and Nong, V. H. Crystal structure of narbonin at 1.8A resolution. *Acta Crystallogr. D Biol. Crystallogr.* 51(2):177-189 (1995).
72. Hennig, M., Schlesier, B., Dauter, Z., Pfeffer, S. and Betzel, C. A TIM barrel protein without enzymatic activity? Crystal-structure of narbonin at 1.8A resolution. *FEBS Lett.* 306(1):80-84 (1992).
73. Hoffmann, J. J. and Meulendijk, P. N. Comparison of reagents for determining the activated partial thromboplastin time. *Thromb. Haemost.* 39(3):640-645 (1978).
74. Holm, L. and Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* 38(suppl 2):W545-W549 (2010).
75. Horigome, D., Satoh, H., Itoh, N., Mitsunaga, K., Oonishi, I., Nakagawa, A. and Uchida, A. Structural mechanism and photoprotective function of water-soluble chlorophyll-binding protein. *J. Biol. Chem.* 282(9):6525-6531 (2007).
76. Houston, D. R., Recklies, A. D., Krupa, J. C. and van Aalten, D. M. F. Structure and ligand-induced conformational change of the 39-kDa glycoprotein from human articular chondrocytes. *J. Biol. Chem.* 278(32):30206-30212 (2003).
77. Huber, R. and Carrell, R. W. Implications of the three-dimensional structure of alpha. 1-antitrypsin for structure and function of serpins. *Biochemistry* 28(23):8951-8966 (1989).
78. Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., Cuche, B. A., De Castro, E., Lachaize, C., Langendijk-Genevaux, P. S. and Sigrist, C. J. A. The 20 years of PROSITE. *Nucleic Acids Res.* 36(suppl 1):D245-D249 (2008).
79. Hurtado-Guerrero, R. and van Aalten, D. M. F. Structure of *Saccharomyces cerevisiae* chitinase 1 and screening-based discovery of potent inhibitors. *Chem. Biol.* 14(5):589-600 (2007).
80. Imoto, T. and Yagishita, K. A simple activity measurement of lysozyme. *Agric. Biol. Chem.* 35(7):1154-1156 (1971).
81. Inagaki, K., Kobayashi, H., Yoshida, R., Kanada, Y., Fukuda, Y., Yagyu, T., Kondo, T., Kurita, N., Kitanaka, T. and Yamada, Y. Suppression of urokinase expression and invasion by a soybean Kunitz trypsin inhibitor are mediated through inhibition of Src-dependent signaling pathways. *J. Biol. Chem.* 280(36):31428-31437 (2005).
82. Inoue, H., Nojima, H. and Okayama, H. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* 96(1):23-28 (1990).
83. Jabeen, T., Sharma, S., Singh, N., Singh, R. K., Verma, A. K., Paramasivam, M., Srinivasan, A. and Singh, T. P. Structure of the zinc-induced heterodimer of two calcium-free isoforms of phospholipase A2 from *Naja naja sagittifera* at 2.7 A resolution. *Acta Crystallogr. D Biol. Crystallogr.* 61(3):302-308 (2005).
84. Johnson, D. J. D., Li, W., Adams, T. E. and Huntington, J. A. Antithrombin-S195A factor Xa-heparin structure reveals the allosteric mechanism of antithrombin activation. *EMBO J.* 25(9):2029-2037 (2006).

85. Karthikeyan, S., Yadav, S., Paramasivam, M., Srinivasan, A. and Singh, T. P. Structure of buffalo lactoferrin at 3.3 A resolution at 277 K. *Acta Crystallogr. D Biol. Crystallogr.* 56(6):684-689 (2000).
86. Kassell, B. Trypsin and chymotrypsin inhibitors from soybeans. *Methods Enzymol.* 19:853-862 (1970).
87. Kaur, N., Dhuna, V., Kamboj, S. S., Agrewala, J. N. and Singh, J. A novel antiproliferative and antifungal lectin from *Amaranthus viridis* Linn seeds. *Protein Pept. Lett.* 13(9):897-905 (2006).
88. Kawamura, K., Shibata, T., Saget, O., Peel, D. and Bryant, P. J. A new family of growth factors produced by the fat body and active on *Drosophila* imaginal disc cells. *Development* 126(2):211-219 (1999).
89. Keyaerts, E., Vijgen, L., Pannecouque, C., Van Damme, E., Peumans, W., Egberink, H., Balzarini, J. and Van Ranst, M. Plant lectins are potent inhibitors of coronaviruses by interfering with two targets in the viral replication cycle. *Antiviral Res.* 75(3):179-187 (2007).
90. Khan, A. R. and James, M. N. Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci.* 7(4):815-836 (1998).
91. Koide, T. and Ikenaka, T. Studies on soybean trypsin inhibitors. *Eur. J. Biochem.* 32(3):417-431 (1973).
92. Komath, S. S., Kavitha, M. and Swamy, M. J. Beyond carbohydrate binding: new directions in plant lectin research. *Org. Biomol. Chem.* 4(6):973-988 (2006).
93. Komiyama, T., Ray, C. A., Pickup, D. J., Howard, A. D., Thornberry, N. A., Peterson, E. P. and Salvesen, G. Inhibition of interleukin-1 beta converting enzyme by the cowpox virus serpin CrmA. An example of cross-class inhibition. *J. Biol. Chem.* 269(30):19331-19337 (1994).
94. Kouzuma, Y., Suetake, M., Kimura, M. and Yamasaki, N. Isolation and primary structure of proteinase inhibitors from *Erythrina variegata* (Linn.) var. Orientalis seeds. *Biosci. Biotechnol. Biochem.* 56(11):1819-1824 (1992).
95. Krauchenco, S., Nagem, R. A. P., da Silva, J. A., Marangoni, S. and Polikarpov, I. Three-dimensional structure of an unusual Kunitz (STI) type trypsin inhibitor from Copaifera langsdorffii. *Biochimie* 86(3):167-172 (2004).
96. Krauchenco, S., Pando, S. C., Marangoni, S. and Polikarpov, I. Crystal structure of the Kunitz (STI)-type inhibitor from *Delonix regia* seeds. *Biochem. Biophys. Res. Commun.* 312(4):1303-1308 (2003).
97. Krowarsch, D., Cierpicki, T., Jelen, F. and Otlewski, J. Canonical protein inhibitors of serine proteases. *Cell. Mol. Life Sci.* 60(11):2427-2444 (2003).
98. Kumar, S., Singh, N., Sinha, M., Dube, D., Singh, S. B., Bhushan, A., Kaur, P., Srinivasan, A., Sharma, S. and Singh, T. P. Crystal structure determination and inhibition studies of a novel xylanase and amylase inhibitor protein (XAIP) from *Scadoxus multiflorus*. *FEBS J.* 277(13):2868-2882 (2010).
99. Kunitz, M. Crystalline soybean trypsin inhibitor. *J. Gen. Physiol.* 29(3):149 (1946).
100. Kunitz, M. Crystalline soybean trypsin inhibitor. *J. Gen. Physiol.* 30(4):291 (1947).
101. Kuo, C. J., Liao, Y. C., Yang, J. H., Huang, L. C., Chang, C. T. and Sung, H. Y. Cloning and characterization of an antifungal class III chitinase from suspension-cultured bamboo (*Bambusa oldhamii)* cells. *J. Agric. Food Chem.* 56(23):11507-11514 (2008).
102. Lacy, D. B., Tepp, W., Cohen, A. C., DasGupta, B. R. and Stevens, R. C. Crystal structure of botulinum neurotoxin type A and implications for toxicity. *Nat. Struct. Biol.* 5(10):898-902 (1998).

103. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227(5259):680-685 (1970).
104. Lapatto, R., Krengel, U., Schreuder, H. A., Arkema, A., de Boer, B., Kalk, K. H., Hol, W. G. J., Grootenhuis, P. D. J., Mulders, J. W. M. and Dijkema, R. X-ray structure of antistasin at 1.9 A resolution and its modelled complex with blood coagulation factor Xa. *EMBO J.* 16(17):5151-5161 (1997).
105. Laskowski Jr, M. and Kato, I. Protein inhibitors of proteinases. *Annu. Rev. Biochem.* 49(1):593-626 (1980).
106. Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26(2):283-291 (1993).
107. Lehle, K., Wrba, A. and Jaenicke, R. *Erythrina caffra* trypsin inhibitor retains its native structure and function after reducing its disulfide bonds. *J. Mol. Biol.* 239(2):276-284 (1994).
108. Leonidas, D. D., Swamy, B. M., Hatzopoulos, G. N., Gonchigar, S. J., Chachadi, V. B., Inamdar, S. R., Zographos, S. E. and Oikonomakos, N. G. Structural basis for the carbohydrate recognition of the *Sclerotium rolfsii* lectin. *J. Mol. Biol.* 368(4):1145-1161 (2007).
109. Lin, Z. and Johnson, M. E. Proposed cation- π mediated binding by factor Xa: a novel enzymatic mechanism for molecular recognition. *FEBS Lett.* 370(1-2):1-5 (1995).
110. Lingaraju, M. H. and Gowda, L. R. A Kunitz trypsin inhibitor of *Entada scandens* seeds: another member with single disulfide bridge. *Biochem. Biophys. Acta* 1784(5):850-855 (2008).
111. Lis, H. Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* 98:137-174 (1998).
112. Liu, B., Li, C., Bian, H., Min, M., Chen, L. and Bao, J. Antiproliferative activity and apoptosis-inducing mechanism of Concanavalin A on human melanoma A375 cells. *Arch. Biochem. Biophys.* 482(1-2):1-6 (2009).
113. Loris, R., De Greve, H., Dao-Thi, M. H., Messens, J., Imberty, A. and Wyns, L. Structural basis of carbohydrate recognition by lectin II from Ulex europaeus, a protein with a promiscuous carbohydrate-binding site1. *J. Mol. Biol.* 301(4):987-1002 (2000).
114. Luiza Vilela Oliva, M., da Silva Ferreira, R., Gasperazzo Ferreira, J., Alessandra Andrade de Paula, C., E Salas, C. and Uemura Sampaio, M. Structural and Functional Properties of Kunitz Proteinase Inhibitors from Leguminosae: A Mini Review. *Curr. Protein Pept. Sci.* 12(5):348-357 (2011).
115. Macek, T., Surá, M., Pavliková, D., Francová, K., Scouten, W. H., Szekeres, M., Sylvestre, M. and Macková, M. Can tobacco have potentially beneficial effect to our health. *Z Naturforsch C* 60(3/4):292-299 (2005).
116. Maignan, S. and Mikol, V. The use of 3D structural data in the design of specific factor Xa inhibitors. *Curr. Top. Med. Chem.* 1(2):161-174 (2001).
117. Manithody, C., Yang, L. and Rezaie, A. R. Role of basic residues of the autolysis loop in the catalytic function of factor Xa. *Biochemistry* 41(21):6780-6788 (2002).
118. Mathialagan, N. and Hansen, T. R. Pepsin-inhibitory activity of the uterine serpins. *Proc. Natl. Acad. Sci. U. S. A.* 93(24):13653-13658 (1996).
119. McLachlan, A. D. Three-fold structural pattern in the soybean trypsin inhibitor (Kunitz). *J. Mol. Biol.* 133(4):557-563 (1979).
120. Meester, P. D., Brick, P., Lloyd, L. F., Blow, D. M. and Onesti, S. Structure of the Kunitz-type soybean trypsin inhibitor (STI): implication for the interactions between members of

the STI family and tissue-plasminogen activator. *Acta Crystallogr. D Biol. Crystallogr.* 54(4):589-597 (1998).
121. Meng, L. and Feldman, L. A rapid TRIzol-based two-step method for DNA-free RNA extraction from Arabidopsis siliques and dry seeds. *Biotechnol. J.* 5(2):183-186 (2010).
122. Meyer, A., Rypniewski, W., Szymański, M., Voelter, W., Barciszewski, J. and Betzel, C. Structure of mistletoe lectin I from *Viscum album* in complex with the phytohormone zeatin. *Biochim. Biophys. Acta* 1784(11):1590-1595 (2008).
123. Minoux, H. and Chipot, C. Cation-π Interactions in Proteins: Can Simple Models Provide an Accurate Description? *J. Am. Chem. Soc.* 121(44):10366-10372 (1999).
124. Mishra, V., Bilgrami, S., Sharma, R. S., Kaur, P., Yadav, S., Krauspenhaar, R., Betzel, C., Voelter, W., Babu, C. R. and Singh, T. P. Crystal structure of himalayan mistletoe ribosome-inactivating protein reveals the presence of a natural inhibitor and a new functionally active sugar-binding site. *J. Biol. Chem.* 280(21):20712-20721 (2005).
125. Mishra, V., Sharma, R. S., Paramasivam, M., Bilgrami, S., Yadav, S., Srinivasan, A., Betzel, C., Babu, C. R. and Singh, T. P. cDNA cloning and characterization of a ribosome inactivating protein of a hemi-parasitic plant (*Viscum album* L.) from North-Western Himalaya (India). *Plant Sci.* 168(3):615-625 (2005).
126. Mitra, D., Mukherjee, S. and Das, A. K. Cyclosporin A binding to Mycobacterium tuberculosis peptidyl-prolyl cis-trans isomerase A-investigation by CD, FTIR and fluorescence spectroscopy. *FEBS Lett.* 580(30):6846-6860 (2006).
127. Mohammadi, M., Chalavi, V., Novakova-Sura, M., Laliberté, J. F. and Sylvestre, M. Expression of bacterial biphenyl-chlorobiphenyl dioxygenase genes in tobacco plants. *Biotechnol. Bioeng.* 97(3):496-505 (2007).
128. Mohammadi, M. and Sylvestre, M. Resolving the profile of metabolites generated during oxidation of dibenzofuran and chlorodibenzofurans by the biphenyl catabolic pathway enzymes. *Chem. Biol.* 12(7):835-846 (2005).
129. Mohanty, A. K., Fisher, A. J., Yu, Z., Pradeep, M. A., Janjanam, J. and Kaushik, J. K. Cloning, expression, characterization and crystallization of BRP39, a signalling glycoprotein expressed during mammary gland apoptosis. *Protein Expr. Purif.* 64(2):213-218 (2009).
130. Mohanty, A. K., Singh, G., Paramasivam, M., Saravanan, K., Jabeen, T., Sharma, S., Yadav, S., Kaur, P., Kumar, P. and Srinivasan, A. Crystal structure of a novel regulatory 40-kDa mammary gland protein (MGP-40) secreted during involution. *J. Biol. Chem.* 278(16):14451-14460 (2003).
131. Mosmann, T. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J. Immunol. Methods* 65(1-2):55-63 (1983).
132. Moza, B., Varma, A. K., Buonpane, R. A., Zhu, P., Herfst, C. A., Nicholson, M. J., Wilbuer, A. K., Seth, N. P., Wucherpfennig, K. W. and McCormick, J. K. Structural basis of T-cell specificity and activation by the bacterial superantigen TSST-1. *EMBO J.* 26(4):1187-1197 (2007).
133. Mukherjee, S., Dhar, R. and Das, A. K. Analyzing the catalytic mechanism of protein tyrosine phosphatase PtpB from *Staphylococcus aureus* through site-directed mutagenesis. *Int. J. Biol. Macromol.* 45(5):463-469 (2009).
134. Mukherjee, S., Dutta, D., Saha, B. and Das, A. K. Crystal Structure of Glyceraldehyde-3-Phosphate Dehydrogenase 1 from Methicillin-Resistant *Staphylococcus aureus* MRSA252 Provides Novel Insights into Substrate Binding and Catalytic Mechanism. *J. Mol. Biol.* 401(5):949-968 (2010).

135. Murshudov, G. N., Vagin, A. A. and Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53(3):240-255 (1997).
136. Murzin, A. G., Lesk, A. M. and Chothia, C. [beta]-Trefoil fold:: Patterns of structure and sequence in the Kunitz inhibitors interleukins-1 [beta] and 1 [alpha] and fibroblast growth factors. *J. Mol. Biol.* 223(2):531-543 (1992).
137. Neurath, H. Evolution of proteolytic enzymes. *Science* 224(4647):350-357 (1984).
138. Neurath, H. Proteolytic processing and physiological regulation. *Trends Biochem. Sci.* 14(7):268-271 (1989).
139. Ngai, P. H. K. and Ng, T. B. A lectin with antifungal and mitogenic activities from red cluster pepper (*Capsicum frutescens*) seeds. *Appl. Microbiol. Biotechnol.* 74(2):366-371 (2007).
140. Norioka, N., Hara, S., Ikenaka, T. and Abe, J. Distribution of the Kunitz and the Bowman-Birk Family Proteinase Inhibitors in Leguminous Seeds. *Agric. Biol. Chem.* 52(5):1245-1252 (1988).
141. Oliva, M. L. V., Andrade, S. A., Juliano, M. A., Sallai, R. C., Torquato, R. J., Sampaio, M. U., Pott, V. J. and Sampaio, C. A. M. Kinetic characterization of factor Xa binding using a quenched fluorescent substrate based on the reactive site of factor Xa inhibitor from *Bauhinia ungulata* seeds. *Curr. Med. Chem.* 10(13):1085-1093 (2003).
142. Oliva, M. L. V., Silva, M. C. C., Sallai, R. C., Brito, M. V. and Sampaio, M. U. A novel subclassification for Kunitz proteinase inhibitors from leguminous seeds. *Biochimie* 92(11):1667-1673 (2010).
143. Oliva, M. L. V., Souza-Pinto, J. C., Batista, I. F. C., Araujo, M. S., Silveira, V. F., Auerswald, E. A., Mentele, R., Eckerskorn, C., Sampaio, M. U. and Sampaio, C. A. M. Leucaena leucocephala serine proteinase inhibitor: primary structure and action on blood coagulation, kinin release and rat paw edema1. *Biochim. Biophys. Acta* 1477(1-2):64-74 (2000).
144. Onesti, S., Brick, P. and Blow, D. M. Crystal structure of a Kunitz-type trypsin inhibitor from *Erythrina coffra* seeds. *J. Mol. Biol.* 217(1):153-176 (1991).
145. Otlewski, J., Krowarsch, D. and Apostoluk, W. Protein inhibitors of serine proteinases. *Acta Biochim. Pol.* 46(3):531-565 (1999).
146. Otwinowski, Z. and Minor, W. [20] Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276:307-326 (1997).
147. Owhashi, M., Arita, H. and Hayai, N. Identification of a novel eosinophil chemotactic cytokine (ECF-L) as a chitinase family protein. *J. Biol. Chem.* 275(2):1279-1286 (2000).
148. Payan, F., Flatman, R., Porciero, S., Williamson, G., Juge, N. and Roussel, A. Structural analysis of xylanase inhibitor protein I (XIP-I), a proteinaceous xylanase inhibitor from wheat (*Triticum aestivum*, var. Soisson). *Biochem. J.* 372(Pt 2):399-405 (2003).
149. Payan, F., Leone, P., Porciero, S., Furniss, C., Tahir, T., Williamson, G., Durand, A., Manzanares, P., Gilbert, H. J. and Juge, N. The dual nature of the wheat xylanase protein inhibitor XIP-I. *J. Biol. Chem.* 279(34):36029-36037 (2004).
150. Peumans, W. J. and Van Damme, E. J. Lectins as plant defense proteins. *Plant Physiol.* 109(2):347-352 (1995).
151. Polito, L., Bortolotti, M., Farini, V., Battelli, M. G., Barbieri, L. and Bolognesi, A. Saporin induces multiple death pathways in lymphoma cells with different intensity and timing as compared to ricin. *Int. J. Biochem. Cell Biol.* 41(5):1055-1061 (2009).
152. Potempa, J., Korzus, E. and Travis, J. The serpin superfamily of proteinase inhibitors: structure, function, and regulation. *J. Biol. Chem.* 269(23):15957-15960 (1994).

153. Qasim, M. A., Van Etten, R. L., Yeh, T., Saunders, C., Ganz, P. J., Qasim, S., Wang, L. and Laskowski Jr, M. Despite Having a Common P1 Leu, Eglin C Inhibits alpha-lytic proteinase a Million-fold More Strongly than Does Turkey Ovomucoid Third Domain. *Biochemistry* 45(38):11342-11348 (2006).
154. Quick, A. J. The prothrombin consumption time test. *Am. J. Clin. Pathol.* 45(1):105-109 (1966).
155. Raghavan, S., Reddy, S. R., Tony, K. A., Kumar, C. N., Varma, A. K. and Nangia, A. Sulfinyl group as an intramolecular nucleophile: synthesis of bromohydrins from beta-methyl-gamma,delta-unsaturated sulfoxides with high 1,2-asymmetric induction. *J. Org. Chem.* 67(16):5838-5841 (2002).
156. Rai, R., Sprengeler, P. A., Elrod, K. C. and Young, W. B. Perspectives on factor Xa inhibition. *Curr. Med. Chem.* 8(2):101-119 (2001).
157. Rao, D. H. and Gowda, L. R. Abundant class III acidic chitinase homologue in tamarind (Tamarindus indica) seed serves as the major storage protein. *J. Agric. Food Chem.* 56(6):2175-2182 (2008).
158. Ravichandran, S., Sen, U., Chakrabarti, C. and Dattagupta, J. K. Cryocrystallography of a Kunitz-type serine protease inhibitor: the 90 K structure of winged bean chymotrypsin inhibitor (WCI) at 2.13 A resolution. *Acta Crystallogr. D Biol. Crystallogr.* 55(11):1814-1821 (1999).
159. Rawlings, N. D. Peptidase inhibitors in the MEROPS database. *Biochimie* 92(11):1463-1483 (2010).
160. Reuter, G. and Gabius, H. J. Eukaryotic glycosylation: whim of nature or multipurpose tool? *Cell. Mol. Life Sci.* 55(3):368-422 (1999).
161. Rezaie, A. R. Role of residue 99 at the S2 subsite of factor Xa and activated protein C in enzyme specificity. *J. Biol. Chem.* 271(39):23807-23814 (1996).
162. Rezaie, A. R. and Esmon, C. T. Contribution of residue 192 in factor Xa to enzyme specificity and function. *J. Biol. Chem.* 270(27):16176-16181 (1995).
163. Ribeiro, J. K. C., Cunha, D. D. S., Fook, J. M. and Sales, M. P. New properties of the soybean trypsin inhibitor: Inhibition of human neutrophil elastase and its effect on acute pulmonary injury. *Eur. J. Pharmacol.* 644(1-3):238-244 (2010).
164. Roy, S., Aravind, P., Madhurantakam, C., Ghosh, A. K., Sankaranarayanan, R. and Das, A. K. Crystallization and preliminary X-ray diffraction analysis of a protease inhibitor from the haemolymph of the Indian tasar silkworm *Antheraea mylitta*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 62(7):669-671 (2006).
165. Roy, S., Aravind, P., Madhurantakam, C., Ghosh, A. K., Sankaranarayanan, R. and Das, A. K. Crystal structure of a fungal protease inhibitor from *Antheraea mylitta*. *J. Struct. Biol.* 166(1):79-87 (2009).
166. Roychaudhuri, R., Sarath, G., Zeece, M. and Markwell, J. Reversible denaturation of the soybean Kunitz trypsin inhibitor. *Arch. Biochem. Biophys.* 412(1):20-26 (2003).
167. Roychaudhuri, R., Sarath, G., Zeece, M. and Markwell, J. Stability of the allergenic soybean Kunitz trypsin inhibitor. *Biochim. Biophys. Acta* 1699(1):207-212 (2004).
168. Rüdiger, H. On the physiological role of plant lectins. *Bioscience* 34(2):95-99 (1984).
169. Rüdiger, H. and Gabius, H. J. Plant lectins: occurrence, biochemistry, functions and applications. *Glycoconj. J.* 18(8):589-613 (2001).
170. Rutenber, E., Katzin, B. J., Ernst, S., Collins, E. J., Mlsna, D., Ready, M. P. and Robertus, J. D. Crystallographic refinement of ricin to 2.5 A. *Proteins* 10(3):240-250 (1991).
171. Sá, R. A., Santos, N. D. L., Silva, C. S. B., Napoleão, T. H., Gomes, F. S., Cavada, B. S., Coelho, L. C. B. B., Navarro, D. M. A. F., Bieber, L. W. and Paiva, P. M. G. Larvicidal

activity of lectins from *Myracrodruon urundeuva* on *Aedes aegypti*. *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* 149(3):300-306 (2009).
172. Sampaio, C. A. M., Oliva, M. L. V., Sampaio, M. U., Batista, I. F. C., Bueno, N. R., Tanaka, A. S., Auerswald, E. A. and Fritz, H. Plant serine proteinase inhibitors. Structure and biochemical applications on plasma kallikrein and related enzymes. *Immunopharmacology* 32(1-3):62-66 (1996).
173. Schechter, I. and Berger, A. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* 27(2):157-162 (1967).
174. Schlesier, B., Dijkstra, B. W., Terwisscha van Scheltinga, A. C., Jansonius, J. N. and Hennig, M. Crystal Structure of Concanavalin B at 1.65 Å Resolution. An "Inactivated"Chitinase from Seeds of *Canavalia ensiformis*. *J. Mol. Biol.* 254:237-246 (1995).
175. Schuttelkopf, A. W. and Van Aalten, D. M. F. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* 60(8):1355-1363 (2004).
176. Schwefel, D., Maierhofer, Ç., Beck, J. G., Seeberger, S., Diederichs, K., Mo ller, H. M., Welte, W. and Wittmann, V. Structural basis of multivalent binding to wheat germ agglutinin. *J. Am. Chem. Soc.* 132(25):8704-8719 (2010).
177. Seifert, G., Jesse, P., Laengler, A., Reindl, T., Lüth, M., Lobitz, S., Henze, G., Prokop, A. and Lode, H. N. Molecular mechanisms of mistletoe plant extract-induced apoptosis in acute lymphoblastic leukemia in vivo and in vitro. *Cancer Lett.* 264(2):218-228 (2008).
178. Shackelton, L. M., Mann, D. M. and Millis, A. J. T. Identification of a 38-kDa heparin-binding glycoprotein (gp38k) in differentiating vascular smooth muscle cells as a member of a group of proteins associated with tissue remodeling. *J. Biol. Chem.* 270(22):13076-13083 (1995).
179. Sharma, M., Ethayathulla, A. S., Jabeen, T., Singh, N., Sarvanan, K., Yadav, S., Sharma, S., Srinivasan, A. and Singh, T. P. Crystal structure of a highly acidic neurotoxin from scorpion *Buthus tamulus* at 2.2A resolution reveals novel structural features. *J. Struct. Biol.* 155(1):52-62 (2006).
180. Sharma, S., Jasti, J., Kumar, J., Mohanty, A. K. and Singh, T. P. Crystal structure of a proteolytically generated functional monoferric C-lobe of bovine lactoferrin at 1.9 A resolution. *J. Mol. Biol.* 331(2):485-496 (2003).
181. Sharma, S., Karthikeyan, S., Betzel, C. and Singh, T. P. Isolation, purification, crystallization and preliminary X-ray analysis of 1-bungarotoxin from *Bungarus caeruleus* (Indian common krait). *Acta Crystallogr. D Biol. Crystallogr.* 55(5):1093-1094 (1999).
182. Sharon, N. Lectins. *Sci. Am.* 236(6):108-116 (1977).
183. Shewry, P. R. Plant storage proteins. *Biol. Rev.* 70(3):375-426 (1995).
184. Shewry, P. R. and Lucas, J. A. Plant proteins that confer resistance to pests and pathogens. *Adv. Bot. Res.* 26:135-192 (1997).
185. Singh, M., Kumar, P. and Karthikeyan, S. Structural basis for pH dependent monomer-dimer transition of 3, 4-dihydroxy 2-butanone-4-phosphate synthase domain from *M. tuberculosis*. *J. Struct. Biol.* 174(2):374-384 (2011).
186. Somashekar, D. and Joseph, R. A new spectrophotometric method of assay for chitosanase based on calcofluor white dye binding. *Carbohyd. Polym.* 34(4):343-346 (1997).
187. Song, H. K. and Suh, S. W. Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator1. *J. Mol. Biol.* 275(2):347-363 (1998).

188. Stein, P. E. and Carrell, R. W. What do dysfunctional serpins tell us about molecular mobility and disease? *Nat. Struct. Biol.* 2(2):96-113 (1995).
189. Sumner, J. B. and Howell, S. F. Identification of Hemagglutinin of Jack Bean with Concanavalin A. *J. Bacteriol.* 32(2):227-237 (1936).
190. Sun, Y. J., Chang, N. C. A., Hung, S. I., Chang, A. C., Chou, C. C. and Hsiao, C. D. The crystal structure of a novel mammalian lectin, Ym1, suggests a saccharide binding site. *J. Biol. Chem.* 276(20):17507-17514 (2001).
191. Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H. and Blow, D. M. Crystal structure of the complex of porcine trypsin with soybean trypsin inhibitor (Kunitz) at 2.6 resolution. *Biochemistry* 13(20):4212-4228 (1974).
192. Tai, H. H., Pelletier, C. and Beardmore, T. Total RNA isolation from Picea mariana dry seed. *Plant Mol. Biol. Rep.* 22(1):93-93 (2004).
193. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28(10):2731-2739 (2011).
194. Terwisscha van Scheltinga, A. C., Armand, S., Kalk, K. H., Isogai, A., Henrissat, B. and Dijkstra, B. W. Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and x-ray structure of a complex with allosamidin evidence for substrate assisted catalysis. *Biochemistry* 34(48):15619-15623 (1995).
195. Tews, I., van Scheltinga, A. C. T., Perrakis, A., Wilson, K. S. and Dijkstra, B. W. Substrate-assisted catalysis unifies two families of chitinolytic enzymes. *J. Am. Chem. Soc.* 119(34):7954-7959 (1997).
196. Thompson, J. D., Higgins, D. G. and Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22(22):4673-4680 (1994).
197. Thompson, M. G. and Palmer, R. M. Signalling pathways regulating protein turnover in skeletal muscle. *Cell. signal.* 10(1):1-11 (1998).
198. Tina, K. G., Bhadra, R. and Srinivasan, N. PIC: protein interactions calculator. *Nucleic Acids Res.* 35(suppl 2):W473-W476 (2007).
199. Tomar, A. K., Sooch, B. S., Raj, I., Singh, S., Singh, T. P. and Yadav, S. Isolation and identification of Concanavalin A binding glycoproteins from human seminal plasma: A step towards identification of male infertility marker proteins. *Dis. Markers* 31(6):379-386 (2011).
200. Tomar, A. K., Sooch, B. S. and Yadav, S. Computational analysis of Concanavalin A binding glycoproteins of human seminal plasma. *Bioinformation* 7(2):69-75 (2011).
201. Transue, T. R., Smith, A. K., Mo, H., Goldstein, I. J. and Saper, M. A. Structure of benzyl T-antigen disaccharide bound to *Amaranthus caudatus* agglutinin. *Nat. Struct. Biol.* 4(10):779-783 (1997).
202. Travis, J. and Salvesen, G. S. Human plasma proteinase inhibitors. *Annu. Rev. Biochem.* 52(1):655-709 (1983).
203. Van Damme, E. J. M., Culerrier, R., Barre, A., Alvarez, R., Rougé, P. and Peumans, W. J. A novel family of lectins evolutionarily related to class V chitinases: an example of neofunctionalization in legumes. *Plant Physiol.* 144(2):662 (2007).
204. Van Damme, E. J. M., Lannoo, N. and Peumans, W. J. Plant lectins. *Adv. Bot. Res.* 48:107-209 (2008).
205. van Scheltinga, A. C. T., Hennig, M. and Dijkstra, B. W. The 1.8 Å resolution structure of hevamine, a plant chitinase/lysozyme, and analysis of the conserved sequence and structure motifs of glycosyl hydrolase family 18. *J. Mol. Biol.* 262:243-257 (1996).

206. van Wijk, E. M. and Smit, J. L. J. A chromogenic micro-assay for factor X-activity. *Thromb. Res.* 34(3):263-268 (1984).
207. Vandenborre, G., Smagghe, G. and Van Damme, E. J. M. Plant lectins as defense proteins against phytophagous insects. *Phytochemistry* 72(13):1538-1550 (2011).
208. Varela, P. F., Llera, A. S., Mariuzza, R. A. and Tormo, J. Crystal structure of imaginal disc growth factor-2. *J. Biol. Chem.* 277(15):13229-13236 (2002).
209. Vézina, J., Barriault, D. and Sylvestre, M. Family shuffling of soil DNA to change the regiospecificity of *Burkholderia xenovorans* LB400 biphenyl dioxygenase. *J. Bacteriol.* 189(3):779-788 (2007).
210. Vézina, J., Barriault, D. and Sylvestre, M. Diversity of the C-terminal portion of the biphenyl dioxygenase large subunit. *J. Mol. Microbiol. Biotechnol.* 15(2-3):139-151 (2008).
211. Vijayan, M. and Chandra, N. Lectins. *Curr. Opin. Struc. Biol.* 9(6):707-714 (1999).
212. Waheed, A., Hassan, M. I., Van Etten, R. L. and Ahmad, F. Human seminal proteinase and prostate-specific antigen are the same protein. *J. Bioscience* 33(2):195-207 (2008).
213. Wang, S., Yu, Q., Bao, J. and Liu, B. Polygonatum cyrtonema lectin, a potential antineoplastic drug targeting programmed cell death pathways. *Biochem. Biophys. Res. Commun.* 406(4):497-500 (2011).
214. Wang, S. X., Hur, E., Sousa, C. A., Brinen, L., Slivka, E. J. and Fletterick, R. J. The extended interactions and Gla domain of blood coagulation factor Xa. *Biochemistry* 42(26):7959-7966 (2003).
215. Watanabe, T., Kobori, K., Miyashita, K., Fujii, T., Sakai, H., Uchida, M. and Tanaka, H. Identification of glutamic acid 204 and aspartic acid 200 in chitinase A1 of *Bacillus circulans* WL-12 as essential residues for chitinase activity. *J. Biol. Chem.* 268(25):18567-18572 (1993).
216. Wei, A., Alexander, R. S., Duke, J., Ross, H., Rosenfeld, S. A. and Chang, C. H. Unexpected binding mode of tick anticoagulant peptide complexed to bovine factor Xa1. *J. Mol. Biol.* 283(1):147-154 (1998).
217. Wei, Z., Yan, Y., Carrell, R. W. and Zhou, A. Crystal structure of protein Z-dependent inhibitor complex shows how protein Z functions as a cofactor in the membrane inhibition of factor X. *Blood* 114(17):3662-3667 (2009).
218. Wright, C. S. The crystal structure of wheat germ agglutinin at 2-2 A resolution. *J. Mol. Biol.* 111(4):439-457 (1977).
219. Wu, A. M., Lisowska, E., Duk, M. and Yang, Z. Lectins as tools in glycoconjugate research. *Glycoconj. J.* 26(8):899-913 (2009).
220. Yadav, S., Tomar, A. K., Jithesh, O., Khan, M. A., Yadav, R. N., Srinivasan, A. and Singh, T. P. Purification and Partial Characterization of Low Molecular Weight Vicilin-Like Glycoprotein from the Seeds of *Citrullus lanatus*. *Protein J.* 30(8):1-6 (2011).
221. Ye, X. Y., Ng, T. B., Tsang, P. W. K. and Wang, J. Isolation of a homodimeric lectin with antifungal and antiviral activities from red kidney bean (*Phaseolus vulgaris*) seeds. *J. Protein Chem.* 20(5):367-375 (2001).
222. Zabell, A. P. R., Schroff Jr, A. D., Bain, B. E., Van Etten, R. L., Wiest, O. and Stauffacher, C. V. Crystal structure of the human B-form low molecular weight phosphotyrosyl phosphatase at 1.6-A resolution. *J. Biol. Chem.* 281(10):6520-6527 (2006).
223. Zhu, K., Huesing, J. E., Shade, R. E., Bressan, R. A., Hasegawa, P. M. and Murdock, L. L. An insecticidal N-acetylglucosamine-specific lectin gene from *Griffonia simplicifolia* (Leguminosae). *Plant Physiol.* 110(1):195-202 (1996).